Analysis & PDE

Volume 17 No. 2 2024

msp

Analysis & PDE

msp.org/apde

See inside back cover or msp.org/apde for submission instructions.

The subscription price for 2024 is US $440/year for the electronic version, and $690/year (+$65, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscriber address should be sent to MSP.

Analysis & PDE (ISSN 1948-206X electronic, 2157-5045 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online.

APDE peer review and production are managed by EditFlow® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

ANALYSIS AND PDE
Vol. 17 (2024), No. 2, pp. 379–420
DOI: 10.2140/apde.2024.17.379

ON A SPATIALLY INHOMOGENEOUS NONLINEAR FOKKER–PLANCK EQUATION: CAUCHY PROBLEM AND DIFFUSION ASYMPTOTICS

FRANCESCA ANCESCHI AND YUZHE ZHU

We investigate the Cauchy problem and the diffusion asymptotics for a spatially inhomogeneous kinetic model associated to a nonlinear Fokker–Planck operator. We derive the global well-posedness result with instantaneous smoothness effect, when the initial data lies below a Maxwellian. The proof relies on the hypoelliptic analog of classical parabolic theory, as well as a positivity-spreading result based on the Harnack inequality and barrier function methods. Moreover, the scaled equation leads to the fast diffusion flow under the low field limit. The relative phi-entropy method enables us to see the connection between the overdamped dynamics of the nonlinearly coupled kinetic model and the correlated fast diffusion. The global-in-time quantitative diffusion asymptotics is then derived by combining entropic hypocoercivity, relative phi-entropy, and barrier function methods.

1. Introduction

We consider the kinetic Fokker–Planck operator $\mathscr{L}_{\mathrm{FP}} := \nabla_v \cdot (\nabla_v + v)$ and the spatially inhomogeneous nonlinear drift-diffusion model

$$\begin{cases} (\partial_t + v \cdot \nabla_x) f(t, x, v) = \rho_f^{\beta}(t, x)\, \mathscr{L}_{\mathrm{FP}} f(t, x, v), \\ f(0, x, v) = f_{\mathrm{in}}(x, v), \end{cases} \tag{1-1}$$

for an unknown $f(t, x, v) \geq 0$ with $t \in \mathbb{R}_+,\ x \in \mathbb{T}^d$ or $\mathbb{R}^d,\ v \in \mathbb{R}^d$, where $\mathbb{T}^d$ denotes the d-dimensional torus with unit volume, the constant $\beta \in [0, 1]$, and

$$\rho_f(t, x) := \int_{\mathbb{R}^d} f(t, x, v)\, \mathrm{d}v.$$

MSC2020: 35A01, 35B40, 35Q35, 35Q84.
Keywords: nonlinear kinetic Fokker–Planck equation, well-posedness, regularity, diffusion asymptotics.

Given a constant $\epsilon \in (0, 1)$, the equation under the low field scaling $t \mapsto \epsilon^2 t,\ x \mapsto \epsilon x$ reads

$$\begin{cases} (\epsilon \partial_t + v \cdot \nabla_x) f_\epsilon(t, x, v) = \frac{1}{\epsilon} \rho_{f_\epsilon}^{\beta}(t, x)\, \mathscr{L}_{\mathrm{FP}}\, f_\epsilon(t, x, v), \\ f_\epsilon(0, x, v) = f_{\epsilon,\mathrm{in}}(x, v). \end{cases} \tag{1-2}$$

Our aim is to show the global well-posedness and the trend to equilibrium with smoothness a priori estimates for (1-1), and the quantitative asymptotic dynamics of (1-2) as ϵ tends to zero.

1A. ***Main results.*** Let us recall that a classical solution of an evolution equation is a nonnegative function satisfying the equation pointwise everywhere and matching the initial data continuously. Unless otherwise specified, any solution we consider below is intended in the classical sense. For $k \in \mathbb{N}$, define $\mathcal{C}^k(\Omega)$ to be the set of functions having all derivatives of order less than or equal to k continuous in the domain Ω. For $\alpha \in (0, 1)$, we note that $\mathcal{C}^\alpha(\Omega)$ is the classical Hölder space on Ω with exponent α. In addition, we define the measure $\mathrm{d}m := \mathrm{d}x\, \mathrm{d}\mu$, where

$$\mu(v) := (2\pi)^{-d/2} e^{-|v|^2/2} \quad \text{and} \quad \mathrm{d}\mu := \mu\, \mathrm{d}v$$

denote the Gaussian function and the Gaussian measure, respectively. A function that takes the form of $C\mu(v)$ for some constant $C > 0$ is called a Maxwellian.

Theorem 1.1. *Let the space domain Ω_x be equal to $\mathbb{T}^d$ or $\mathbb{R}^d$ and the constants $0 < \lambda < \Lambda$ be given.*

(i) *If $f_{\mathrm{in}} \in \mathcal{C}^0(\Omega_x \times \mathbb{R}^d)$ satisfies $0 \le f_{\mathrm{in}} \le \Lambda\mu$ in $\Omega_x \times \mathbb{R}^d$, then there exists a solution f to the Cauchy problem (1-1) such that $0 \le f \le \Lambda\mu$ in $\mathbb{R}_+ \times \Omega_x \times \mathbb{R}^d$. Moreover, for any $\nu \in (0, 1)$, $k \in \mathbb{N}$, and any compact subset $K \subset (0, T] \times \Omega_x$, there is some constant $C_{T,\nu,k,K} > 0$ depending only on d, β, λ, Λ, T, ν, k, K, and the initial data such that*

$$\|\mu^{-\nu} f\|_{\mathcal{C}^k(K \times \mathbb{R}^d)} \le C_{T,\nu,k,K}.$$

Additionally, if f_{in} is Hölder continuous and $\rho_{f_{\mathrm{in}}} \ge \lambda$ in Ω_x, then the solution that lies below any Maxwellian is unique.

(ii) *For $\Omega_x = \mathbb{T}^d$, if the initial data satisfies $\lambda\mu \le f_{\mathrm{in}} \le \Lambda\mu$ in $\mathbb{T}^d \times \mathbb{R}^d$, then, for any $k \in \mathbb{N}$, there exists some constant $c > 0$ depending only on d, β, λ, Λ and some constant $C_k > 0$ depending additionally on k such that, for any $t \ge 1$,*

$$\left\| \frac{f - \mu \int f_{\mathrm{in}}\, \mathrm{d}x\, \mathrm{d}v}{\sqrt{\mu}} \right\|_{\mathcal{C}^k(\mathbb{T}^d \times \mathbb{R}^d)} \le C_k e^{-ct}.$$

For $\Omega_x = \mathbb{R}^d$, if the initial data satisfies $\lambda\mu \le f_{\mathrm{in}} \le \Lambda\mu$ in $\mathbb{R}^d \times \mathbb{R}^d$ and $f_{\mathrm{in}} - M_1\mu \in L^1(\mathbb{R}^d \times \mathbb{R}^d)$ for some constant $M_1 > 0$, then there is some constant $C' > 0$ depending only on d, β, λ, Λ, M_1 such that

$$\left\| \frac{f - M_1\mu}{\sqrt{\mu}} \right\|_{L^2(\mathbb{R}^d \times \mathbb{R}^d)} \le C'(1 + \|f_{\mathrm{in}} - M_1\mu\|_{L^1(\mathbb{R}^d \times \mathbb{R}^d)}) t^{-d/4}.$$

Remark 1.2. If the general measurable initial data f_{in} satisfies $f_{\mathrm{in}} \le \Lambda\mu$ and an extra locally uniform lower bound assumption (see (4-14) below for a precise description), the existence of solutions still holds in some weak sense, as pointed out in Remark 4.9 below.

In order to describe the diffusion asymptotics of (1-2), we introduce the (Bregman) distance characterized by the relative phi-entropy functional $\mathcal{H}_\beta$.

Definition 1.3. Let $\beta \in [0, 1]$. For any measurable functions $h_1 \geq 0$ and $h_2 > 0$ defined in $\mathbb{T}^d \times \mathbb{R}^d$, the *relative phi-entropy* of h_1 with respect to h_2 is defined by

$$\mathcal{H}_\beta(h_1 \,|\, h_2) := \int_{\mathbb{T}^d \times \mathbb{R}^d} (\varphi_\beta(h_1) - \varphi_\beta(h_2) - \varphi'_\beta(h_2)(h_1 - h_2))\, \mathrm{d}m,$$

where $\varphi_\beta : \mathbb{R}_+ \to \mathbb{R}$ is defined by

$$\varphi_\beta(z) := \frac{1}{1-\beta}(z^{2-\beta} - (2-\beta)z + 1 - \beta)$$

for $\beta \in [0, 1)$ and $\varphi_1(z) := z \log z - z + 1$.

Theorem 1.4. *Let the constants $\alpha_0 \in (0, 1)$ and $0 < \lambda < \Lambda$ be given, and consider a sequence of functions $\{f_{\epsilon,\mathrm{in}}\}_{\epsilon \in (0,1)} \subset \mathcal{C}^{\alpha_0}(\mathbb{T}^d \times \mathbb{R}^d)$ satisfying $0 \leq f_{\epsilon,\mathrm{in}} \leq \Lambda\mu$ in $\mathbb{T}^d \times \mathbb{R}^d$ and $\rho_{f_{\epsilon,\mathrm{in}}} \geq \lambda$ in $\mathbb{T}^d$. Let f_ϵ be the solution to (1-2) associated with the initial data $f_{\epsilon,\mathrm{in}}$.*

(i) *If there exists some constant $\epsilon' \in (0, 1)$ and some function $\rho_{\mathrm{in}} \in \mathcal{C}^{\alpha_0}(\mathbb{T}^d)$ valued in $[\lambda, \Lambda]$ such that*

$$\mathcal{H}_\beta(\mu^{-1} f_{\epsilon,\mathrm{in}} \,|\, \rho_{\mathrm{in}}) \leq \epsilon',$$

then there exist constants $M, m > 0$ depending only on d, β, λ, Λ, α_0, $\|\rho_{\mathrm{in}}\|_{\mathcal{C}^{\alpha_0}(\mathbb{T}^d)}$, and $\|f_{\epsilon,\mathrm{in}}\|_{\mathcal{C}^{\alpha_0}(\mathbb{T}^d \times \mathbb{R}^d)}$ such that, for any $T > 0$,

$$\|\mu^{-1} f_\epsilon - \rho\|_{L^\infty([0,T]; L^2(\mathbb{T}^d \times \mathbb{R}^d, \mathrm{d}m))} \leq M e^{MT} (\epsilon + \epsilon')^m,$$

where $\rho(t, x)$ for $(t, x) \in \mathbb{R}_+ \times \mathbb{T}^d$ is the solution to the fast diffusion equation

$$\begin{cases} \partial_t \rho(t, x) = \nabla_x \cdot (\rho^{-\beta}(t, x) \nabla_x \rho(t, x)), \\ \rho(0, x) = \rho_{\mathrm{in}}(x). \end{cases} \tag{1-3}$$

(ii) *If we additionally assume that $f_{\epsilon,\mathrm{in}} \geq \lambda\mu$ in $\mathbb{T}^d \times \mathbb{R}^d$, then there exist some constants $M', m' > 0$ with the same dependence as M and m such that*

$$\|\mu^{-1} f_\epsilon - \rho\|_{L^\infty(\mathbb{R}_+; L^2(\mathbb{T}^d \times \mathbb{R}^d, \mathrm{d}m))} \leq M'(\epsilon + \epsilon')^{m'}.$$

1B. *Strategy and background.*

1B1. *Cauchy problem of the nonlinear model.* The well-posedness of the nonlinear model (1-1) was first studied in [Imbert and Mouhot 2021] mixing Hölder and Sobolev spaces in the torus, and in [Liao et al. 2018] under the regime of perturbation to the global equilibrium in the whole space. We develop well-posedness with rough initial data by means of the combination of the hypoelliptic analog of the parabolic theory with a positivity-spreading result; in particular, the technique we employ allows us to drop the smallness and lower bound assumptions asserted in Theorem 1.1. In addition, the global behavior of solutions to (1-1) is derived under the assumption of upper and lower bounds on the initial data only.

When the drift-diffusion coefficient ρ_f^β in (1-1) is proportional to the local mass of the solution — that is when $\beta = 1$ — (1-1) and (1-2) have the same quadratic homogeneity as the Landau equation, but simpler global bounds and conservation laws. Due to the complex structure of the Landau equation, most of the existing results for its classical solutions are about the global theory under the near Maxwellian equilibrium regime [Guo 2002; Kim et al. 2020] and about the local well-posedness associated with low regularity and nonperturbative initial data [Henderson et al. 2019; 2020a]. By contrast, the boundedness from above and from below by Maxwellians of the initial data will be preserved along time for the solutions to (1-1) and (1-2), and the lack of conservation of momentum and energy of (1-2) reduces its hydrodynamic limit to the fast diffusion flow (1-3) rather than the Navier–Stokes dynamics of the scaling limit of the Landau equation, which makes its Cauchy problem and global behavior more tractable in a very general setting.

To address the nonlinear Cauchy problem subject to only one requirement that the initial data lies below a Maxwellian, we propose a method that involves several ingredients. First, in Section 3 we carry out a preliminary study on the linear counterpart of (1-1) — that is the Cauchy problem associated to the Kolmogorov operator

$$\mathscr{L}_1 := \partial_t + v \cdot \nabla_x - \operatorname{tr}(A(t,x,v) D_v^2 \cdot) + B(t,x,v) \cdot \nabla_v, \tag{1-4}$$

where the coefficients including the entries of the positive definite $d \times d$ real symmetric matrix A and the d-dimensional vector B are Hölder continuous (B is not necessarily bounded over $v \in \mathbb{R}^d$). Even if the well-posedness theory for the Cauchy problem associated to the linear operator (1-4) was already well developed in some sense in the existing literature (see [Anceschi and Polidoro 2020; Manfredini 1997]), the Hölder spaces (see Definition 2.3) considered in those works are different from those studied in [Imbert and Mouhot 2021; Imbert and Silvestre 2021] (see Definition 2.1), which are the ones we study. Indeed, in contrast to [Imbert and Mouhot 2021], the (Schauder-type) a priori estimates proved in the previous literature are weaker and not appropriate for bootstrap arguments proving higher regularity for nonlinear problems (see Section 4C).

Secondly, the treatment of the existence issue for (1-1) in Hölder spaces is based on a fixed-point argument, where the compactness is provided by hypoelliptic regularization results; see Section 4B. A breakthrough on such a priori estimates for spatially inhomogeneous kinetic equations with a quasilinear diffusive structure in velocity was obtained in [Golse et al. 2019] and [Henderson and Snelson 2020; Imbert and Mouhot 2021], where the authors prove the kinetic (hypoelliptic) counterparts of the De Giorgi–Nash–Moser theory and the Schauder theory for classical elliptic equations (see for instance [Gilbarg and Trudinger 2001]), respectively. One may refer to [Mouhot 2018] for a summary. Armed with the Schauder estimate developed in [Imbert and Mouhot 2021] in kinetic Hölder spaces and the bootstrap procedure developed in [Imbert and Silvestre 2022] adapted to this case, we are then able to derive instantaneous $\mathcal{C}^\infty$ regularization for the solutions to (1-1) in Section 4C, provided that the solution is bounded from above and bounded away from vacuum, which guarantees the ellipticity in the velocity variable for (1-1).

Thirdly, in order to remove the lower bound assumption on the initial data, in Section 4A we establish a self-generating lower bound result showing that the positivity of solutions spreads everywhere instantaneously. Its proof is based on repeated applications of the spreading of positivity forward in time (see Lemma 4.5) and the spreading for all velocities (see Lemma 4.6), as proposed in

[Henderson et al. 2020b]. On the one hand, the barrier function argument will be used in the same spirit as [Henderson et al. 2020b] to show Lemma 4.5. Indeed, a *lower* (*resp. upper*) *barrier* for a certain equation is a subsolution (resp. supersolution) of the equation which bounds its solution from below (resp. above) on the boundary; it then follows from the maximum principle that the barrier function performs as a lower (resp. upper) bound of the solution. On the other hand, combining the local Harnack inequality obtained in [Golse et al. 2019] with the construction of a Harnack chain yields Lemma 4.6. We remark that the idea of the Harnack chain was first used in [Moser 1964], and an example of its application to Kolmogorov equations can be found in [Anceschi et al. 2019]. Essentially, the spreading of positivity can be seen as a lower bound estimate of the fundamental solution, which is thus related to the result in [Henderson et al. 2019], where the authors applied a probabilistic method.

A subtle point of the lower bound result lies in the possibilities of the degeneracy of solutions as $t \to 0^+$ or $t \to \infty$, which leads to two delicate issues. First, with the same difficulty as mentioned in [Henderson et al. 2020a], in order to prove the uniqueness of the Cauchy problem (1-1), the nondegeneracy of diffusion up to the initial time is required so that the a priori estimates can be still applicable. We remark that, generally speaking, deriving uniqueness of solutions to nonlinear equations in rough spaces is always a classical difficulty, and the presence of a vacuum sometimes gives rise to nonuniqueness phenomenon even for the limiting equation (1-3); see for instance [Daskalopoulos and Kenig 2007]. Under the additional assumptions of Hölder continuity and absence of vacuum on the initial data, we achieve the uniqueness by using the scaling argument and Grönwall's lemma, since the Hölder estimate around the initial time implies that the integrand in the inequality of Grönwall's type is improved to be integrable with respect to the time variable; see the proof of Proposition 4.11 for more details. Second, we are only able to show the convergence to equilibrium if the drift-diffusion coefficient ρ_f^β decays slower than t^{-1} as $t \to \infty$ in Proposition 5.1. Therefore, an additional lower Maxwellian bound on the initial data is imposed in Theorem 1.1(ii) and Theorem 1.4(ii) to ensure the solutions will be away from the vacuum uniformly along time. It would be expected that such additional lower bound assumption could be removed, especially when β is small.

1B2. *Long time behavior.* The drift-diffusion operator $\mathscr{L}_{\mathrm{FP}}$ acts only on the velocity variable and ceases to be dissipative on its unique steady state μ, which also ensures that the null space of $\mathscr{L}_{\mathrm{FP}}$ is spanned by μ and the conservation law of mass is satisfied. Consequently, the convergence to equilibrium is to be expected. With the help of the global smoothness a priori estimates, we are able to pass from the exponential convergence to equilibrium in the L^2-framework to the uniform convergence in $\mathcal{C}^\infty$ in Section 5A, when the spatial domain is compact — that is the periodic box $\mathbb{T}^d$. Therein, the L^2-convergence is obtained by the L^2-hypocoercivity under a macro-micro (fluid-kinetic) decomposition scheme, which suggests the construction of some proper entropy (Lyapunov) functional that would provide an equivalent L^2-norm for solutions. The key ingredient is to control the macroscopic part by means of the microscopic part in view of the decomposition. This hypocoercive theory was studied in [Esposito et al. 2013; Dolbeault et al. 2015; Hérau 2018] via different approaches, while their ideas are essentially the same. In [Esposito et al. 2013], the authors intended to develop the nonlinear energy estimate in an L^2-to-L^∞ framework. In [Dolbeault et al. 2015] and [Hérau 2018], the authors studied

the L^2-hypocoercivity theory in an abstract setting and in the framework of pseudodifferential calculus, respectively. In addition, if the spatial domain is $\mathbb{R}^d$ — meaning that it is not confined to a compact region — then the convergence rate slows down to an algebraic decay, for which the hypocoercive theory was captured in [Bouin et al. 2020]. We remark that the L^2-framework allows us to avoid some difficulties from the nonlinearity of the operator $\rho_f^\beta \mathscr{L}_{\mathrm{FP}} f$, in contrast with H^1-entropic hypocoercivity methods (see for instance [Villani 2009]).

1B3. *Diffusion asymptotics.* The diffusion approximation serves as a simplification of collisional kinetic equations when the mean-free path is much smaller than the typical length of observation in a long time scale. This approximation for linear Fokker–Planck models can be traced back to [Degond and Mas-Gallic 1987], where the authors applied the Hilbert expansion method. One is also able to achieve the diffusion limit for (1-2) in some weak sense by applying a similar strategy to the one given in [El Ghani and Masmoudi 2010]. However, weak convergence is sometimes not appropriate for application, as a precise description of the convergence is not given. Still, the nonlinearity of the term $\rho_{f_\epsilon}^\beta \mathscr{L}_{\mathrm{FP}} f_\epsilon$ in (1-2) associated with nonperturbative initial data reveals some difficulties when deriving a quantitative convergence.

In order to overcome this difficulty, in Section 5B we will rely on the phi-entropy of solutions relative to their limit to see the finite-time asymptotics on the torus. The relative entropy method, which heavily relies on the regularity of solutions to the target equation, has become an effective tool in the study of hydrodynamic limits since [Bardos et al. 1993; Yau 1991] (see also [Saint-Raymond 2009]). The method applied to the diffusion asymptotics of the kinetic Fokker–Planck equation of the type with linear diffusion can be found in [Markou 2017]. The so-called phi-entropy (relative to the global equilibrium) was used to study the convergence of certain kinds of Fokker–Planck equations; see for instance [Arnold et al. 2001; Dolbeault and Li 2018]. Finally, combining the barrier function method with a careful treatment of the regularity estimate of the target equation enables us to deal with the asymptotic dynamics for the cases associated with general Hölder continuous initial data.

1C. *Physical motivation.* The spatially inhomogeneous Fokker–Planck equation (1-1) arises from modeling the evolution of some system of a large number of interacting particles from the statistical mechanical point of view. These models appear for instance in the study of plasma physics and biological dynamics; see [Chavanis 2008; Villani 2002]. Its solution can be interpreted as the probability density of the particles lying at the position x at time t with velocity v. The scaled model (1-2) for small ϵ describes the evolution of the particle density in the small mean-free path and long-time regime, where the nondimensional parameter $\epsilon \in (0, 1)$ designates the ratio between the mean-free path (microscopic scale) and the typical macroscopic length. The limiting equation (1-3) characterizes its macroscopic dynamics.

From the perspective of a stochastic process $\{(X_t, V_t) : t \geq 0\}$ driven by a Brownian motion $\{\mathcal{B}_t : t \geq 0\}$

$$\begin{cases} \mathrm{d}X_t = V_t \,\mathrm{d}t, \\ \mathrm{d}V_t = \rho_f^\beta(t, X_t) V_t \,\mathrm{d}t + \sqrt{2\rho_f^\beta(t, X_t)}\,\mathrm{d}\mathcal{B}_t, \end{cases}$$

the dual equation describing the dynamics of $\{(X_t, V_t) : t \geq 0\}$ is given by (1-1); see the review paper [Chandrasekhar 1943]. Indeed, the nonlinear term $\rho_f^\beta \mathscr{L}_{\mathrm{FP}} f$ models the collisional interaction of the

particles, where the mobility of these particles is hampered by their aggregation. More precisely, the nonlinear dependence on the drift-diffusion coefficient ρ_f^β translates the fact that the effect of friction in the interaction is positively correlated to the local mass of particles occupying the position x at time t. Moreover, the low field scaling $t \mapsto \epsilon^2 t,\ x \mapsto \epsilon x$ of (1-1) formally implies (1-2). As ϵ tends to zero, its spatial diffusion phenomena are characterized by (1-3).

Regarding its physical interpretation, we point out that the factor multiplying the time derivative in (1-2) takes into account the long time scale. The inverse of the factor multiplying $\rho_{f_\epsilon}^\beta\, \mathscr{L}_{\mathrm{FP}}\, f_\epsilon$ stands for the scaled average distance traveled by particles between each collision, and it is usually referred to as mean-free path. In the small mean-free path regime, it was noticed in [Chandrasekhar 1943] that the spatial variation occurs significantly only under the long time scale that is consistent with the particle motion. In such an overdamped process, also called a low field limit or diffusion limit, the statistics of the particle motion translates into the macroscopic behavior of the particle system.

Finally, we recall that the associated phi-entropy introduced in Definition 1.3 is also known as Tsallis entropy in the physics community, which generalizes the Boltzmann–Gibbs entropy (the phi-entropy with $\beta = 1$) in nonextensive statistical mechanics [Tsallis 1988]. It gives some hints for the formulation of the correlated diffusion, where the index β measures the degree of nonextensivity and nonlocality of the system; see [Tsallis 2009].

1D. ***Organization of the paper.*** The article is organized as follows. In Section 2, we recall some basic notions related to kinetic Hölder spaces that are adapted to the Fokker–Planck equations. Section 3 is devoted to the study of the linear Fokker–Planck equation with Hölder continuous coefficients. The well-posedness result Theorem 1.1(i) is proved in Section 4. The asymptotic behaviors, including Theorem 1.1(ii) and Theorem 1.4, are proved in Section 5.

2. Preliminaries

This section is devoted to basic notation, including the invariant structure and the kinetic Hölder space for the equations we are concerned with. Instead of the usual parabolic scaling and translations, the invariant scaling and transformation associated with the Kolmogorov operator $\mathscr{L}_1$ (see (1-4)) is replaced by kinetic scaling and Galilean transformations, respectively. It then turns out that the appropriate Hölder space as well as its norm should be adapted to the new scaling and transformation.

2A. ***The geometry associated to Kolmogorov operators.*** Let $z := (t, x, v) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d$. We define the *kinetic scaling*

$$S_r(t, x, v) := (r^2 t, r^3 x, r v) \quad \text{for } r > 0$$

and the *Galilean transformation*

$$(t_0, x_0, v_0) \circ (t, x, v) := (t_0 + t,\ x_0 + x + t v_0,\ v_0 + v) \quad \text{for } (t_0, x_0, v_0) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d.$$

With respect to the product $\circ$, we are able to define the inverse of z as $z^{-1} := (-t, -x + t v, -v)$. In view of this structure of scaling and transformation, it is natural to define the cylinder centered at the origin of

radius $r > 0$ as

$$Q_r := (-r^2, 0] \times B_{r^3}(0) \times B_r(0).$$

More generally, the cylinder centered at $z_0 = (t_0, x_0, v_0)$ with radius r is defined by

$$Q_r(z_0) := \{z_0 \circ S_r(z) : z \in Q_1\} = \{(t, x, v) : t_0 - r^2 < t \le t_0,\ |x - x_0 - (t - t_0)v_0| < r^3,\ |v - v_0| < r\}.$$

Roughly speaking, for fixed $z_0 \in \mathbb{R}^{1+2d}$, the Kolmogorov operator $\mathscr{L}_1$ is invariant under the kinetic scaling and left-invariant under the Galilean transformation. It means that if f is a solution to the equation $\mathscr{L}_1 f = 0$ in $Q_r(z_0)$, then $f(z_0 \circ S_r(\cdot))$ solves an equation of the same ellipticity class in Q_1.

In addition, the associated quasinorm $\|\cdot\|$ is defined by

$$\|z\| := \max\{|t|^{1/2}, |x|^{1/3}, |v|\},$$

and we notice that $\|S_r(z)\| = r\|z\|$ and $\|z_0 \circ z\| \le 3(\|z_0\| + \|z\|)$. For further information on the non-Euclidean geometry associated to Kolmogorov operators, one may refer to [Anceschi and Polidoro 2020; Imbert and Silvestre 2021].

2B. ***Kinetic Hölder spaces and differential operators.*** The proper kinetic Hölder space and kinetic degree of basic differential operators should be adapted to the above definitions such that they are homogeneous under these transformations. Their definitions were introduced in [Imbert and Silvestre 2021] (see also [Imbert and Mouhot 2021]), and we recall them below.

Given a monomial $m(t, x, v) = t^{k_0} x_1^{k_1} \cdots x_d^{k_d} v_1^{k_{d+1}} \cdots v_d^{k_{2d}}$, we define its *kinetic degree*

$$\deg_{\mathrm{kin}}(m) = 2k_0 + 3\sum_{j=1}^{d} k_j + \sum_{j=d+1}^{2d} k_j.$$

Any polynomial $p \in \mathbb{R}[t, x, v]$ can be uniquely written as a linear combination of monomials, and its kinetic degree $\deg_{\mathrm{kin}}(p)$ is defined by the maximal kinetic degree of the monomials appearing in p. This definition is justified by the fact that $p(S_r(z)) = r^{\deg_{\mathrm{kin}}(p)} p(z)$.

Definition 2.1. Let the constant $\alpha > 0$ and the open subset $\Omega \subset \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d$ be given. We say a function $f : \Omega \to \mathbb{R}$ is $\mathcal{C}_l^\alpha$-continuous at a point $z_0 \in \Omega$ if there exists a polynomial $p_0 \in \mathbb{R}[t, x, v]$ with $\deg_{\mathrm{kin}}(p_0) < \alpha$ and a constant $C > 0$ such that, for any $z \in \Omega$ with $z_0 \circ z \in \Omega$,

$$|f(z_0 \circ z) - p_0(z)| \le C\|z\|^\alpha. \tag{2-1}$$

If this property holds for any $z_0,\ z$ on each compact subset of Ω, then we say $f \in \mathcal{C}_l^\alpha(\Omega)$. If the constant C in (2-1) is uniformly bounded for any $z_0,\ z \in \Omega$, we define the smallest value of C as the seminorm $[f]_{\mathcal{C}_l^\alpha(\Omega)}$ and the norm $\|f\|_{\mathcal{C}_l^\alpha(\Omega)} := [f]_{\mathcal{C}_l^0(\Omega)} + [f]_{\mathcal{C}_l^\alpha(\Omega)}$, where we additionally define $\mathcal{C}_l^0(\Omega) := \mathcal{C}^0(\Omega)$, the space of continuous functions on Ω, with the norm $\|f\|_{\mathcal{C}_l^0(\Omega)} := [f]_{\mathcal{C}_l^0(\Omega)} := \|f\|_{\mathcal{C}^0(\Omega)} = \|f\|_{L^\infty(\Omega)}$.

Remark 2.2. For $\alpha \in [0, 1)$, this $\mathcal{C}_l^\alpha$-continuity is equivalent to the standard definition of $\mathcal{C}^\alpha$-continuity with respect to the distance $\|\cdot\|$. The subscript "l" of $\mathcal{C}_l$ stems from the definition of Hölder continuity above, which is given in terms of a left-invariant distance with respect to the group structure of $\circ$.

We also mention another kind of Hölder space suitable for the study of Kolmogorov operators that was first used in [Manfredini 1997].

Definition 2.3. Let $\alpha \in [0, 1)$ and $\Omega \subset \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d$ be given. The space $\mathcal{C}^{2+\alpha}_{\text{kin}}(\Omega)$ consists of functions $f \in \mathcal{C}^0_l(\Omega)$ such that $D^2_v f$, $(\partial_t + v \cdot \nabla_x) f \in \mathcal{C}^\alpha_l(\Omega)$, and is equipped with the norm

$$\|f\|_{\mathcal{C}^{2+\alpha}_{\text{kin}}(\Omega)} := \|f\|_{\mathcal{C}^0_l(\Omega)} + \|D^2_v f\|_{\mathcal{C}^\alpha_l(\Omega)} + \|(\partial_t + v \cdot \nabla_x) f\|_{\mathcal{C}^\alpha_l(\Omega)}.$$

The consistency between these two definitions is given by [Imbert and Silvestre 2021, Lemma 2.7] (see also [Imbert and Mouhot 2021, Lemma 2.4]), a result that we state here.

Lemma 2.4. *Let $\alpha \in (0, 1)$ and $f \in \mathcal{C}^{2+\alpha}_l(Q_2)$. Then there exists some constant $C > 0$ depending only on the dimension d such that*

$$\|\nabla_v f\|_{\mathcal{C}^\alpha_l(Q_1)} \le C\|f\|_{\mathcal{C}^{1+\alpha}_l(Q_2)} \quad \textit{and} \quad \|D^2_v f\|_{\mathcal{C}^\alpha_l(Q_1)} + \|(\partial_t + v \cdot \nabla_x) f\|_{\mathcal{C}^\alpha_l(Q_1)} \le C\|f\|_{\mathcal{C}^{2+\alpha}_l(Q_2)}.$$

Remark 2.5. For $\alpha > 2$, one can easily check that the polynomial p_0 in (2-1) has the form

$$p_0(t, x, v) = f(z_0) + (\partial_t + v_0 \cdot \nabla_x) f(z_0)t + \nabla_v f(z_0) \cdot v + \tfrac{1}{2} D^2_v f(z_0) v \cdot v + \cdots .$$

In particular, if $\alpha \in (2, 3)$, the polynomial expansion is independent of the x-variable.

Remark 2.6. A subtle difference between $\mathcal{C}^2_l$ and $\mathcal{C}^2_{\text{kin}}$ comes from the fact that, for $f \in \mathcal{C}^2_l$, we have $D^2_v f$ and $(\partial_t + v \cdot \nabla_x) f$ lying in L^∞ rather than $\mathcal{C}^0$.

We will also employ the following notions of weighted Hölder norms in Section 3.

Definition 2.7. Let $z = (t, x, v) \in \Omega := (0, T] \times \mathbb{R}^d \times \mathbb{R}^d$ with $T \in \mathbb{R}_+$. For $f \in \mathcal{C}^\alpha_l(\Omega)$ with $\alpha > 0$ and $\sigma \in \mathbb{R}$, we define

$$[f]^{(\sigma)}_0 := \sup_{z \in \Omega} \iota^\sigma [f]_{\mathcal{C}^0_l(Q_\iota(z))}, \quad [f]^{(\sigma)}_\alpha := \sup_{z \in \Omega} \iota^{\alpha+\sigma} [f]_{\mathcal{C}^\alpha_l(Q_\iota(z))}, \quad \|f\|^{(\sigma)}_\alpha := [f]^{(\sigma)}_0 + [f]^{(\sigma)}_\alpha,$$

where $\iota := \min\{1, t^{1/2}\}$ measures the distance between z and the (parabolic) boundary of Ω.

2C. *Other notation.* Throughout the article, B_R denotes the Euclidean ball in $\mathbb{R}^d$ centered at the origin with radius $R > 0$. We employ the Japanese bracket defined as $\langle \cdot \rangle := (1 + |\cdot|^2)^{1/2}$. By abuse of notation, $\langle \cdot \rangle$ will also denote the velocity mean in Section 5.

Moreover, we assume $0 < \lambda < \Lambda$. We denote by C a *universal* constant — that is to say a constant depending only on β, d, λ, Λ, α, σ, α_0 specified in context. Finally, we write $X \lesssim Y$ to mean that $X \le CY$ for some universal constant $C > 0$, and $X \lesssim_q Y$ to mean that $X \le C_q Y$ for some $C_q > 0$ depending only on universal constants and the quantity q.

3. Kolmogorov–Fokker–Planck equation

This section is devoted to the study of the Cauchy problem associated to the operator (1-4),

$$\begin{cases} \mathscr{L}_1 f := (\partial_t + v \cdot \nabla_x) f - \operatorname{tr}(A D^2_v f) - B \cdot \nabla_v f = s & \text{in } (0, T] \times \mathbb{R}^d \times \mathbb{R}^d, \\ f(0, x, v) = f_{\text{in}}(x, v) & \text{in } \mathbb{R}^d \times \mathbb{R}^d, \end{cases} \tag{3-1}$$

where the $d \times d$ symmetric matrix $A(t, x, v)$ and the d-dimensional vector $B(t, x, v)$ satisfy the condition

$$\begin{cases} A\xi \cdot \xi \geq \lambda |\xi|^2 & \text{for any } \xi \in \mathbb{R}^d, \\ \|A\|_{\mathcal{C}_l^\alpha} + \|B\|_{\mathcal{C}_l^\alpha} \leq \Lambda, \end{cases} \tag{3-2}$$

where $\alpha \in (0, 1)$ and the norm $\|\cdot\|_{\mathcal{C}_l^\alpha(\Omega)}$ of matrix denotes the summation of the norm of each entry. The boundedness condition at infinity means that the solution shall be bounded, which is intended for the validity of maximum principle; see the proof of Lemma A.1 below.

The aim of this section is to solve the Cauchy problem (3-1) by virtue of the weighted Hölder norm (Definition 2.7) and by means of the standard continuity method combined with Schauder-type estimates. One may refer to [Gilbarg and Trudinger 2001, Subsection 6.5] for the corresponding treatment in classical elliptic theory.

Throughout this section we work with the domain $\Omega := (0, T] \times \mathbb{R}^d \times \mathbb{R}^d$, with $T \in \mathbb{R}_+$. We shed light on the fact that all of the results below can be restricted to $(0, T] \times \mathbb{T}^d \times \mathbb{R}^d$ whenever required.

3A. ***Schauder estimates.*** In order to apply the continuity method, first of all one needs to prove a global a priori estimate for solutions to (3-1) with respect to the weighted Hölder norm. In the kinetic setting, we have at our disposal the interior Schauder estimates proved in [Imbert and Mouhot 2021, Theorem 3.9].

Proposition 3.1 (interior Schauder estimate). *Let the constant $\alpha \in (0, 1)$ be given and the cylinder $Q_{2r}(z_0)$ be a subset of Ω with $r \in (0, 1]$. If f satisfies (1-4) with condition (3-2) in $Q_{2r}(z_0)$ and $s \in \mathcal{C}_l^\alpha(Q_{2r}(z_0))$, then we have*

$$r^{2+\alpha}[f]_{\mathcal{C}_l^{2+\alpha}(Q_r(z_0))} \lesssim \|f\|_{L^\infty(Q_{2r}(z_0))} + r^{2+\alpha}[s]_{\mathcal{C}_l^\alpha(Q_{2r}(z_0))}. \tag{3-3}$$

In particular, the right-hand side controls $r^2\|(\partial_t + v \cdot \nabla_x) f\|_{L^\infty(Q_r(z_0))} + r^2\|D_v^2 f\|_{L^\infty(Q_r(z_0))}$.

First of all, we enhance this result to a global estimate for the Cauchy problem (3-1) under a vanishing condition for the initial data.

Proposition 3.2 (global Schauder estimate). *Let $\Omega = (0, T] \times \mathbb{R}^d \times \mathbb{R}^d$ and the constants $\alpha \in (0, 1)$ and $\sigma \in (0, 2)$ be universal, $s \in \mathcal{C}_l^\alpha(\Omega)$ such that $\|s\|_\alpha^{(2-\sigma)} < \infty$, and f be a bounded solution to the Cauchy problem (3-1) under condition (3-2) in Ω. If the initial data f_{in} equals 0, then we have*

$$\|f\|_{2+\alpha}^{(-\sigma)} \lesssim \|s\|_\alpha^{(2-\sigma)}.$$

Proof. In view of Proposition 3.1, it suffices to deal with the estimates around the initial time. Without loss of generality, we assume $T \leq 1$.

Let $z_0 = (t_0, x_0, v_0) \in \Omega$ and $2r = t_0^{1/2}$. Applying the interior Schauder estimate (3-3) yields

$$r^{2+\alpha}[f]_{\mathcal{C}_l^{2+\alpha}(Q_r(z_0))} \lesssim \|f\|_{L^\infty(Q_{2r}(z_0))} + r^{2+\alpha}[s]_{\mathcal{C}_l^\alpha(Q_{2r}(z_0))}.$$

It then follows from the arbitrariness of z_0 that, for any $\sigma < 2$ such that $[f]_0^{(-\sigma)} < \infty$,

$$[f]_{2+\alpha}^{(-\sigma)} \lesssim [f]_0^{(-\sigma)} + [s]_\alpha^{(2-\sigma)}. \tag{3-4}$$

With $\sigma \in (0, 2)$, observing that

$$\begin{aligned} \mathscr{L}_1\Big(\frac{2}{\sigma}[s]_0^{(2-\sigma)}t^{\sigma/2} \pm f\Big) &= [s]_0^{(2-\sigma)}t^{\sigma/2-1} \pm s \geq 0 \quad \text{in } \Omega, \\ \frac{2}{\sigma}[s]_0^{(2-\sigma)}t^{\sigma/2} \pm f &= 0 \qquad\qquad\qquad\qquad \text{on } \{t=0\}, \end{aligned}$$

we apply the maximum principle (Lemma A.1) to the function $(2/\sigma)[s]_0^{(2-\sigma)}t^{\sigma/2} \pm f$ to deduce that $\pm t^{-\sigma/2} f \leq (2/\sigma)[s]_0^{(2-\sigma)}$, which means

$$[f]_0^{(-\sigma)} \lesssim [s]_0^{(2-\sigma)}.$$

Combining this estimate with (3-4), we get the desired result. □

3B. ***Cauchy problem for the linear equation.*** The goal of this subsection is to prove the well-posedness of the Cauchy problem (3-1) with Hölder continuous coefficients.

Proposition 3.3. *Let $\Omega = (0, T] \times \mathbb{R}^d \times \mathbb{R}^d$ and the constants $\alpha \in (0, 1)$ and $\sigma \in (0, 2)$ be universal. Assume that*

$$\begin{cases} A\xi \cdot \xi \geq \lambda|\xi|^2 \quad \textit{for any } \xi \in \mathbb{R}^d, \\ \|A\|_{\mathcal{C}_l^\alpha(\Omega)} + \|\langle v\rangle^{-1}B\|_{\mathcal{C}_l^\alpha(\Omega)} \leq \Lambda. \end{cases} \tag{3-5}$$

Then, for any $s \in \mathcal{C}_l^\alpha(\Omega)$ such that $\|s\|_\alpha^{(2-\sigma)} < \infty$ and $f_{\text{in}} \in \mathcal{C}^0(\mathbb{R}^d \times \mathbb{R}^d)$, there exists a unique bounded solution $f \in \mathcal{C}_l^{2+\alpha}(\Omega)$ to the Cauchy problem (3-1).

Remark 3.4. In contrast with (3-2), condition (3-5) is weaker, which allows the coefficients of B to not necessarily be bounded globally. This fact will be applied to the Ornstein–Uhlenbeck operator $\mathscr{L}_{\text{OU}} = (\nabla_v - v) \cdot \nabla_v$ in Section 4B.

The simplest possible setting of (3-1) under condition (3-5) is recovered by choosing $A = I$ and $B = 0$, which turns out to be the classical Kolmogorov operator $\mathscr{L}_0 := \partial_t + v \cdot \nabla_x - \Delta_v$. This operator was first studied in [Kolmogoroff 1934], where its fundamental solution was calculated explicitly as

$$\Gamma(t, x, v) = \begin{cases} \left(\dfrac{\sqrt{3}}{2\pi t^2}\right)^d \exp\left(-\dfrac{3\left|x + \frac{1}{2}tv\right|^2}{t^3} - \dfrac{|v|^2}{4t}\right) & \text{for } t > 0, \\ 0 & \text{for } t \leq 0. \end{cases} \tag{3-6}$$

One can easily see that Γ is smooth outside of its pole (the origin). In fact, in this latter case the following result holds.

Lemma 3.5. *Let $\Omega = (0, T] \times \mathbb{R}^d \times \mathbb{R}^d$ and $\alpha \in (0, 1)$. For any $s \in \mathcal{C}_l^\alpha(\Omega)$ such that $\|s\|_\alpha^{(2-\sigma)} < \infty$, the function*

$$f(t, x, v) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \Gamma((\tau, \xi, \eta)^{-1} \circ (t, x, v))s(\tau, \xi, \eta)\, \mathrm{d}\tau\, \mathrm{d}\xi\, \mathrm{d}\eta \tag{3-7}$$

is the unique bounded solution in $\mathcal{C}_l^{2+\alpha}(\Omega)$ to (3-1) with $\mathscr{L}_1$ replaced by $\mathscr{L}_0$ and $f_{\text{in}} = 0$.

Remark 3.6. When the spatial domain is $\mathbb{T}^d$, one can apply Green's function

$$G(t,x,v) := \sum_{n\in\mathbb{Z}^d} \Gamma(t,x+n,v),$$

which is well defined due to the decay of Γ.

We are now in a position to apply the standard continuity method to derive Proposition 3.3.

Proof of Proposition 3.3. We split the proof into three steps. In the first step, we establish the case for vanishing initial data under the stronger assumption (3-2). We point out that the assumption on the coefficient B can be weakened in the second step. Finally, we deal with general continuous initial data.

Step 1. Assume $f_{\text{in}} = 0$ and condition (3-2) holds. Let the constant $\sigma \in (0,2)$ be fixed and consider the Banach space $\mathcal{Y} := (\mathcal{C}_l^{2+\alpha}(\Omega), \|\cdot\|_{2+\alpha}^{(-\sigma)})$. In particular, every function lying in $\mathcal{Y}$ vanishes at $t=0$.

For $\tau \in [0,1]$, we define the operator $\mathscr{L}_\tau := (1-\tau)\mathscr{L}_0 + \tau\mathscr{L}_1$, which can be written in the form

$$\mathscr{L}_\tau = \partial_t + v\cdot\nabla_x - \operatorname{tr}(A_\tau D_v^2 \cdot) - \tau B\cdot\nabla_v,$$

where its coefficients $A_\tau := (1-\tau)I + \tau A$ and τB still satisfy condition (3-2) (with λ and Λ replaced by $\min\{1,\lambda\}$ and $\max\{1,\Lambda\}$, respectively). For any $w \in \mathcal{Y}$, we have

$$\|\mathscr{L}_\tau w\|_\alpha^{(2-\sigma)} \lesssim (1+\|A_\tau\|_\alpha^{(2)})\|w\|_{2+\alpha}^{(-\sigma)} + \|B\|_\alpha^{(2)}\|w\|_{1+\alpha}^{(-\sigma)} \lesssim \|w\|_{2+\alpha}^{(-\sigma)}. \tag{3-8}$$

Let the set $\mathcal{I}$ be the collection of $\tau \in [0,1]$ such that the Cauchy problem (3-9) is solvable for any $s \in \mathcal{C}_l^\alpha(\Omega)$ with $\|s\|_\alpha^{(2-\sigma)} < \infty$: there is a unique bounded solution $f \in \mathcal{Y}$ satisfying

$$\begin{cases} \mathscr{L}_\tau f = s & \text{in } \Omega, \\ f(0,x,v) = 0 & \text{in } \mathbb{R}^d\times\mathbb{R}^d. \end{cases} \tag{3-9}$$

By Lemma 3.5, we see that $0 \in \mathcal{I}$; in particular, $\mathcal{I}$ is not empty.

It now suffices to show that $1 \in \mathcal{I}$. Pick $\tau_0 \in \mathcal{I}$. Then the global Schauder estimate provided by Proposition 3.2 implies that, for any $s \in \mathcal{C}_l^\alpha(\Omega)$ with $\|s\|_\alpha^{(2-\sigma)} < \infty$, $f = \mathscr{L}_{\tau_0}^{-1} s$ satisfies

$$\|\mathscr{L}_{\tau_0}^{-1} s\|_{2+\alpha}^{(-\sigma)} \lesssim \|s\|_\alpha^{(2-\sigma)}. \tag{3-10}$$

For any $w \in \mathcal{Y}$, since $\tau_0 \in \mathcal{I}$ and (3-8) holds, the following Cauchy problem is solvable for any $s \in \mathcal{C}_l^\alpha(\Omega)$ with $\|s\|_\alpha^{(2-\sigma)} < \infty$:

$$\begin{cases} \mathscr{L}_{\tau_0} f = s + (\tau-\tau_0)(\mathscr{L}_0 - \mathscr{L}_1)w & \text{in } \Omega, \\ f(0,x,v) = 0 & \text{in } \mathbb{R}^d\times\mathbb{R}^d. \end{cases}$$

Thus, we can define the mapping $F : \mathcal{Y} \to \mathcal{Y}$ by setting $F(w) = f$. Armed with (3-10) and (3-8), there exists a universal constant $C > 0$ such that, for any $u, w \in \mathcal{Y}$,

$$\|F(u) - F(w)\|_{2+\alpha}^{(-\sigma)} \le C|\tau-\tau_0|\|(\mathscr{L}_0 - \mathscr{L}_1)(u-w)\|_\alpha^{(2-\sigma)} \le C|\tau-\tau_0|\|u-w\|_{2+\alpha}^{(-\sigma)}.$$

Hence F is a contraction mapping, provided that $|\tau-\tau_0| \le \delta := (2C)^{-1}$. Then, F gives a unique fixed point $f \in \mathcal{Y}$, which is the unique bounded solution to the Cauchy problem (3-9) in $\mathcal{Y}$. By dividing the interval $[0,1]$ into subintervals of length less than δ, we conclude that $1 \in \mathcal{I}$.

Step 2. If $f_{\text{in}} = 0$ and condition (3-5) holds, we approximate the coefficient B by $B_n := B\varrho_n$, where $\varrho_n(v) := \varrho_1(v/n)$ for $v \in \mathbb{R}^d$, $n \in \mathbb{N}_+$, and $\varrho_1 \in \mathcal{C}_c^\infty(B_2)$ is valued in $[0, 1]$ such that $\varrho_1 \equiv 1$ in B_1. Then, for each $n \in \mathbb{N}_+$, the result obtained in the previous step provides a bounded solution f_n to (3-1) with B replaced by B_n. Indeed, applying the maximum principle (Lemma A.1) to the function $\pm f - e^t \sup_\Omega |s|$ implies $\sup_\Omega |f_n| \le e^T \sup_\Omega |s|$. Thanks to the interior Schauder estimate (Proposition 3.1), for any compact subset $K \subset \Omega$, we have that $\{f_n\}_{n \ge N}$ is precompact in $\mathcal{C}^2_{\text{kin}}(K)$, provided that N (depending on K) is large enough. Sending $n \to \infty$ in the equation satisfied by f_n yields that the limit function $f \in \mathcal{C}_l^{2+\alpha}(\Omega)$ is a bounded solution to (3-1), which satisfies $\sup_\Omega |f| \le e^T \sup_\Omega |s|$.

Step 3. For general $f_{\text{in}} \in \mathcal{C}^0(\mathbb{R}^d \times \mathbb{R}^d)$, we approximate f_{in} uniformly as $\varepsilon \to 0$ by a sequence of smooth functions $\{f_{\text{in}}^\varepsilon\}$ on $\mathbb{R}^d \times \mathbb{R}^d$. Thus, the function $f - f_{\text{in}}^\varepsilon$ is a solution to (3-1), with the source term equal to $s - \mathscr{L}_1 f_{\text{in}}^\varepsilon$, and associated with the vanishing initial data. The procedure presented in the previous steps ensures a unique bounded solution f^ε to (3-1) for each $f_{\text{in}}^\varepsilon$.

The uniform convergence of $\{f_{\text{in}}^\varepsilon\}$ and the maximum principle (Lemma A.1) implies the uniform convergence of $\{f^\varepsilon\}$. We may denote its limit by $f \in \mathcal{C}^0(\overline{\Omega})$, which satisfies $f(0, x, v) = f_{\text{in}}(x, v)$ on $\mathbb{R}^d \times \mathbb{R}^d$. The interior Schauder estimate again implies that $\{f^\varepsilon\}$ is precompact in $\mathcal{C}^2_{\text{kin}}(K)$ for any compact subset $K \subset \Omega$; then sending $\varepsilon \to 0$ gives the solution $f \in \mathcal{C}_l^{2+\alpha}(\Omega)$ to (3-1). Its uniqueness is again given by the maximum principle. This concludes the proof. □

4. Well-posedness of the nonlinear model

This section is devoted to the proof of Theorem 1.1(i), including a self-generating lower bound given in Section 4A, the existence and uniqueness given in Section 4B, and a smoothness a priori estimate given in Section 4C.

First, we recast the Cauchy problem (1-1) in terms of $g(t, x, v) := \mu(v)^{-1/2} f(t, x, v)$, an unknown function, with $g_{\text{in}}(x, v) := \mu(v)^{-1/2} f_{\text{in}}(x, v)$ as follows:

$$\begin{cases} (\partial_t + v \cdot \nabla_x) g = \mathcal{R}[g]\mathcal{U}[g], \\ g(0, x, v) = g_{\text{in}}(x, v), \end{cases} \tag{4-1}$$

where $\mathcal{R}[g]$ and $\mathcal{U}[g]$ on the right-hand side are defined by

$$\mathcal{R}[g] := \left(\int_{\mathbb{R}^d} g\mu^{1/2}\,\mathrm{d}v\right)^\beta \quad \text{and} \quad \mathcal{U}[g] := \mu^{-1/2}\nabla_v(\mu\nabla_v(\mu^{-1/2}g)) = \Delta_v g + \left(\tfrac{1}{2}d - \tfrac{1}{4}|v|^2\right)g.$$

The main advantage of this formulation is that it allows us to get rid of the first-order term in v, and the zeroth-order term is bounded, since g is bounded from above by a Maxwellian.

For convenience, we are also concerned with the substitution $h(t, x, v) := \mu(v)^{-1} f(t, x, v)$ and the Ornstein–Uhlenbeck operator $\mathscr{L}_{\text{OU}} := (\nabla_v - v) \cdot \nabla_v$. Equation (1-1) is then equivalent to

$$(\partial_t + v \cdot \nabla_x) h(t, x, v) = \mathscr{R}_h(t, x)\, \mathscr{L}_{\text{OU}}\, h(t, x, v), \quad \mathscr{R}_h(t, x) := \left(\int_{\mathbb{R}^d} h(t, x, v)\,\mathrm{d}\mu\right)^\beta. \tag{4-2}$$

In contrast with (1-1), the zeroth-order term disappears. Let us begin by exhibiting the global bounds of solutions to (4-2) in $(0, T) \times \mathbb{T}^d \times \mathbb{R}^d$, which is a variant of [Imbert and Mouhot 2021, Lemma 4.1].

Lemma 4.1 (global bounds). *Let $a(t,x) \in L^\infty((0,T)\times\mathbb{T}^d)$ be nonnegative. Assume that, in the sense of distributions, $h(t,x,v) \in L^\infty((0,T); H^1(\mathbb{T}^d\times\mathbb{R}^d, \mathrm{d}m))$ satisfies $(\partial_t + v\cdot\nabla_x)h = a\,\mathscr{L}_{\mathrm{OU}}\,h$ in $(0,T)\times\mathbb{T}^d\times\mathbb{R}^d$. If $h(0,\cdot,\cdot)\le\Lambda$ in $\mathbb{T}^d\times\mathbb{R}^d$, then $h\le\Lambda$ in $(0,T)\times\mathbb{T}^d\times\mathbb{R}^d$; if $h(0,\cdot,\cdot)\ge\lambda$ in $\mathbb{T}^d\times\mathbb{R}^d$, then $h\ge\lambda$ in $(0,T)\times\mathbb{T}^d\times\mathbb{R}^d$.*

Proof. Integrating the equation $(\partial_t + v\cdot\nabla_x)(h-\Lambda) = a\,\mathscr{L}_{\mathrm{OU}}(h-\Lambda)$ against $(h-\Lambda)_+$ yields

$$\frac{1}{2}\int_{\mathbb{T}^d\times\mathbb{R}^d}[(h(t,\cdot,\cdot)-\Lambda)_+^2-(h(0,\cdot,\cdot)-\Lambda)_+^2]\,\mathrm{d}m = -\int_{[0,t]\times\mathbb{T}^d\times\mathbb{R}^d} a|\nabla_v(h-\Lambda)_+|^2\,\mathrm{d}t\,\mathrm{d}m \le 0$$

for any $t\in[0,T]$. This means that the upper bound is preserved along time. Similarly, the lower bound can be obtained by integrating the equation $(\partial_t + v\cdot\nabla_x)(\lambda-h) = a\,\mathscr{L}_{\mathrm{OU}}(\lambda-h)$ against $(\lambda-h)_+$. □

In particular, the above result preserving global bounds holds for solutions to (4-2) and (5-1) in $(0,T)\times\mathbb{T}^d\times\mathbb{R}^d$. We will also apply such result to the substitution $g=\mu^{1/2}h$ appearing in Section 4B. Unless otherwise specified, throughout this section we set the domain $\Omega := (0,T]\times\mathbb{T}^d\times\mathbb{R}^d$ with $T\in\mathbb{R}_+$. Nevertheless, as specified in Remark 4.4, Corollary 4.10, and Proposition 4.11 below, the results of this section also hold if the spatial domain is $\mathbb{R}^d$.

4A. ***Self-generating lower bound.*** Throughout this subsection, we assume that the bounded solution h of (1-1) lies below the universal constant Λ, which is guaranteed by Lemma 4.1 if the initial data lies below Λ. The aim of this subsection is to show the following positivity-spreading result. We remark that this proposition only relies on the mixing structure of the classical parabolic-type maximum principle and the transport operator, but not on the structure of the mass conservation.

Proposition 4.2 (lower bound). *Let $\delta>0$, $\underline{T}\in(0,T)$, and h be a bounded solution to (4-2) in Ω satisfying*

$$h(0,x,v)\ge\delta\mathbb{1}_{\{|x-x_0|<r,|v-v_0|<r\}} \tag{4-3}$$

for some $(x_0,v_0)\in\mathbb{T}^d\times\mathbb{R}^d$. Then, there exist two positive continuous functions $\eta_1(t)$ and $\eta_2(t)$ on $(0,T]$ depending only on universal constants, T, δ, r, and v_0 such that, for any $(t,x,v)\in\Omega$,

$$h(t,x,v)\ge\eta_1(t)e^{-\eta_2(t)|v|^2}. \tag{4-4}$$

Remark 4.3. In particular, the functions $\eta_1(t)$ and $\eta_2(t)$ are positive and bounded on any compact subset of $(0,T]$, but η_1 might degenerate to zero and η_2 may go to infinity as t tends to zero or infinity.

Remark 4.4. If one is concerned with the problem in the whole space — that is $\Omega=(0,T]\times\mathbb{R}^d\times\mathbb{R}^d$ — we can proceed along the same lines as the proof in Appendix B to see that (4-3) implies the lower bound

$$h(t,x,v)\ge\eta_1(t,x)^{-1}e^{-\eta_2(t,x)|v|^4}, \tag{4-5}$$

where the functions $\eta_1(t,x)$ and $\eta_2(t,x)$ on $(0,T]\times\mathbb{R}^d$ are positive, continuous and only depend on universal constants, T, δ, r, and v_0. Compared with (4-4), $\eta_1(t,x)$ and $\eta_2(t,x)$ lose the uniformity in x as $\mathbb{R}^d$ is not compact (see Step 3 of the proof of the proposition in Appendix B). In addition, the exponential tail with respect to v cannot be improved to a Gaussian type, since there is no uniform-in-x lower bound on the local mass $\int h\,\mathrm{d}\mu$ such that Step 4 in Appendix B fails.

We note that the proof of the proposition is composed mainly of two lemmas. On the one hand, Lemma 4.5 extends the lower bounds forward a short time from a neighborhood of any given point in $\mathbb{T}^d \times \mathbb{R}^d$ and at any given time. On the other hand, Lemma 4.6 is used to spread the lower bound for all velocities. The spreading of the lower bound in space is given by selecting the proper velocity to transport the positivity which is guaranteed by Lemma 4.5. By applying these lemmas repeatedly, as proposed in [Henderson et al. 2020b], we are able to control the solution from below for any finite time. We postpone the full proof of Proposition 4.2, obtained by combining these two lemmas, until Appendix B.

Lemma 4.5 (lower bound forward in time). *Let $\delta, \tau, r \in (0, 1]$ and h be a bounded solution to (4-2) in Ω with*

$$h(0, x, v) \geq \delta \mathbb{1}_{\{|x-x_0|<r, |v-v_0|<r/\tau\}}$$

for some $(x_0, v_0) \in \mathbb{T}^d \times \mathbb{R}^d$. Then there exists some universal constant $c_0 > 0$ such that

$$h \geq \tfrac{1}{8}\delta \mathbb{1}_{\mathcal{P}}, \quad \mathcal{P} := \big\{t \leq \min\{T,\ \tau,\ c_0\langle \tau r^{-1}\rangle^{-2}\langle v_0\rangle^{-2}\},\ |x - x_0 - tv| < \tfrac{1}{2}r,\ |v - v_0| < \tfrac{1}{2}r\tau^{-1}\big\}.$$

Proof. Let us consider the barrier function

$$\underline{h}(t, x, v) := -C_0 t + \tfrac{1}{2}\delta(1 - r^{-2}|x - x_0 - tv|^2 - \tau^2 r^{-2}|v - v_0|^2),$$

with the constant $C_0 > 0$ to be determined. The region $\mathcal{Q} := \{t \leq \min\{T, \tau\}, |x-x_0-tv|^2+\tau^2|v-v_0|^2 < r^2\}$ contains $\mathcal{P}$. A direct computation yields

$$|\mathscr{L}_{\mathrm{OU}}\, \underline{h}| \leq |\Delta_v \underline{h}| + |v \cdot \nabla_v \underline{h}| \lesssim \delta\langle \tau r^{-1}\rangle^2 \langle v_0\rangle^2 \quad \text{in } \mathcal{Q}.$$

By choosing $C_0 := (1/(8c_0))\delta\langle \tau r^{-1}\rangle^2\langle v_0\rangle^2$ for some (small) universal constant $c_0 > 0$, we have

$$(\partial_t + v \cdot \nabla_x - \mathscr{R}_h \mathscr{L}_{\mathrm{OU}})\underline{h} \leq -C_0 + \Lambda^\beta |\mathscr{L}_{\mathrm{OU}}\, \underline{h}| < 0 \quad \text{in } \mathcal{Q}. \tag{4-6}$$

In addition, $\underline{h}(t, x, v) \geq \frac{1}{8}\delta$ in $\big\{t \leq c_0\langle \tau r^{-1}\rangle^{-2}\langle v_0\rangle^{-2}, |x - x_0 - tv|^2 + \tau^2|v - v_0|^2 < \frac{1}{2}r^2\big\} \supset \mathcal{P}$.

Applying the classical maximum principle to $\underline{h} - h$ in $\mathcal{Q}$ after observing that $\underline{h} - h \leq 0$ on the parabolic boundary $\{t = 0\} \cap \mathcal{Q}$ and $\{t \leq \min\{T, \tau\},\ |x - x_0 - tv|^2 + \tau^2|v - v_0|^2 = r^2\}$ yields the result. □

The spreading of lower bound to all velocities relies on the construction of a Harnack chain through iterative application of the local Harnack inequality [Golse et al. 2019, Theorem 1.6]. Although some coefficients of (4-2) are unbounded globally over $v \in \mathbb{R}^d$, we remark that their local boundedness is sufficient for us to achieve the result through a careful study on the rescaling during the construction of the Harnack chain.

Lemma 4.6 (lower bound for all velocities). *Let $\delta > 0$, $T, R \in (0, 1]$, $\underline{T} \in (0, T)$, and h be a bounded solution to (4-2) in Ω such that, for any $t \in [0, T]$,*

$$h(t, x, v) \geq \delta \mathbb{1}_{\{|x-x_0-tv_0|<R, |v-v_0|<R\}} \tag{4-7}$$

for some $(x_0, v_0) \in \mathbb{T}^d \times \mathbb{R}^d$. Then there exists some (large) constant $\underline{C} > 0$ depending only on universal constants, $\underline{T}$, δ, R, and v_0 such that, for any $t \in [\underline{T}, T]$, we have

$$h(t, x, v) \geq \underline{C}^{-1} e^{-\underline{C}|v|^4} \mathbb{1}_{\{|x-x_0-tv_0|<R/2\}}. \tag{4-8}$$

Proof. For any $z := (t, x, v) \in \left\{t \in [\underline{T}, T], |x - x_0 - tv_0| < \frac{1}{2}R, v \in \mathbb{R}^d\right\}$, we will construct a finite sequence of points to reach z from the region $\{t \le T, |x - x_0 - tv_0| < R, |v - v_0| < R\}$, where the solution is positive by assumption. In particular, x does not exit this region. The nonlocality of the coefficient $\mathscr{R}_h$, with assumption (4-7), implies the nondegeneracy of the diffusion in velocity so that the positivity of the solution h propagates over $v \in \mathbb{R}^d$ in a localized space region.

Step 1. Iterate the Harnack inequality. For $i \in \{1, 2, \dots, N+1\}$ with $N \in \mathbb{N}$, we define $z_{N+1} := z$ and $z_i := (t_i, x_i, v_i)$ by the relation

$$z_i = z_{i+1} \circ S_r\left(-\tau_1,\ 0,\ -\tau_2 \frac{v - v_0}{|v - v_0|}\right),$$

where the parameters $N, r, \tau_1, \tau_2 > 0$ will be determined next. Consider the function for $\tilde{z} := (\tilde{t}, \tilde{x}, \tilde{v}) \in Q_1$:

$$h_i(\tilde{z}) := h(z_i \circ S_r(\tilde{z})) = h(t_i + r^2\tilde{t}, x_i + r^3\tilde{x} + r^2\tilde{t}v_i, v_i + r\tilde{v}).$$

We observe that, if the following is true for any $\tilde{z} \in Q_1$:

$$t_{i+1} + r^2\tilde{t} \in [0, T], \quad Nr\tau_2 \le |v - v_0|, \tag{4-9}$$

$$|x_{i+1} + r^3\tilde{x} + r^2\tilde{t}v_{i+1} - x_0 - (t_{i+1} + r^2\tilde{t})v_0| < R, \tag{4-10}$$

then, for $1 \le i \le N$, the function $h_{i+1}(\tilde{z})$ satisfies the equation

$$(\partial_{\tilde{t}} + \tilde{v} \cdot \nabla_{\tilde{x}})h_i = \mathscr{R}_h(z_i \circ S_r(\tilde{z}))(\Delta_{\tilde{v}}h_i - r(v_i + r\tilde{v}) \cdot \nabla_{\tilde{v}}h_i) \quad \text{in } Q_1,$$

where the coefficients satisfy

$$\delta^\beta R^{d\beta} \lesssim \mathscr{R}_h \lesssim 1 \quad \text{and} \quad |r(v_i + r\tilde{v})| \le r(1 + |v_0| + |v - v_0|) \le 1,$$

provided that $r \le (1 + |v_0| + |v - v_0|)^{-1}$. Applying the Harnack inequality [Golse et al. 2019, Theorem 1.6] to h_i implies that there exist constants $c_0, \tau_1 \in (0, 1)$ depending only on universal constants, δ, and R such that, for any $\tau_2 \in [0, 1 - \tau_1]$ and $1 \le i \le N$, we have

$$h(z_{i+1}) = h_{i+1}(0, 0, 0) \ge c_0 h_{i+1}\left(-\tau_1, 0, -\tau_2 \frac{v - v_0}{|v - v_0|}\right) = c_0 h(z_i). \tag{4-11}$$

Hence it remains to determine the chain $\{z_i\}_{1 \le i \le N+1}$ such that conditions (4-9) and (4-10) hold and the point z_1 stays in the region $\{(t, x, v) : t \le T, |x - x_0 - tv_0| < R, |v - v_0| < R\}$.

Step 2. Determine the Harnack chain (including N, r, and τ_2) from a proper starting time t_1. For $M > 0$, we set

$$t_1 := \max\left\{\tfrac{1}{2}\underline{T}, t - \tfrac{1}{8}R(1 + |v_0| + |v - v_0|)^{-1}\right\} \quad \text{and} \quad r := \frac{R}{M}(1 + |v_0| + |v - v_0|)^{-2}.$$

Recalling that $T, R \in (0, 1]$, by choosing

$$M \ge \frac{2}{\underline{T}} + \frac{\tau_1}{1 - \tau_1}\left(8 + \frac{2}{\underline{T}}\right),$$

we have

$$r^2 \le \tfrac{1}{2}\underline{T} \quad \text{and} \quad \tau_2 := \frac{r\tau_1|v - v_0|}{t - t_1} \le 1 - \tau_1.$$

To determine the parameter $M > 0$, we point out that there exists some constant $\bar{C}$ depending only on universal constants, $\underline{T}$, δ, R, and v_0 such that $M \le \bar{C}$ and

$$N := \frac{t - t_1}{r^2 \tau_1} \in \mathbb{N}^+.$$

Thus, $N r \tau_2 = |v - v_0|$. This setting then guarantees condition (4-9).

It also follows from the iteration relation that $v_1 = v_0$, and, for $1 \le i \le N+1$,

$$t_i = t_1 + (i-1) r^2 \tau_1, \quad v_i = v_0 + (i-1) r \tau_2 \frac{v - v_0}{|v - v_0|}, \quad x_i = x - r^2 \tau_1 \sum_{j=i}^{N} v_{j+1}. \tag{4-12}$$

Step 3. Determine the starting point x_1. For any $1 \le i \le N$, we estimate the departure distance from the expression (4-12)

$$|x_{i+1} - x_1 - (t_{i+1} - t_1) v_0| = \tfrac{1}{2} i(i+1) r^3 \tau_1 \tau_2 \le N^2 r^3 \tau_1 \tau_2 = (t - t_1)|v - v_0| \le \tfrac{1}{8} R.$$

Therefore, for any $x \in B_{R/2}(x_0 + t v_0)$, there exists some $x_1 \in B_{5R/8}(x_0 + t_1 v_0)$ such that $x_{N+1} = x$. In this setting, for any $1 \le i \le N$, we also have

$$\begin{aligned} &|x_{i+1} + r^3 \tilde{x} + r^2 \tilde{t} v_{i+1} - x_0 - (t_{i+1} + r^2 \tilde{t}) v_0| \\ &\qquad \le |x_{i+1} - x_1 - (t_{i+1} - t_1) v_0| + |x_1 - x_0 - t_1 v_0| + r^2 |r \tilde{x} + \tilde{t} v_{i+1} - \tilde{t} v_0| \\ &\qquad \le \frac{R}{8} + \frac{5R}{8} + r^2 (1 + |v - v_0|) < \frac{3R}{4} + \frac{R^2}{M^2} < R. \end{aligned}$$

Thus, condition (4-10) ensures the inequality (4-11) is satisfied for $1 \le i \le N$, which yields

$$h(t, x, v) \ge c_0^N h(t_1, x_1, v_0) \ge \delta e^{-N \log 1/c_0}.$$

Recalling that $c_0 \in (0, 1)$ appears in (4-11) and $N \le T \bar{C}^2 (1 + |v_0| + |v - v_0|)^4 / (\tau_1 R^2)$, we obtain the desired result (4-8). □

4B. ***Existence and uniqueness.*** Let us begin by summarizing some basic a priori estimates for solutions to (4-1).

Lemma 4.7 (Hölder estimates). *Let $\Omega_x = \mathbb{T}^d$ or $\mathbb{R}^d$, and let g be a solution to (4-1) in $[0, T] \times \Omega_x \times \mathbb{R}^d$ satisfying*

$$\mathcal{R}[g] \ge \lambda \quad \text{in } [0, T] \times \Omega_x \qquad \text{and} \qquad 0 \le g_{\text{in}} \le \Lambda \mu^{1/2} \quad \text{in } \Omega_x \times \mathbb{R}^d.$$

(i) *Let $\underline{T} \in (0, T)$ and $\delta \in \left(0, \frac{1}{2}\right)$. There exists some universal constant $\alpha \in (0, 1)$ such that, for any $Q_{2r}(z_0) \subset [\underline{T}, T] \times \Omega_x \times \mathbb{R}^d$, we have*

$$\|g\|_{\mathcal{C}_l^{2+\alpha}(Q_r(z_0))} \lesssim_{\underline{T}, \delta} \mu^{\delta}(v_0). \tag{4-13}$$

(ii) *If $g_{\text{in}} \in \mathcal{C}^{\alpha_0}(\Omega_x \times \mathbb{R}^d)$ with (universal) $\alpha_0 \in (0, 1)$, then, for any $\delta \in \left(0, \frac{1}{2}\right)$, there exists some universal constant $\alpha \in (0, 1)$ such that*

$$\|g\|_{\mathcal{C}_l^{\alpha}([0,T] \times \Omega_x \times B_1(v_0))} \lesssim_{\delta} (1 + [g_{\text{in}}]_{\mathcal{C}^{\alpha_0}(\Omega_x \times \mathbb{R}^d)}) \mu^{\delta}(v_0).$$

We remark that, armed with Lemma 4.1, the assertions (i) and (ii) in the above lemma directly follow from [Imbert and Mouhot 2021, Proposition 4.4] and [Zhu 2021, Corollary 4.6], respectively.

Proposition 4.8 (existence). *For any $g_{\text{in}} \in \mathcal{C}^0(\mathbb{T}^d \times \mathbb{R}^d)$ such that $0 \le g_{\text{in}} \le \Lambda\mu^{1/2}$ in $\mathbb{T}^d \times \mathbb{R}^d$, there exists a (classical) solution g to the Cauchy problem (4-1) satisfying $0 \le g \le \Lambda\mu^{1/2}$ in Ω.*

Remark 4.9. For any given nonnegative continuous function g_{in} that is not identically zero, there is some point $(x_0, v_0) \in \mathbb{T}^d \times \mathbb{R}^d$ and some constants $\delta, r > 0$ such that

$$g_{\text{in}}(x, v) \ge \delta \mathbb{1}_{\{|x-x_0|<r, |v-v_0|<r\}} \quad \text{in } \mathbb{T}^d \times \mathbb{R}^d. \tag{4-14}$$

We will see that the upper bound $g_{\text{in}} \le \Lambda\mu^{1/2}$ and the lower bound (4-14) assumptions on the initial data g_{in} (which could be discontinuous) are sufficient to ensure the existence of a solution $g \in \mathcal{C}^2_{\text{kin}}(\Omega)$ in the weak sense such that, for any $\phi \in \mathcal{C}_c^\infty([0, T) \times \mathbb{T}^d \times \mathbb{R}^d)$,

$$\int_{\mathbb{T}^d \times \mathbb{R}^d} g_{\text{in}}\phi|_{t=0} = \int_\Omega \big\{-g(\partial_t + v \cdot \nabla_x)\phi + \mathcal{R}[g]\nabla_v g \cdot \nabla_v \phi - \mathcal{R}[g]\big(\tfrac{1}{2}d - \tfrac{1}{4}|v|^2\big)g\phi\big\}. \tag{4-15}$$

As solutions become regular instantaneously, the difference between the weak solution and the classical one lies only in the continuity around the initial time.

Proof. Let us assume that g_{in} satisfies (4-14) for some point $(x_0, v_0) \in \mathbb{T}^d \times \mathbb{R}^d$ and some constants $\delta, r > 0$. By Proposition 4.2, for any solution g to (4-1) and for any $\underline{T} \in (0, T)$, there is some $\lambda_* > 0$ depending only on universal constants, $\underline{T}$, T, δ, r, and v_0 such that

$$\mathcal{R}[g](t, x) \ge \lambda_* \quad \text{in } [\underline{T}, T] \times \mathbb{T}^d. \tag{4-16}$$

<u>Step 1</u>. We first approximate the initial data g_{in} by $g^\epsilon_{\text{in}} := g_{\text{in}} * \varrho_\varepsilon + \varepsilon\mu^{1/2}$, where $\varepsilon \in (0, 1]$, $\varrho_\varepsilon(x, v) := \varepsilon^{-2d}\varrho_1(x/\varepsilon, v/\varepsilon)$ with $(x, v) \in \mathbb{T}^d \times \mathbb{R}^d$, and $\varrho_1 \in \mathcal{C}_c^\infty(B_1 \times B_1)$ is a nonnegative bump function such that $\int_{\mathbb{R}^{2d}} \varrho_1 = 1$. Then

$$\varepsilon\mu^{1/2} \le g^\varepsilon_{\text{in}} \le (1 + \Lambda)\mu^{1/2} \quad \text{in } T^d \times \mathbb{R}^d.$$

Let us fix $\varepsilon \in (0, 1]$. In order to establish the existence of a solution to (4-1) associated with the initial data $g^\varepsilon_{\text{in}}$, we find a fixed point of the mapping $F : w \mapsto g$ defined by solving the Cauchy problem

$$\begin{cases} (\partial_t + v \cdot \nabla_x)g = \mathcal{R}[w]\mathcal{U}[g] & \text{in } \Omega, \\ g(0, \cdot, \cdot) = g^\varepsilon_{\text{in}} & \text{in } \mathbb{T}^d \times \mathbb{R}^d \end{cases} \tag{4-17}$$

on the closed convex subset $\mathcal{K}$ of the Banach space $\mathcal{C}^\gamma_l(\overline{\Omega})$,

$$\mathcal{K} := \{w \in \mathcal{C}^\gamma_l(\overline{\Omega}) : \|w\|_{\mathcal{C}^\gamma_l(\Omega)} \le \mathcal{N},\ \varepsilon\mu^{1/2} \le w \le (1 + \Lambda)\mu^{1/2} \text{ in } \overline{\Omega}\},$$

where the constants $\gamma \in (0, 1)$ and $\mathcal{N} > 0$ are to be determined. We remark that (4-17) is equivalent to

$$(\partial_t + v \cdot \nabla_x)(\mu^{-1/2}g) = \mathcal{R}[w]\,\mathscr{L}_{\text{OU}}(\mu^{-1/2}g).$$

By Lemma 4.1 and the fact that $\mathcal{R}[w] \ge \varepsilon$, we have $\varepsilon\mu^{1/2} \le g \le (1 + \Lambda)\mu^{1/2}$ in $\overline{\Omega}$. In particular, for any $w \in \mathcal{K}$, we have the following for the lower-order term: $\big|\mathcal{R}[w]\big(\frac{1}{2}d - \frac{1}{4}|v|^2\big)g\big| \lesssim 1$. Thus, the global Hölder estimate [Zhu 2021, Corollary 4.6] implies that there exist some constants $\gamma \in (0, 1)$ and $\mathcal{N} > 0$ depending

only on universal constants and ε such that $\|g\|_{\mathcal{C}_l^{2\gamma}(\Omega)} \le \mathcal{N}$, which also implies that the lower-order term $\big|\mathcal{R}[w]\big(\frac{1}{2}d - \frac{1}{4}|v|^2\big)g\big|$ is bounded in $\mathcal{C}_l^{2\gamma}(\Omega)$. It then follows from Proposition 3.3 with the interior Schauder estimate (Proposition 3.1) that the mapping $F : \mathcal{K} \to \mathcal{K} \cap \mathcal{C}_l^{2\gamma}(\overline{\Omega}) \cap \mathcal{C}_l^{2+2\gamma}(\Omega)$ is well defined. In addition, with the help of the Arzelà–Ascoli theorem, we know that $F(\mathcal{K})$ is precompact in $\mathcal{C}_l^{\gamma}(\overline{\Omega})$.

As far as the continuity of F is concerned, we take a sequence $\{w_n\}$ converging to w_∞ in $\mathcal{C}_l^{\gamma}(\overline{\Omega})$. Since $\{F(w_n)\}$ is precompact in $\mathcal{C}_l^{\gamma}(\overline{\Omega})$, there exists a converging subsequence whose limit is $g_\infty \in \mathcal{C}_l^{\gamma}(\overline{\Omega})$ which satisfies $g_\infty(0, \cdot, \cdot) = g_{\text{in}}^{\varepsilon}$ in $\mathbb{T}^d \times \mathbb{R}^d$. In view of the interior Schauder estimate (Proposition 3.1), $\{F(w_n)\}$ is precompact in $\mathcal{C}_{\text{kin}}^2(K)$ for any compact subset $K \subset \Omega$ and $g_\infty \in \mathcal{C}_{\text{kin}}^2(\Omega) \cap \mathcal{C}^0(\overline{\Omega})$. Sending $n \to \infty$ in (4-17) satisfied by $(w, g) = (w_n, F(w_n))$, we see that (4-17) also holds for the pair of limits $(w, g) = (w_\infty, g_\infty)$. Then, applying the maximum principle (Lemma A.1) to

$$\begin{cases} (\partial_t + v \cdot \nabla_x)(\mu^{-1/2}(g_\infty - F(w_\infty))) = \mathcal{R}[w_\infty]\,\mathscr{L}_{\text{OU}}(\mu^{-1/2}(g_\infty - F(w_\infty))) & \text{in } \Omega, \\ (g_\infty - F(w_\infty))(0, \cdot, \cdot) = 0 & \text{in } \mathbb{T}^d \times \mathbb{R}^d, \end{cases}$$

we arrive at $g_\infty = F(w_\infty)$.

Then, for every $\varepsilon \in (0, 1]$, we are allowed to apply the Schauder fixed-point theorem (see for instance [Gilbarg and Trudinger 2001, Corollary 11.2]) to get $g^\varepsilon \in \mathcal{C}_{\text{kin}}^2(\Omega) \cap \mathcal{C}^0(\overline{\Omega})$ such that $F(g^\varepsilon) = g^\varepsilon$, which is a (classical) solution to (4-1) associated with the initial data $g_{\text{in}}^{\varepsilon}$.

Step 2. Passage to the limit. Recalling the lower bound (4-16) on the coefficient and the higher-order Hölder estimate given by Lemma 4.7(i), for any $\underline{T} \in (0, T)$, we point out that $\{g^\varepsilon\}$ is uniformly bounded in $\mathcal{C}_l^{2+\alpha_*}([\underline{T}, T] \times \mathbb{T}^d \times \mathbb{R}^d)$ for some constant $\alpha_* \in (0, 1)$ with the same dependence as λ_*. Hence g^ε converges uniformly to g in $\mathcal{C}_{\text{kin}}^2([\underline{T}, T] \times \mathbb{T}^d \times \mathbb{R}^d)$, up to a subsequence.

Write the equation satisfied by g^ε in the weak formulation: that is, for any $\phi \in \mathcal{C}_c^\infty(\overline{\Omega})$,

$$\begin{aligned} &\int_{\mathbb{T}^d \times \mathbb{R}^d} [g^\varepsilon(T, x, v)\phi(T, x, v) - g_{\text{in}}^{\varepsilon}(x, v)\phi(0, x, v)] \\ &\qquad = \int_\Omega \big\{ g^\varepsilon(\partial_t + v \cdot \nabla_x)\phi - \mathcal{R}[g^\varepsilon]\nabla_v g^\varepsilon \cdot \nabla_v \phi + \mathcal{R}[g^\varepsilon]\big(\tfrac{1}{2}d - \tfrac{1}{4}|v|^2\big) g^\varepsilon \phi \big\}. \end{aligned} \tag{4-18}$$

Combining the energy estimate derived by choosing $\phi = g^\varepsilon$ above with the upper bound of g^ε provided by Lemma 4.1, we have

$$\int_\Omega |\mathcal{R}[g^\varepsilon]\nabla_v g^\varepsilon|^2 \lesssim \int_\Omega \mathcal{R}[g^\varepsilon]|\nabla_v g^\varepsilon|^2 \le \frac{1}{2}\int_{\mathbb{T}^d \times \mathbb{R}^d} |g_{\text{in}}^{\varepsilon}|^2 + \int_\Omega \mathcal{R}[g^\varepsilon]\big(\tfrac{1}{2}d - \tfrac{1}{4}|v|^2\big)|g^\varepsilon|^2 \lesssim 1.$$

Therefore, after passing to a subsequence, $\mathcal{R}[g^\varepsilon]\nabla_v g^\varepsilon$ converges weakly in $L^2(\Omega)$. On account of its local uniform convergence, we know that its weak limit is $\mathcal{R}[g]\nabla_v g$. In addition, since $\mu^{-1/2}g^\varepsilon$ is uniformly bounded, by their local uniform convergence, we can also derive that the sequences g^ε and $\mathcal{R}[g^\varepsilon]\big(\frac{1}{2}d - \frac{1}{4}|v|^2\big)g^\varepsilon$ converge to g and $\mathcal{R}[g]\big(\frac{1}{2}d - \frac{1}{4}|v|^2\big)g$, respectively, weakly in $L^2(\Omega)$, up to a subsequence. Then, for any $\phi \in \mathcal{C}_c^\infty([0, T) \times \mathbb{T}^d \times \mathbb{R}^d)$, sending $\varepsilon \to 0$ in (4-18) gives (4-15).

Furthermore, if the initial data g_{in} is continuous, then the barrier function method shows that the continuity around the initial time depends only on the upper bound of the solution and the continuity of g_{in}; see the derivation of the estimate (5-30) of a general type in Section 5B. Indeed, by (5-30)

(with $\epsilon = 1$, $R = |v_1|$, $h_\epsilon = \mu^{-1/2} g$, and $h_{\epsilon,\mathrm{in}} = \mu^{-1/2} g_{\mathrm{in}}$), we see that, for any fixed $\delta \in (0, 1)$ and $(x_1, v_1) \in \mathbb{T}^d \times \mathbb{R}^d$ and for any $(t, x, v) \in [0, \delta/(4(1+|v_1|))] \times B_\delta(x_1, v_1)$,

$$\begin{aligned} |g(t,x,v) - g_{\mathrm{in}}(x_1,v_1)| &\lesssim \delta^{-2}\langle v_1\rangle^2 \mu^{1/2}(v_1) t + \delta^{-2}\mu^{1/2}(v_1)(|x - x_1 - tv|^2 + |v - v_1|^2) \\ &\qquad + \mu^{1/2}(v_1) \sup_{B_\delta(x_1,v_1)} |g_{\mathrm{in}}(x,v) - g_{\mathrm{in}}(x_1,v_1)| \\ &\lesssim \delta^{-2}(t + |x - x_1|^2 + \mu^{1/2}(v_1)|v - v_1|^2) \\ &\qquad + \sup_{B_\delta(x_1,v_1)} |g_{\mathrm{in}}(x,v) - g_{\mathrm{in}}(x_1,v_1)|. \end{aligned} \tag{4-19}$$

This implies the continuity of the solution g around $t = 0$ and finishes the proof. □

One may extend the above existence result to the case where the spacial domain $\mathbb{T}^d$ is replaced by $\mathbb{R}^d$.

Corollary 4.10. *For any $g_{\mathrm{in}} \in \mathcal{C}^0(\mathbb{R}^d \times \mathbb{R}^d)$ such that $0 \le g_{\mathrm{in}} \le \Lambda\mu^{1/2}$ in $\mathbb{R}^d \times \mathbb{R}^d$, there exists a solution g to the Cauchy problem (4-1) satisfying $0 \le g \le \Lambda\mu^{1/2}$ in $(0, T] \times \mathbb{R}^d \times \mathbb{R}^d$. If additionally $h_{\mathrm{in}} = \mu^{-1/2} g_{\mathrm{in}}$ satisfies $h_{\mathrm{in}} \ge \lambda$ and $h_{\mathrm{in}} - M_1 \in L^1(\mathbb{R}^d \times \mathbb{R}^d, \mathrm{d}m)$ for some constant $M_1 > 0$, then $h = \mu^{-1/2} g$ satisfies $h \ge \lambda$ in $(0, T] \times \mathbb{R}^d \times \mathbb{R}^d$ and*

$$\|h - M_1\|_{L_t^\infty([0,T]; L^1(\mathbb{R}^d \times \mathbb{R}^d, \mathrm{d}m))} \le \|h_{\mathrm{in}} - M_1\|_{L^1(\mathbb{R}^d \times \mathbb{R}^d, \mathrm{d}m)}. \tag{4-20}$$

Proof. For $R > 1$, we set $g_{\mathrm{in}}^R = g_{\mathrm{in}} \mathbb{1}_{[-R+R^{-1}, R-R^{-1}]^d}$ for $x \in [-R, R]^d$ with periodic extension to $\mathbb{R}^d$. In the light of Proposition 4.8, we take a solution g^R to (4-1) associated with the initial data g_{in}^R in $(0, T] \times [-R, R]^d \times \mathbb{R}^d$, where $[-R, R]^d$ is considered as a periodic box. After extracting a subsequence, we define the function $g := \lim_{R\to\infty} g^R$ in $(0, T] \times \mathbb{R}^d \times \mathbb{R}^d$ pointwise; furthermore, since $0 \le \mu^{-1/2} g^R \le \Lambda$ in $(0, T] \times [-R, R]^d \times \mathbb{R}^d$, we know that the limiting function satisfies $0 \le \mu^{-1/2} g \le \Lambda$ in $(0, T] \times \mathbb{R}^d \times \mathbb{R}^d$. Similarly, $\mu^{-1/2} g_{\mathrm{in}} \ge \lambda$ in $\mathbb{R}^d \times \mathbb{R}^d$ implies that $\mu^{-1/2} g \ge \lambda$ in $(0, T] \times \mathbb{R}^d \times \mathbb{R}^d$.

Since the initial data is continuous unless it is identically zero, we assume that $g_{\mathrm{in}} \ge \delta \mathbb{1}_{\{|x-x_0|<r, |v-v_0|<r\}}$ for some point $(x_0, v_0) \in \mathbb{R}^d \times \mathbb{R}^d$ and some constants $\delta, r > 0$. Consider $R > |x_0| + r$. Applying the lower bound of the solution given by (4-5) yields that, for any compact subset $K \subset (0, T] \times \mathbb{R}^d \times \mathbb{R}^d$, the coefficient $\mathcal{R}[g^R]$ is greater than or equal to λ_*, where the constant $\lambda_* > 0$ only depends on universal constants, δ, r, v_0, and K. In view of the higher-order Hölder estimate given by Lemma 4.7(i), we know that g^R uniformly converges to g in $\mathcal{C}^2_{\mathrm{kin}}(K)$, up to a subsequence. Additionally, due to the estimate derived in (4-19), the limiting function g is a solution to (4-1) that matches the initial data g_{in} continuously.

As for (4-20), we notice that the function $(h^R - M_1)_\pm$ with $h^R := \mu^{-1/2} g^R$ satisfies

$$(\partial_t + v \cdot \nabla_x)(h^R - M_1)_\pm \le \mathscr{R}_{h^R} \mathscr{L}_{\mathrm{OU}}(h^R - M_1)_\pm \quad \text{in } (0, T] \times [-R, R]^d \times \mathbb{R}^d.$$

Integrating the equation against the function $\int_{[-R,R]^d} (h^R - M_1)_\pm \, \mathrm{d}x$ yields

$$\int_{[-R,R]^d \times \mathbb{R}^d} (h^R(t, \cdot, \cdot) - M_1)_\pm \, \mathrm{d}m - \int_{[-R,R]^d \times \mathbb{R}^d} (h^R(0, \cdot, \cdot) - M_1)_\pm \, \mathrm{d}m \le 0.$$

Sending $R \to \infty$, we acquire

$$\|(h - M_1)_\pm\|_{L_t^\infty([0,T]; L^1(\mathbb{R}^d \times \mathbb{R}^d, \mathrm{d}m))} \le \|(h_{\mathrm{in}} - M_1)_\pm\|_{L^1(\mathbb{R}^d \times \mathbb{R}^d, \mathrm{d}m)},$$

which implies the estimate (4-20) as asserted. □

The following proposition concerned with the uniqueness of the Cauchy problem (4-1) is derived from a Grönwall-type argument. The standard scaling technique and the Hölder estimate up to the initial time given by Lemma 4.7(ii) can improve the integrability with respect to t in the energy estimate so that Grönwall's inequality becomes admissible; see (4-25) for the precise expression. This kind of phenomena was also noticed in [Henderson et al. 2020a] (see the remarks in §1.4.2). The global energy estimate of (4-1) is not available when the spatial domain is unbounded, since there is no decay of the solution as $|x| \to \infty$. To work it out, we take advantage of the idea originated from the uniformly local space used in [Henderson et al. 2019; Kato 1975]. We note that such a technique is not necessary when working with the periodic box $\mathbb{T}^d$.

Proposition 4.11 (uniqueness). *Let the domain Ω_x be $\mathbb{T}^d$ or $\mathbb{R}^d$, the constant $\alpha_0 \in (0, 1)$, and the functions $0 \le g_1, g_2 \lesssim \mu^{1/2}$ be two solutions to (4-1) in $(0, T] \times \Omega_x \times \mathbb{R}^d$ associated with the same initial data $g_{\mathrm{in}} \in \mathcal{C}^{\alpha_0}(\Omega_x \times \mathbb{R}^d)$ such that*

$$\int_{\mathbb{R}^d} g_{\mathrm{in}} \mu^{1/2} \, \mathrm{d}v \ge \lambda \quad \text{in } \Omega_x \qquad \text{and} \qquad 0 \le g_{\mathrm{in}} \lesssim \mu^{1/2} \quad \text{in } \Omega_x \times \mathbb{R}^d.$$

Then $g_1 = g_2$ in $[0, T] \times \Omega_x \times \mathbb{R}^d$.

Proof. In view of the lower bound given by Lemma 4.5 and Proposition 4.2, we know that there is some constant $\lambda_* \in (0, 1)$ depending only on universal constants, T, and the initial data such that

$$\int_{\mathbb{R}^d} g_i \mu^{1/2} \, \mathrm{d}v \ge \lambda_* \quad \text{in } [0, T] \times \Omega_x, \quad i = 1, 2. \tag{4-21}$$

Therefore, we may assume $T = \Lambda^{-1}$ with $\Lambda > 1$. Let us set the difference

$$\tilde{g} := \exp\left(-\tfrac{1}{8}|v|^2 t\right)(g_1 - g_2).$$

We have to show that $\tilde{g}$ is identically zero.

In view of (4-1), a direct computation yields that the function $\tilde{g}$ satisfies

$$\begin{aligned}(\partial_t + v \cdot \nabla_x)\tilde{g} + \tfrac{1}{8}|v|^2 \tilde{g} \\ = \exp\left(-\tfrac{1}{8}|v|^2 t\right)(\mathcal{R}[g_1] - \mathcal{R}[g_2])\mathcal{U}[g_1] + \mathcal{R}[g_2]\left(\mathcal{U}[\tilde{g}] + \tfrac{1}{2} t v \cdot \nabla_v \tilde{g} + \left(\tfrac{1}{4} d t + \tfrac{1}{16}|v|^2 t^2\right)\tilde{g}\right),\end{aligned} \tag{4-22}$$

with the initial condition $\tilde{g}(0, x, v) = 0$ in $\Omega_x \times \mathbb{R}^d$.

Let $y \in \mathbb{R}^d$. We introduce a cut-off function $\phi_y(x) := \phi(x - y)$, where $\phi \in C_c^\infty(\mathbb{R}^d)$ is valued in $[0, 1]$ such that $\phi|_{B_1} \equiv 1$, $\phi|_{B_2^c} \equiv 0$, and $|\nabla \phi| \lesssim 1$ in $\mathbb{R}^d$. For any $t \in (0, T]$, integrating (4-22) against $\phi_y^2 \tilde{g}$ in $\Omega_x \times \mathbb{R}^d$ and applying integration by parts yields

$$\begin{aligned}\frac{1}{2}\int_{\Omega_x \times \mathbb{R}^d} \phi_y^2 \tilde{g}^2(t) = \int_0^t \int_{\Omega_x \times \mathbb{R}^d} \Big\{ (v \cdot \nabla \phi_y)\phi_y \tilde{g}^2 - \tfrac{1}{8}|v|^2 \phi_y^2 \tilde{g}^2 + \exp\left(-\tfrac{1}{8}|v|^2 t\right)(\mathcal{R}[g_1] - \mathcal{R}[g_2])\mathcal{U}[g_1]\phi^2 \tilde{g} \\ - \mathcal{R}[g_2]\left(|\nabla_v(\mu^{-1/2}\tilde{g})|^2 \mu \phi_y^2 + \tfrac{1}{4} d t \phi_y^2 \tilde{g}^2 - \left(\tfrac{1}{4} d t + \tfrac{1}{16}|v|^2 t^2\right)\phi_y^2 \tilde{g}^2\right)\Big\}.\end{aligned}$$

Since $\mathcal{R}[g_2] \in [0, \Lambda]$, for any $t \in (0, T]$, we have

$$\frac{1}{2}\int_{\Omega_x \times \mathbb{R}^d} \phi_y^2 \tilde{g}^2(t) \le \int_0^t \int_{\Omega_x \times \mathbb{R}^d} \left\{ |v|\,|\nabla \phi_y| \phi_y \tilde{g}^2 - \tfrac{1}{16}|v|^2 \phi_y^2 \tilde{g}^2 + \mu^{-1/4}|\mathcal{R}[g_1] - \mathcal{R}[g_2]|\,|\mathcal{U}[g_1]|\phi_y^2 |\tilde{g}| \right\}.$$

Due to the elementary inequality $|z^\beta - 1| \le |z-1|$ (with $z \in \mathbb{R}^+$) and the lower bound estimate in (4-21), as well as the boundedness assumption on g_1 and g_2, we have

$$|\mathcal{R}[g_1] - \mathcal{R}[g_2]| \le \mathcal{R}[g_1]^{(\beta-1)/\beta} |\mathcal{R}[\tilde{g}]|^{1/\beta} \le \frac{1}{\lambda_*} \int_{\mathbb{R}^d} |\tilde{g}(t,x,\cdot)| \mu^{1/2} \lesssim_{\lambda_*} 1 \quad \text{in } [0,1] \times \Omega_x.$$

It then follows that, for any $t \in (0, T]$,

$$\begin{aligned}
\frac{1}{2} \int_{\Omega_x \times \mathbb{R}^d} \phi_y^2 \tilde{g}^2(t) &\le \int_0^t \int_{\Omega_x \times \mathbb{R}^d} \big(|v| |\nabla \phi_y| \phi_y \tilde{g}^2 - \tfrac{1}{16} |v|^2 \phi_y^2 \tilde{g}^2 \big) \\
&\qquad + \frac{1}{\lambda_*} \int_0^t \|\mu^{-3/8} \mathcal{U}[g_1]\|_{L^\infty_{x,v}} \int_{\Omega_x \times \mathbb{R}^d_v} \phi_y^2 |\tilde{g}(t,x,v)| \mu^{1/8} \int_{\mathbb{R}^d_\xi} |\tilde{g}(t,x,\xi)| \mu^{1/2} \, \mathrm{d}\xi \\
&\lesssim_{\lambda_*} \int_0^t \int_{\Omega_x \times \mathbb{R}^d} |\nabla \phi_y|^2 \tilde{g}^2 + \int_0^t \|\mu^{-3/8} \mathcal{U}[g_1]\|_{L^\infty_{x,v}} \int_{\Omega_x \times \mathbb{R}^d} \phi_y^2 \tilde{g}^2, \qquad (4\text{-}23)
\end{aligned}$$

where we used the Cauchy–Schwarz inequality and Hölder's inequality in the last line. Recalling that $\phi_y(x) = \phi(x-y) \in C_c^\infty(\mathbb{R}^d)$ and $|\nabla \phi| \lesssim 1$ in $\mathbb{R}^d$, we have

$$\sup_{y \in \mathbb{R}^d} \int_{\Omega_x \times \mathbb{R}^d} |\nabla \phi_y|^2 \tilde{g}^2 \lesssim \sup_{y \in \mathbb{R}^d} \int_{\Omega_x \times \mathbb{R}^d} \phi_y^2 \tilde{g}^2.$$

By the definition of $\mathcal{U}[g_1]$ and the upper bound $g_1 \lesssim \mu^{1/2}$,

$$\|\mu^{-3/8} \mathcal{U}[g_1]\|_{L^\infty_{x,v}} \lesssim 1 + \|\mu^{-3/8} \Delta_v g_1\|_{L^\infty_{x,v}}.$$

Hence, for any $t \in (0, T]$, taking supremum over $y \in \mathbb{R}^d$ in (4-23), we obtain

$$\sup_{y \in \mathbb{R}^d} \int_{\Omega_x \times \mathbb{R}^d} \phi_y^2 \tilde{g}^2(t) \lesssim_{\lambda_*} \int_0^t (1 + \|\mu^{-3/8} \Delta_v g_1\|_{L^\infty_{x,v}}) \sup_{y \in \mathbb{R}^d} \int_{\Omega_x \times \mathbb{R}^d} \phi_y^2 \tilde{g}^2. \qquad (4\text{-}24)$$

Now we have to consider the pointwise estimate on $D_v^2 g_1$. Let $z_0 = (t_0, x_0, v_0) \in (0, T] \times \Omega_x \times \mathbb{R}^d$ and $2r = t_0^{1/2}$. In view of (4-21), Lemma 4.7(ii) implies that there exists some constant $\alpha_* \in (0,1)$ with the same dependence as λ_* such that

$$\|g_1\|_{C_l^{\alpha_*}([0,T] \times \Omega_x \times B_1(v_0))} \lesssim_{\lambda_*} 1 + [g_{\mathrm{in}}]_{C^{\alpha_0}(\Omega_x \times \mathbb{R}^d)}.$$

Then, applying the interior Schauder estimate (Proposition 3.1) and the upper bound $g_1 \lesssim \mu^{1/2}$ yields

$$\begin{aligned}
\|D_v^2 g_1\|_{L^\infty(Q_r(z_0))} &\lesssim_{\lambda_*} r^{-2} \|g_1 - g_1(z_0)\|_{L^\infty(Q_{2r}(z_0))} + r^{\alpha_*} \big[\mathcal{R}[g_1] \big(\tfrac{1}{2} d - \tfrac{1}{4} |v|^2 \big) g_1 \big]_{C_l^{\alpha_*}(Q_{2r}(z_0))} \\
&\lesssim_{\lambda_*} r^{-2+\alpha_*/4} \mu^{3/8}(v_0) [g_1]^{1/4}_{C_l^{\alpha_*}(Q_{2r}(z_0))} + \mu^{3/8}(v_0) [g_1]^{1/4}_{C_l^{\alpha_*}(Q_{2r}(z_0))} \\
&\lesssim_{\lambda_*} t_0^{-1+\alpha_*/8} \mu^{3/8}(v_0) (1 + [g_{\mathrm{in}}]^{1/4}_{C^{\alpha_0}(\Omega_x \times \mathbb{R}^d)}).
\end{aligned}$$

By the arbitrariness of z_0, we know that, for any $s \in (0, T]$,

$$\|\mu^{-3/8} \Delta_v g_1(s)\|_{L^\infty(\Omega_x \times \mathbb{R}^d)} \lesssim_{\lambda_*} (1 + [g_{\mathrm{in}}]^{1/4}_{C^{\alpha_0}(\Omega_x \times \mathbb{R}^d)}) s^{-1+\alpha_*/8}.$$

Dragging this estimate into (4-24) yields that for any $t \in (0, T]$,

$$\sup_{y\in\mathbb{R}^d} \int_{\Omega_x\times\mathbb{R}^d} \phi_y^2 \tilde{g}^2(t) \le C_* \int_0^t \mathrm{d}s(1+s^{-1+\alpha_*/8}) \sup_{y\in\mathbb{R}^d} \int_{\Omega_x\times\mathbb{R}^d} \phi_y^2 \tilde{g}^2(s), \tag{4-25}$$

where the constant $C_* > 0$ depends only on universal constants and the initial data. The desired result is then given by Grönwall's inequality. □

4C. *Global regularity.* The instantaneous smoothness a priori estimate in Theorem 1.1(i) is made up of the lower bound given by Proposition 4.2 and the following proposition.

Proposition 4.12. *Let $\Omega_x = \mathbb{T}^d$ or $\mathbb{R}^d$, let $\underline{T} \in (0, T)$, and let the function g be a solution to (4-1) in $(0, T) \times \mathbb{T}^d \times \mathbb{R}^d$ such that*

$$\mathcal{R}[g] \ge \lambda \quad \text{in } [\underline{T}/4, T] \times \Omega_x \qquad \text{and} \qquad 0 \le g \le \Lambda\mu^{1/2} \quad \text{in } [0, T] \times \Omega_x \times \mathbb{R}^d. \tag{4-26}$$

Then, for any $\nu \in \left(0, \frac{1}{2}\right)$ and $k \in \mathbb{N}$, we have

$$\|\mu^{-\nu} g\|_{\mathcal{C}^k([\underline{T},T]\times\Omega_x\times\mathbb{R}^d)} \le C_{\underline{T},\nu,k}$$

for some constant $C_{\underline{T},\nu,k} > 0$ depending only on universal constants, $\underline{T}$, ν, and k.

Generally speaking, if g is a solution to (4-1) in $(0, T] \times \Omega_x \times \mathbb{R}^d$ constructed by Proposition 4.8 (with $\Omega_x = \mathbb{T}^d$) or Corollary 4.10 (with $\Omega_x = \mathbb{R}^d$), then the uniform positivity assumption (4-26) should be replaced by

$$\mathcal{R}[g] \ge \lambda_{t,x} \quad \text{in } (0, T] \times \Omega_x \qquad \text{and} \qquad 0 \le g \le \Lambda\mu^{1/2} \quad \text{in } \Omega_x \times \mathbb{R}^d,$$

where $\lambda_{t,x} > 0$ may degenerate to zero as $t \to 0$ or $t + |x| \to \infty$; see Proposition 4.2 and Remark 4.4. As an immediate consequence of the above proposition, for any $\nu \in \left(0, \frac{1}{2}\right)$, $k \in \mathbb{N}$, and for any compact subset $K \subset (0, T] \times \Omega_x$, there exists some constant $C_{\nu,k,K} > 0$ depending only on universal constants, ν, k, and K such that

$$\|\mu^{-\nu} g\|_{\mathcal{C}^k(K\times\mathbb{R}^d)} \le C_{\nu,k,K},$$

which is exactly the assertion in Theorem 1.1(i).

In order to show the higher regularity, we will apply the bootstrap procedure developed in [Imbert and Silvestre 2022] which was intended for the non-cut-off Boltzmann equation. The classical bootstrap iteration proceeds by differentiating the equation, using a priori estimates to the new equation to improve the regularity of solutions, and repeating the procedure. Nevertheless, since $\mathcal{C}_l^{2+\alpha} \not\subset \mathcal{C}_x^1$ for any $\alpha \in (0, 1)$ by their definitions, the hypoelliptic structure of (4-1) does not gain enough regularity in the x-variable which disables the x-differentiation at each iteration. Indeed, the Schauder-type estimate provided by Lemma 4.7(i) only shows that the solution to (4-1) belongs to $\mathcal{C}^{(2+\alpha)/3}$ with respect to the x-variable. In order to overcome it, we have to apply estimates to increments of the solution to recover a full derivative. From now on, for $y \in \mathbb{R}^d$ and $w \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d$, we denote the spatial increment by

$$\delta_y g(z) := g(w \circ (0, y, 0)) - g(w).$$

Let us proceed with the proof of the regularity estimate.

Proof of Proposition 4.12. We are going to show that, for any multi-index $k := (k_t, k_x, k_v) \in \mathbb{N} \times \mathbb{N}^d \times \mathbb{N}^d$ and $\nu \in \left(0, \frac{1}{2}\right)$, there exists some constant $\alpha_k \in (0, 1)$ depending only on $|k|$ such that, for any $Q_r(z_0) \subset \left[\frac{1}{2}\underline{T}, T\right] \times \Omega_x \times \mathbb{R}^d$,

$$\|\partial_t^{k_t} \partial_x^{k_x} \partial_v^{k_v} g\|_{\mathcal{C}_l^{2+\alpha_k}(Q_r(z_0))} \lesssim_{\underline{T},\nu,k} \mu^\nu(v_0). \tag{4-27}$$

For simplicity, we will omit the domain in estimates below, since the estimates can be always localized around the center z_0.

Step 0. The case of $k = (0, 0, 0)$ in (4-27) is a direct consequence of Lemma 4.7(i).

Step 1. We will establish that (4-27) holds for any differential operators of the type $\partial_x^{k_x}$. It suffices to show that, for any $n \in \mathbb{N}$, $k_x \in \mathbb{N}^d$ with $|k_x| = n$, $\nu \in \left(0, \frac{1}{2}\right)$, and $y \in B_{r^3/4}$,

$$\|\delta_y \partial_x^{k_x} g\|_{\mathcal{C}_l^{2+\alpha_n}} \lesssim_{\underline{T},\nu,n} |y| \mu^\nu(v_0). \tag{4-28}$$

Indeed, sending $y \to 0$ in (4-28) will complete this step.

Based on an induction on $|k_x| = n$, we suppose that (4-28) holds for any $|k_x| \le n - 1$, which implies, for any $k_x \in \mathbb{N}^d$ with $|k_x| \le n$,

$$\|\partial_x^{k_x} g\|_{\mathcal{C}_l^{2+\alpha_n}} \lesssim_{\underline{T},\nu,n} \mu^\nu(v_0). \tag{4-29}$$

We remark that the induction here begins with (4-29) for $|k_x| = 0$, which holds due to the previous step.

Let $q := \delta_y \partial_x^{k_x} g$ with $|k_x| = n$. Lemma C.2 and (4-29) gives

$$\|q\|_{\mathcal{C}_l^{\alpha_n}} \lesssim \|\partial_x^{k_x} g\|_{\mathcal{C}_l^{2+\alpha_n}} \|(0, y, 0)\|^2 \lesssim_{\underline{T},\nu,n} |y|^{2/3} \mu^\nu(v_0). \tag{4-30}$$

Therefore, we have to enhance the exponent $\frac{2}{3}$ on the right-hand side to 1; as a sacrifice, the Hölder exponent on the left-hand side will decrease.

Set $\tau_y g(w) := g(w \circ (0, y, 0))$ for $y \in \mathbb{R}^d$ and $w \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d$. A direct computation shows that q satisfies

$$(\partial_t + v \cdot \nabla_x) q = \mathcal{R}[g]\mathcal{U}[q] + \sum_{\substack{|i| \le n \\ i \le k_x}} \delta_y \hat{D}_i \mathcal{R}[g]\mathcal{U}[\tau_y D_i g] + \sum_{\substack{|i| \le n-1 \\ i \le k_x}} \hat{D}_i \mathcal{R}[g]\mathcal{U}[\delta_y D_i g], \tag{4-31}$$

where the multi-indices $i \le k_x$ mean each component of i is less than or equal to the corresponding component of k, and $\hat{D}_i$ denotes the differential operator satisfying $\partial_t^{k_t} D_x^{k_x} = \hat{D}_i \circ D_i$.

In view of (4-29), (4-30), and the induction hypothesis, each term in the summations on the right-hand side of (4-31) is bounded in $\mathcal{C}_l^{\alpha_n}$ by $C_n \|(0, y, 0)\|^2 \mu^{\nu'}(v_0)$ for any $\nu' \in (0, \nu)$. Then, by the interior Schauder estimate (Proposition 3.1),

$$\|q\|_{\mathcal{C}_l^{2+\alpha_n}} \lesssim_{\underline{T},\nu',n} \|(0, y, 0)\|^2 \mu^{\nu'}(v_0). \tag{4-32}$$

Combining Lemma C.1 with (4-30) and (4-32), we obtain (4-28).

Step 2. For the case $k_v = 0$ in (4-27), we proceed with a bidimensional induction on $(m, n) = (k_t, |k_x|)$ such that, for any $\nu \in \left(0, \frac{1}{2}\right)$,

$$\|\partial_t^{k_t} D_x^{k_x} g\|_{\mathcal{C}_l^{2+\alpha_{m,n}}} \lesssim_{\underline{T},\nu,m,n} \mu^\nu(v_0). \tag{4-33}$$

Based on the previous step ($m = 0$), we have to show that (4-33) holds for $k_t = m \geq 1$ and $|k_x| = n$ under the induction hypothesis that (4-33) holds for any $k_t \leq m - 1$ and $|k_x| \leq n + 1$.

With $k_t = m > 0$ and $|k_x| = n$, set $q := \partial_t^{k_t} D_x^{k_x} g$. Then,

$$(\partial_t + v \cdot \nabla_x) q = \mathcal{R}[g]\mathcal{U}[q] + \sum_{\substack{i \leq (k_t, k_x, 0) \\ i \neq (k_t, k_x, 0)}} \hat{D}_i \mathcal{R}[g]\mathcal{U}[D_i g], \tag{4-34}$$

where we use the notation $\hat{D}_i$ for the differential operator satisfying $\partial_t^{k_t} D_x^{k_x} = \hat{D}_i \circ D_i$.

By the induction hypothesis, each term in the remainder (the summation on the right-hand side of (4-34)) with $i \neq (0, 0, 0)$ can be controlled in $\mathcal{C}_l^{\alpha_{m,n}}$. It now suffices to deal with the exceptional term $\partial_t^{k_t} D_x^{k_x} \mathcal{R}[g]$ so that the whole remainder can be controlled in $\mathcal{C}_l^{\alpha_{m,n}}$; then (4-33) follows from the interior Schauder estimate (Proposition 3.1). To this end, using Lemma 2.4 and the induction hypothesis with the pair $(m - 1, n)$ yields

$$\|(\partial_t + v \cdot \nabla_x)\partial_t^{m-1} D_x^{k_x} g\|_{\mathcal{C}_l^{\alpha_{m,n}}} \lesssim_{\underline{T}, \nu, m, n} \mu^{\nu}(v_0). \tag{4-35}$$

Due to the induction hypothesis with the pair $(m - 1, n + 1)$, for any $\nu' \in (0, \nu)$,

$$\mu^{-\nu'}(v_0)\|(v \cdot \nabla_x)\partial_t^{m-1} D_x^{k_x} g\|_{\mathcal{C}_l^{2+\alpha_{m,n}}} \lesssim_{\nu, \nu'} \mu^{-\nu}(v_0)\|\partial_t^{m-1}\nabla_x D_x^{k_x} g\|_{\mathcal{C}_l^{2+\alpha_{m,n}}} \lesssim_{\underline{T}, \nu, m, n} 1. \tag{4-36}$$

Then, (4-35) and (4-36) produce the bound on $\mu^{-\nu'}(v_0)\|q\|_{\mathcal{C}_l^{\alpha_{m,n}}}$.

<u>Step 3</u>. Similarly, to show (4-27) for any differential operator $\partial_t^{k_t} D_x^{k_x} D_v^{k_v}$, we proceed with a bidimensional induction on $(m, n) = (k_t + |k_x|, k_v)$ such that, for any $\nu \in \left(0, \frac{1}{2}\right)$,

$$\|\partial_t^{k_t} D_x^{k_x} D_v^{k_v} g\|_{\mathcal{C}_l^{2+\alpha_{m,n}}} \lesssim_{\underline{T}, \nu, m, n} \mu^{\nu}(v_0). \tag{4-37}$$

The case $n = 0$ is treated in the previous step. By Lemma 2.4 and the induction hypothesis (4-37) with $k_t + |k_x| = m$ and $|k_v| = n - 1,\ n \geq 1$, we have

$$\|\partial_v \partial_t^{k_t} \partial_x^{k_x} \partial_v^{k_v} g\|_{\mathcal{C}_l^{\alpha_{m,n}}} \lesssim \|\partial_t^{k_t} \partial_x^{k_x} \partial_v^{k_v} g\|_{\mathcal{C}_l^{1+\alpha_{m,n}}} \lesssim_{\underline{T}, \nu, m, n} \mu^{\nu}(v_0).$$

Computing the equation satisfied by $\partial_v \partial_t^{k_t} \partial_x^{k_x} \partial_v^{n-1} g$ and proceeding as in the previous step, we conclude the proof. □

5. Diffusion asymptotics

This section is devoted to the study of the global-in-time quantitative diffusion asymptotics which consists of the (uniform-in-ϵ) convergence towards the equilibrium over long times and of the finite-time asymptotics, including the results of Theorem 1.1(ii) and Theorem 1.4.

We first introduce the required notation. For any scalar or vector-valued function $\Psi \in L^1(\mathbb{R}^d, \mathrm{d}\mu)$, we denote its velocity mean by

$$\langle \Psi \rangle := \int_{\mathbb{R}^d} \Psi(v)\, \mathrm{d}\mu.$$

For any pair of functions (scalars, vectors, or $d \times d$ matrices) $\Psi_1, \Psi_2 \in L^2(\mathbb{T}^d \times \mathbb{R}^d, \mathrm{d}m)$, we denote their L^2 inner product with respect to the measure $\mathrm{d}m$ by

$$(\Psi_1, \Psi_2) := \int_{\mathbb{T}^d \times \mathbb{R}^d} \Psi_1(x, v)\Psi_2(x, v)\,\mathrm{d}m,$$

where the multiplication between the pair in the integrand is replaced by the scalar contraction product if Ψ_1 and Ψ_2 are a pair of vectors or matrices.

Recalling our notation for the Ornstein–Uhlenbeck operator $\mathscr{L}_{\mathrm{OU}} = (\nabla_v - v) \cdot \nabla_v$, we apply the substitutions $f_\epsilon = \mu h_\epsilon$ and $f_{\epsilon,\mathrm{in}} = \mu h_{\epsilon,\mathrm{in}}$ in (1-2) and obtain

$$\begin{cases} (\epsilon\partial_t + v \cdot \nabla_x)h_\epsilon(t, x, v) = \frac{1}{\epsilon}\langle h_\epsilon\rangle^\beta(t, x)\,\mathscr{L}_{\mathrm{OU}}\, h_\epsilon(t, x, v), \\ h_\epsilon(0, x, v) = h_{\epsilon,\mathrm{in}}(x, v). \end{cases} \tag{5-1}$$

In this setting, by applying integration by parts, for any $h_1, h_2 \in \mathcal{C}_c^\infty(\mathbb{T}^d \times \mathbb{R}^d)$, we get

$$(h_1, \mathscr{L}_{\mathrm{OU}}\, h_2) = -(\nabla_v h_1, \nabla_v h_2).$$

We will use this identity repeatedly in the computation below. Then, the operator $\mathscr{L}_{\mathrm{OU}}$ is self-adjoint with respect to the inner product $(\,\cdot\,,\cdot\,)$, and the bracket $\langle\,\cdot\,\rangle$ is a projection on the null space of $\mathscr{L}_{\mathrm{OU}}$. Moreover, as the total mass is conserved, we define

$$M_0 := \int_{\mathbb{T}^d \times \mathbb{R}^d} h_\epsilon\,\mathrm{d}m = \int_{\mathbb{T}^d} \langle h_\epsilon\rangle\,\mathrm{d}x. \tag{5-2}$$

Proceeding with the macro-micro (fluid-kinetic) decomposition, we define the orthogonal complement of the projection $\langle\,\cdot\,\rangle$ of h_ϵ as

$$h_\epsilon^\perp(t, x, v) := h_\epsilon(t, x, v) - \langle h_\epsilon\rangle(t, x).$$

In this framework, the local mass $\langle h_\epsilon\rangle$ is the macroscopic (fluid) part and the complement $h_\epsilon^\perp$ is the microscopic (kinetic) part. In addition, taking the bracket $\langle\,\cdot\,\rangle$ after multiplying the equation in (5-1) with 1 and v leads to the macroscopic equations

$$\epsilon\partial_t\langle h_\epsilon\rangle + \nabla_x \cdot \langle v h_\epsilon\rangle = 0, \tag{5-3}$$

$$\epsilon\partial_t\langle v h_\epsilon\rangle + \nabla_x \cdot \langle v^{\otimes 2} h_\epsilon\rangle = -\frac{1}{\epsilon}\langle h_\epsilon\rangle^\beta\langle v h_\epsilon\rangle, \tag{5-4}$$

where $\langle v h_\epsilon\rangle$ and $\langle v^{\otimes 2} h_\epsilon\rangle$ represent the local momentum and the stress tensor, respectively.

5A. ***Long time behavior.*** Our aim is to establish the (uniform-in-ϵ) exponential decay towards the equilibrium M_0 for (5-1). In particular, when $\epsilon = 1$, it sets up the exponential convergence in each order derivative based on the smoothness a priori estimates given in Section 4C.

We note that the classical coercive method is not applicable in our case to obtain the convergence to equilibrium due to the degeneracy of the ellipticity of the spatially inhomogeneous equation. Indeed, the Poincaré inequality only produces a spectral gap on the orthogonal complement of the projection $\langle\,\cdot\,\rangle$; see (5-7). As mentioned in Section 1B, there are several ways to achieve the long-time asymptotics. We mainly follow the argument in [Esposito et al. 2013] (see also [Kim et al. 2020]) in a simpler scenario. It also allows us to see some similarity among [Dolbeault et al. 2015; Esposito et al. 2013; Hérau 2018].

Proposition 5.1. *Let the function $\lambda_t : \mathbb{R}_+ \to [0, \Lambda]$ satisfy $\lambda_t' \le 0$ on $\mathbb{R}_+$. If h_ϵ is a solution to (5-1) in $\mathbb{R}_+ \times \mathbb{T}^d \times \mathbb{R}^d$ associated with the initial data $0 \le h_{\epsilon,\mathrm{in}} \le \Lambda$ and satisfying*

$$\langle h_\epsilon\rangle^\beta(t,x) \ge \lambda_t \quad \text{in } \mathbb{R}_+ \times \mathbb{T}^d \qquad \text{and} \qquad \int_{\mathbb{R}_+} (\lambda_t + \lambda_t')\,\mathrm{d}t = \infty, \tag{5-5}$$

then the solution h_ϵ converges to the state M_0 in $L^2(\mathrm{d}m)$ as $t \to \infty$. More precisely, there exists some universal constant $c > 0$ such that, for any $t > 0$, we have

$$\|h_\epsilon(t,\cdot,\cdot) - M_0\|^2_{L^2(\mathrm{d}m)} \lesssim \|h_{\epsilon,\mathrm{in}} - M_0\|^2_{L^2(\mathrm{d}m)} \exp\biggl(-c\int_0^t (\lambda_s + \lambda_s')\,\mathrm{d}s\biggr). \tag{5-6}$$

Proof. Since the velocity mean of the microscopic part vanishes, $\langle h_\epsilon^\perp\rangle = 0$, using (5-1) and the Poincaré inequality yields

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|h_\epsilon - M_0\|^2_{L^2(\mathrm{d}m)} &= \frac{1}{\epsilon^2}(\langle h_\epsilon\rangle^\beta \mathscr{L}_{\mathrm{OU}}\, h_\epsilon, h_\epsilon - M_0) = -\frac{1}{\epsilon^2}(\langle h_\epsilon\rangle^\beta \nabla_v h_\epsilon^\perp, \nabla_v h_\epsilon^\perp)\\ &\le -\frac{\lambda_t}{\epsilon^2}\|\nabla_v h_\epsilon^\perp\|^2_{L^2(\mathrm{d}m)} \lesssim -\frac{\lambda_t}{\epsilon^2}\|h_\epsilon^\perp\|^2_{L^2(\mathrm{d}m)}.\end{aligned} \tag{5-7}$$

Now we have to recover a new entropy that would give some bound on the projection $\langle h_\epsilon\rangle - M_0$.

For every test function $v\cdot\Psi(t,x)\mu$, with a vector-valued function $\Psi \in H^1_{t,x}(\mathbb{R}^+ \times \mathbb{T}^d, \mathbb{R}^d)$, we write the weak formulation of (5-1) as

$$\frac{\mathrm{d}}{\mathrm{d}t}(v\cdot\Psi, h_\epsilon) = \frac{1}{\epsilon}(v^{\otimes 2} : \nabla_x\Psi, h_\epsilon) + (v\cdot\partial_t\Psi, h_\epsilon) + \frac{1}{\epsilon^2}(\langle h_\epsilon\rangle^\beta \mathscr{L}_{\mathrm{OU}}\, v\cdot\Psi, h_\epsilon).$$

Taking the macro-micro decomposition into account, from the above expression we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(v\cdot\Psi, h_\epsilon^\perp) &= \frac{1}{\epsilon}(|v_1|^2\,\mathrm{tr}(\nabla_x\Psi), \langle h_\epsilon\rangle - M_0) + \frac{1}{\epsilon}(v^{\otimes 2} : \nabla_x\Psi, h_\epsilon^\perp)\\ &\quad + (v\cdot\partial_t\Psi, h_\epsilon^\perp) - \frac{1}{\epsilon^2}(\langle h_\epsilon\rangle^\beta v\cdot\Psi, h_\epsilon^\perp).\end{aligned} \tag{5-8}$$

Let us now introduce an auxiliary function $u(t,x)$: for any fixed $t \in \mathbb{R}_+$, defined $u(t,x)$ as the solution of the following elliptic equation under the compatibility condition (5-2):

$$-\Delta_x u = \langle h_\epsilon\rangle - M_0 \quad \text{in } \mathbb{T}^d, \tag{5-9}$$

whose elliptic estimate states

$$\|\nabla_x u\|_{L^2_x} + \|\nabla_x^2 u\|_{L^2_x} \lesssim \|\langle h_\epsilon\rangle - M_0\|_{L^2_x}. \tag{5-10}$$

In addition, observing that $\langle v h_\epsilon\rangle = \langle v h_\epsilon^\perp\rangle$, from (5-3) we get

$$\epsilon\partial_t\langle h_\epsilon\rangle + \nabla_x\cdot\langle v h_\epsilon^\perp\rangle = 0.$$

Combining this macroscopic relation with (5-9), we have

$$\int_{\mathbb{T}^d} |\nabla_x(\partial_t u)|^2 = \int_{\mathbb{T}^d} \partial_t u\,\partial_t\langle h_\epsilon\rangle = -\frac{1}{\epsilon}\int_{\mathbb{T}^d} \partial_t u\,\nabla_x\cdot\langle v h_\epsilon^\perp\rangle = \frac{1}{\epsilon}\int_{\mathbb{T}^d} \nabla_x(\partial_t u)\cdot\langle v h_\epsilon^\perp\rangle.$$

It then follows from Hölder's inequality that

$$\|\nabla_x(\partial_t u)\|_{L^2_x} \le \frac{1}{\epsilon}\|\langle v h_\epsilon^\perp\rangle\|_{L^2_x} \lesssim \frac{1}{\epsilon}\|h_\epsilon^\perp\|_{L^2(\mathrm{d}m)}. \tag{5-11}$$

Choosing $\Psi = \nabla_x u$ in (5-8) yields

$$-\frac{1}{\epsilon}(|v_1|^2\Delta_x u, \langle h_\epsilon\rangle - M_0) + \frac{\mathrm{d}}{\mathrm{d}t}(v\cdot\nabla_x u, h_\epsilon^\perp) \lesssim \left(\frac{1}{\epsilon}\|\nabla_x^2 u\|_{L^2_x} + \|\nabla_x(\partial_t u)\|_{L^2_x} + \frac{1}{\epsilon^2}\|\nabla_x u\|_{L^2_x}\right)\|h_\epsilon^\perp\|_{L^2(\mathrm{d}m)}.$$

Applying (5-9)–(5-11), we have

$$\frac{1}{\epsilon}\|\langle h_\epsilon\rangle - M_0\|_{L^2_x}^2 + \frac{\mathrm{d}}{\mathrm{d}t}(v\cdot\nabla_x u, h_\epsilon^\perp) \lesssim \frac{1}{\epsilon^2}\|\langle h_\epsilon\rangle - M_0\|_{L^2_x}\|h_\epsilon^\perp\|_{L^2(\mathrm{d}m)} + \frac{1}{\epsilon}\|h_\epsilon^\perp\|_{L^2(\mathrm{d}m)}^2.$$

By the Cauchy–Schwarz inequality, we arrive at

$$\|\langle h_\epsilon\rangle - M_0\|_{L^2_x}^2 + \epsilon\frac{\mathrm{d}}{\mathrm{d}t}(v\cdot\nabla_x u, h_\epsilon^\perp) \lesssim \frac{1}{\epsilon^2}\|h_\epsilon^\perp\|_{L^2(\mathrm{d}m)}^2. \tag{5-12}$$

Then, (5-12) combined with (5-7) implies

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}_\epsilon(t) &\lesssim -\frac{1-\delta}{\epsilon^2}\|h_\epsilon^\perp\|_{L^2(\mathrm{d}m)}^2 - \delta\lambda_t\|\langle h_\epsilon\rangle - M_0\|_{L^2_x}^2 + \delta\epsilon\lambda_t'(v\cdot\nabla_x u, h_\epsilon^\perp)\\ &\le -\delta\lambda_t\|h_\epsilon - M_0\|_{L^2(\mathrm{d}m)}^2 - \delta\lambda_t'|(v\cdot\nabla_x u, h_\epsilon^\perp)|,\end{aligned}$$

where the constant $\delta \in \left(0, \frac{1}{2}\right)$ will be determined and the modified entropy $\mathcal{E}_\epsilon$ is defined by

$$\mathcal{E}_\epsilon(t) := \|h_\epsilon - M_0\|_{L^2(\mathrm{d}m)}^2 + \delta\epsilon\lambda_t(v\cdot\nabla_x u, h_\epsilon^\perp).$$

We note that (5-10) also implies

$$|(v\cdot\nabla_x u, h_\epsilon^\perp)| \lesssim \|\langle h_\epsilon\rangle - M_0\|_{L^2_x}\|h_\epsilon^\perp\|_{L^2(\mathrm{d}m)} \le \|h_\epsilon - M_0\|_{L^2(\mathrm{d}m)}^2. \tag{5-13}$$

It means that the modified entropy $\mathcal{E}_\epsilon$ is equivalent (independent of ϵ) to the square of the $L^2(\mathrm{d}m)$-distance between h_ϵ and M_0, when the constant $\delta > 0$ is sufficiently small.

Hence we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}_\epsilon(t) \lesssim -(\lambda_t + \lambda_t')\mathcal{E}_\epsilon(t).$$

The conclusion (5-6) then follows from Grönwall's inequality and the equivalence between $\mathcal{E}_\epsilon(t)$ and $\|h_\epsilon(t,\cdot,\cdot) - M_0\|_{L^2(\mathrm{d}m)}^2$. □

We pointed out that the elliptic estimate (5-10) for the Poisson equation (5-9) used in the above proof resulting from a Poincaré-type inequality essentially relies on the compactness of the spatial domain. It was shown in [Bouin et al. 2020] that the related elliptic estimate can be recovered by applying the Nash inequality [1958] when the spatial domain is the whole space $\mathbb{R}^d$, whose argument is under an abstract setting. Inspired by the proof of Proposition 5.1 above, we are also able to make the construction of [Bouin et al. 2020] precise to see that the argument still works for the nonlinear equation (5-1). We remark that the following algebraic decay rate is optimal in the sense that it is the same as in the linear case; see Appendix A of [Bouin et al. 2020].

Proposition 5.2. *Assume the initial data $h_{\epsilon,\mathrm{in}}$ is valued in $[\lambda, \Lambda]$ and satisfies $h_{\epsilon,\mathrm{in}} - M_1 \in L^1(\mathbb{R}^{2d}, \mathrm{d}m)$ for some universal constant $M_1 > 0$. Let the function h_ϵ valued in $[\lambda, \Lambda]$ be a solution to (5-1) in $\mathbb{R}_+ \times \mathbb{R}^{2d}$ associated with $h_\epsilon|_{t=0} = h_{\epsilon,\mathrm{in}}$. Then, for any $t > 0$,*

$$\|h_\epsilon - M_1\|_{L^2(\mathbb{R}^{2d},\,\mathrm{d}m)} \lesssim (1 + \|h_{\epsilon,\mathrm{in}} - M_1\|_{L^1(\mathbb{R}^{2d},\,\mathrm{d}m)})t^{-d/4}.$$

Proof. By the same derivation of (5-7) and (5-8) as in the proof of Proposition 5.1, we have the microscopic coercivity

$$\frac{\mathrm{d}}{\mathrm{d}t}\|h_\epsilon - M_1\|^2_{L^2(\mathbb{R}^{2d},\,\mathrm{d}m)} \lesssim -\frac{1}{\epsilon^2}\|h_\epsilon^\perp\|^2_{L^2(\mathbb{R}^{2d},\,\mathrm{d}m)}, \tag{5-14}$$

and the identity from the macro-micro decomposition

$$\begin{aligned}-(|v_1|^2\Delta_x w, \langle h_\epsilon\rangle - M_0)_W + \epsilon\frac{\mathrm{d}}{\mathrm{d}t}(v\cdot\nabla_x w, h_\epsilon^\perp)_W \\ = (v^{\otimes 2} : \nabla_x^2 w, h_\epsilon^\perp)_W + \epsilon(v\cdot\partial_t\nabla_x w, h_\epsilon^\perp)_W - \frac{1}{\epsilon}(\langle h_\epsilon\rangle^\beta v\cdot\nabla_x w, h_\epsilon^\perp)_W,\end{aligned} \tag{5-15}$$

where $(\,\cdot\,,\,\cdot\,)_W$ denotes the $L^2(\mathbb{R}^{2d}, \mathrm{d}m)$ inner product, and the function $w(t,x) \in L_t^\infty([0,T]; L_x^1 \cap L_x^2(\mathbb{R}^d))$ is chosen to be the solution of the following elliptic equation associated with the constant $\Theta := \langle |v_1|^2\rangle$ and the macroscopic source $\langle h_\epsilon\rangle - M_1$:

$$w - \Theta\Delta_x w = \langle h_\epsilon\rangle - M_1 \quad \text{in } \mathbb{R}^d. \tag{5-16}$$

The elliptic estimate is derived by integrating (5-16) against $-\Theta\Delta_x w$, so that

$$\begin{aligned}\Theta\|\nabla_x w\|^2_{L^2_x(\mathbb{R}^d)} + \Theta^2\|\nabla_x^2 w\|^2_{L^2_x(\mathbb{R}^d)} &= (-\Theta\Delta_x w, \langle h_\epsilon\rangle - M_1)_W \\ &= (\langle h_\epsilon\rangle - M_1 - w, \langle h_\epsilon\rangle - M_1)_W =: \mathcal{A}.\end{aligned} \tag{5-17}$$

It also follows from the same derivation as (5-11) that

$$\|\nabla_x(\partial_t w)\|_{L^2_x(\mathbb{R}^d)} \lesssim \frac{1}{\epsilon}\|h_\epsilon^\perp\|_{L^2(\mathbb{R}^{2d},\,\mathrm{d}m)}.$$

Combining the above two estimates with (5-15), we obtain

$$\mathcal{A} + \epsilon\frac{\mathrm{d}}{\mathrm{d}t}(v\cdot\nabla_x w, h_\epsilon^\perp)_W \lesssim \frac{1}{\epsilon}\mathcal{A}^{1/2}\|h_\epsilon^\perp\|_{L^2(\mathbb{R}^{2d},\,\mathrm{d}m)} + \|h_\epsilon^\perp\|^2_{L^2(\mathbb{R}^{2d},\,\mathrm{d}m)},$$

which implies from the Cauchy–Schwarz inequality that

$$\mathcal{A} + \epsilon\frac{\mathrm{d}}{\mathrm{d}t}(v\cdot\nabla_x w, h_\epsilon^\perp)_W \lesssim \frac{1}{\epsilon^2}\|h_\epsilon^\perp\|^2_{L^2(\mathbb{R}^{2d},\,\mathrm{d}m)}.$$

Denoting the modified entropy by

$$\mathscr{E}_\epsilon(t) := \|h_\epsilon - M_1\|^2_{L^2(\mathbb{R}^{2d},\,\mathrm{d}m)} + \delta\epsilon(v\cdot\nabla_x w, h_\epsilon^\perp)_W$$

with a sufficiently small constant $\delta > 0$ and using (5-14), we conclude that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{E}_\epsilon(t) \lesssim -\mathcal{A} - \|h_\epsilon^\perp\|^2_{L^2(\mathbb{R}^{2d},\,\mathrm{d}m)}. \tag{5-18}$$

Recalling the similar estimate (5-13), we see that $\mathscr{E}_\epsilon$ is equivalent to the square of the $L^2(\mathbb{R}^{2d}, \mathrm{d}m)$-distance between h_ϵ and M_1. It thus suffices to recover $\langle h_\epsilon\rangle - M_1$ by means of $\mathcal{A}$. By (5-16) and the convexity of $|\cdot|$, we know that $|w|$ is a subsolution in the sense that

$$|w| - \Theta\Delta_x|w| \le |\langle h_\epsilon\rangle - M_1| \quad \text{in } \mathbb{R}^d,$$

and hence $\|w\|_{L^1_x(\mathbb{R}^d)} \le \|\langle h_\epsilon\rangle - M_1\|_{L^1_x(\mathbb{R}^d)}$. With the aid of Corollary 4.10,

$$\|w\|_{L^1_x(\mathbb{R}^d)} \le \|h - M_1\|_{L^1(\mathbb{R}^{2d},\,\mathrm{d}m)} \le \|h_{\epsilon,\mathrm{in}} - M_1\|_{L^1(\mathbb{R}^{2d},\,\mathrm{d}m)}.$$

Applying (5-17) and the Nash inequality $\|w\|^{d+2}_{L^2_x(\mathbb{R}^d)} \lesssim \|w\|^2_{L^1_x(\mathbb{R}^d)}\|\nabla_x w\|^d_{L^2_x(\mathbb{R}^d)}$, we then acquire

$$\begin{aligned}\|\langle h_\epsilon\rangle - M_1\|^2_{L^2_x(\mathbb{R}^d)} &\lesssim \mathcal{A} + \|w\|^2_{L^2_x(\mathbb{R}^d)} \lesssim \mathcal{A} + \|h_{\epsilon,\mathrm{in}} - M_1\|^{4/(d+2)}_{L^1(\mathbb{R}^{2d},\,\mathrm{d}m)}\|\nabla_x w\|^{2d/(d+2)}_{L^2_x(\mathbb{R}^d)}\\ &\lesssim (\mathcal{A}^{2/(d+2)} + \|h_{\epsilon,\mathrm{in}} - M_1\|^{4/(d+2)}_{L^1(\mathbb{R}^{2d},\,\mathrm{d}m)})\mathcal{A}^{d/(d+2)}.\end{aligned}$$

Since $\|h_\epsilon - M_1\|^2_{L^2_x(\mathbb{R}^d)} \le (\Lambda + M_1)\|h_{\epsilon,\mathrm{in}} - M_1\|_{L^1(\mathbb{R}^{2d},\,\mathrm{d}m)}$, in both cases

$$\mathcal{A} + \|h_\epsilon^\perp\|^2_{L^2(\mathbb{R}^{2d},\,\mathrm{d}m)} \lesssim \|h_\epsilon - M_1\|^2_{L^2(\mathbb{R}^{2d},\,\mathrm{d}m)},$$

we conclude that

$$\|h_\epsilon - M_1\|^2_{L^2(\mathbb{R}^{2d},\,\mathrm{d}m)} \lesssim (1 + \|h_{\epsilon,\mathrm{in}} - M_1\|^{4/(d+2)}_{L^1(\mathbb{R}^{2d},\,\mathrm{d}m)})(\mathcal{A} + \|h_\epsilon^\perp\|^2_{L^2(\mathbb{R}^{2d},\,\mathrm{d}m)})^{d/(d+2)}.$$

Combining this with (5-18) and the equivalence between $\mathscr{E}_\epsilon$ and $\|h_\epsilon - M_1\|^2_{L^2(\mathbb{R}^{2d},\,\mathrm{d}m)}$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{E}_\epsilon(t) \lesssim -(1 + \|h_{\epsilon,\mathrm{in}} - M_1\|^{4/d}_{L^1(\mathbb{R}^{2d},\,\mathrm{d}m)})^{-1}\mathscr{E}_\epsilon(t)^{1+2/d},$$

Since $\mathscr{E}_\epsilon(0) \lesssim \|h_{\epsilon,\mathrm{in}} - M_1\|_{L^1(\mathbb{R}^{2d},\,\mathrm{d}m)}$, we arrive at

$$\mathscr{E}_\epsilon(t) \lesssim [\mathscr{E}_\epsilon(0)^{-2/d} + (1 + \|h_{\epsilon,\mathrm{in}} - M_1\|^{4/d}_{L^1(\mathbb{R}^{2d},\,\mathrm{d}m)})^{-1}t]^{-d/2} \lesssim (1 + \|h_{\epsilon,\mathrm{in}} - M_1\|^2_{L^1(\mathbb{R}^{2d},\,\mathrm{d}m)})t^{-d/2}. \qquad \square$$

As far as the case $\epsilon = 1$ is concerned, we conclude the result of convergence to equilibrium.

Proof of Theorem 1.1(ii). Consider $g := \mu^{1/2}h$. In view of Proposition 4.8 and Corollary 4.10 with the assumption on initial data, we know that $\lambda \le \mu^{-1/2}g \le \Lambda$ in $\mathbb{R}_+ \times \Omega_x \times \mathbb{R}^d$ for $\Omega_x = \mathbb{T}^d$ or $\mathbb{R}^d$. By applying Proposition 5.1 to $h = \mu^{-1/2}g$ with $\lambda_t = \lambda$ and $\Omega_x = \mathbb{T}^d$, we have an universal constant $c > 0$ such that

$$\|g(t) - M_0\mu^1/2\|_{L^2(\mathbb{T}^d\times\mathbb{R}^d)} \lesssim e^{-2ct}.$$

Combining this with the Sobolev embedding and the interpolation, we derive the following for any $k \in \mathbb{N}$ with $k \ge d$:

$$\begin{aligned}\|g(t) - M_0\mu^{1/2}\|_{\mathcal{C}^k(\mathbb{T}^d\times\mathbb{R}^d)} &\lesssim_k \|g(t) - M_0\mu^{1/2}\|_{H^{2k}(\mathbb{T}^d\times\mathbb{R}^d)}\\ &\lesssim_k \|g(t) - M_0\mu^{1/2}\|^{1/2}_{H^{4k}(\mathbb{T}^d\times\mathbb{R}^d)}\|g(t) - M_0\mu^{1/2}\|^{1/2}_{L^2(\mathbb{T}^d\times\mathbb{R}^d)}.\end{aligned}$$

Since the H^{4k}-norm on the right-hand side is bounded due to the global regularity estimate given by Proposition 4.12, we obtain the exponential convergence to equilibrium in each order derivative.

The asserted result in the case $\Omega_x = \mathbb{R}^d$ is a direct consequence of Proposition 5.2. As a side remark, one is also able to upgrade the long-time convergence to higher-order derivatives of solutions by means of the global regularity estimate and interpolation as above; it yet gives the algebraic decay rate that is not optimal. □

5B. ***Finite-time asymptotics.*** The study of macroscopic dynamics for the nonlinear kinetic model (5-1) in this subsection relies on the regularity of the target equation (1-3). On account of this, let us begin with mentioning some standard results for (1-3) without proof. If the initial data satisfies $\lambda \le \rho_{\mathrm{in}} \le \Lambda$, then such bounds are preserved along times, $\lambda \le \rho \le \Lambda$, in the same spirit as Lemma 4.1. Combining the parabolic De Giorgi–Nash–Moser theory with Schauder theory, we know that the solution ρ is smooth for any positive time. We state the a priori estimate precisely as follows, where its behavior near the initial time is taken into account in view of the standard scaling technique.

Lemma 5.3. *Let $\rho_{\mathrm{in}} \in \mathcal{C}^{\alpha_0}(\mathbb{T}^d)$ be valued in $[\lambda, \Lambda]$ with $\alpha_0 \in (0, 1)$, and let ρ be the solution to (1-3) in $\mathbb{R}_+ \times \mathbb{T}^d$. Then there is some universal constant $\alpha \in (0, 1)$ such that*

$$\|\rho\|_{\mathcal{C}^{\alpha}(\mathbb{R}_+\times\mathbb{T}^d)} \lesssim 1 + \|\rho_{\mathrm{in}}\|_{\mathcal{C}^{\alpha_0}(\mathbb{T}^d)}. \tag{5-19}$$

Moreover, there exists some constant $C_\rho > 0$ depending only on universal constants and $\|\rho_{\mathrm{in}}\|_{\mathcal{C}^{\alpha_0}(\mathbb{T}^d)}$ such that, for any $t \in (0, 1]$ and $x \in \mathbb{T}^d$, we have

$$t^{(1-\alpha)/2}|\nabla_x\rho(t,x)| + t^{(2-\alpha)/2}|\partial_t\rho(t,x)| + t^{(2-\alpha)/2}|\nabla_x^2\rho(t,x)| + t^{(3-\alpha)/2}|\partial_t\nabla_x\rho(t,x)| \le C_\rho, \tag{5-20}$$

and, for any $t \ge 1$, we have

$$\|\nabla_x\rho(t,\cdot)\|_{L^\infty(\mathbb{T}^d)} + \|\partial_t\rho(t,\cdot)\|_{L^\infty(\mathbb{T}^d)} + \|\nabla_x^2\rho(t,\cdot)\|_{L^\infty(\mathbb{T}^d)} + \|\partial_t\nabla_x\rho(t,\cdot)\|_{L^\infty(\mathbb{T}^d)} \lesssim 1. \tag{5-21}$$

We measure the distance between solutions to the scaled nonlinear kinetic model (1-2) and solutions to the fast diffusion equation (1-3) by the relative phi-entropy functional $\mathcal{H}_\beta$ (see Definition 1.3). The following lemma shows the effectiveness of the relative phi-entropy for measuring L^2-distance by virtue of the uniform convexity of φ_β. It can be seen as a simple version of the Csiszár–Kullback inequality on the relative entropy. We give its statement below with a proof taken from [Dolbeault and Li 2018, Proposition 2.1] for the sake of completeness.

Lemma 5.4. *Let h_1 and h_2 be two functions valued in $[0, \Lambda]$. Then we have*

$$\mathcal{H}_\beta(h_1|h_2) \ge \left(1 - \tfrac{1}{2}\beta\right)\Lambda^{-\beta}\|h_1 - h_2\|^2_{L^2(\mathrm{d}m)}. \tag{5-22}$$

If we additionally assume the lower bound $h_1, h_2 \ge \lambda$, then

$$\mathcal{H}_\beta(h_1|h_2) \le \left(1 - \tfrac{1}{2}\beta\right)\lambda^{-\beta}\|h_1 - h_2\|^2_{L^2(\mathrm{d}m)}.$$

Proof. Since $\varphi_\beta(1) = \varphi'_\beta(1) = 0$ and $\beta \in [0, 1]$, for any $z \in \mathbb{R}_+$, there exists $\xi_z \in \mathbb{R}_+$ lying between 1 and z such that

$$\varphi_\beta(z) = \tfrac{1}{2}\varphi''_\beta(\xi_z)(z-1)^2 = \tfrac{1}{2}(2-\beta)\xi_z^{-\beta}(z-1)^2.$$

Since $\min\{z, 1\} \le \xi_z \le \max\{z, 1\}$, we have

$$\int_{\mathbb{T}^d\times\mathbb{R}^d} \max\{h_1^{-\beta}, h_2^{-\beta}\}|h_1-h_2|^2\,\mathrm{d}m \le \frac{2}{2-\beta}\mathcal{H}_\beta(h_1|h_2) \le \int_{\mathbb{T}^d\times\mathbb{R}^d} \min\{h_1^{-\beta}, h_2^{-\beta}\}|h_1-h_2|^2\,\mathrm{d}m,$$

which implies the desired results by using the boundedness of h_1 and h_2. □

Let us consider the finite-time diffusion asymptotics.

Proposition 5.5. *Let $\rho_{\mathrm{in}} \in \mathcal{C}^{\alpha_0}(\mathbb{T}^d)$ be valued in $[\lambda, \Lambda]$ with $\alpha_0 \in (0, 1)$, and let the sequence of functions $\{h_{\epsilon,\mathrm{in}}\}_{\epsilon\in(0,1)} \subset \mathcal{C}^{\alpha_0}(\mathbb{T}^d \times \mathbb{R}^d)$ satisfy*

$$\langle h_{\epsilon,\mathrm{in}}\rangle \ge \lambda \quad \text{in } \mathbb{T}^d \qquad \text{and} \qquad 0 \le h_{\epsilon,\mathrm{in}} \le \Lambda \quad \text{in } \mathbb{T}^d \times \mathbb{R}^d.$$

Let h_ϵ be the solutions of (5-1) associated with this initial data. Then there exist some universal constants $\alpha \in (0, 1)$ and $C > 0$, and some constant $C_\rho > 0$ depending only on universal constants and $\|\rho_{\mathrm{in}}\|_{\mathcal{C}^{\alpha_0}(\mathbb{T}^d)}$ such that, for any $\epsilon \in (0, 1)$ and for any $t \in [\underline{T}, 1]$ with $\underline{T} \in (0, 1)$, the following estimate holds:

$$\mathcal{H}_\beta(h_\epsilon|\rho)(t) \le C_\rho \mathcal{H}_\beta(h_\epsilon|\rho)(\underline{T}) + C_\rho\epsilon(t^{(\alpha-1)/2} + \epsilon t^{(\alpha-2)/2}), \tag{5-23}$$

where $\rho(t, x)$ is the solution to (1-3) associated with the initial data ρ_{in}, and, for any $t \ge 1$, we have

$$\mathcal{H}_\beta(h_\epsilon|\rho)(t) \le [\mathcal{H}_\beta(h_\epsilon|\rho)(t)(1) + C\epsilon(1 + t^{1/2})]e^{Ct}. \tag{5-24}$$

Proof. For $\beta \in [0, 1)$, the phi-entropy of h_ϵ relative to ρ reads

$$\mathcal{H}_\beta(h_\epsilon|\rho) = \mathcal{H}_\beta(h_\epsilon|1) - \mathcal{H}_\beta(\rho|1) - \frac{2-\beta}{1-\beta}(\langle h_\epsilon\rangle - \rho, \rho^{1-\beta} - 1).$$

As far as the entropy $\mathcal{H}_\beta(h_\epsilon|1)$ is concerned, the entropy dissipation is derived by (5-1), integration by parts, and using Hölder's inequality $\langle\nabla_v h_\epsilon\rangle^2 \le \langle h_\epsilon\rangle^\beta\langle h_\epsilon^{-\beta}|\nabla_v h_\epsilon|^2\rangle$:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}_\beta(h_\epsilon|1) &= \frac{2-\beta}{1-\beta}(h^{1-\beta}, h_t) = -\frac{2-\beta}{\epsilon^2}(h_\epsilon^{-\beta}\nabla_v h_\epsilon, \langle h_\epsilon\rangle^\beta\nabla_v h_\epsilon)\\ &\le -\frac{2-\beta}{\epsilon^2}\|\langle\nabla_v h_\epsilon\rangle\|_{L^2_x}^2 = -\frac{2-\beta}{\epsilon^2}\|\langle v h_\epsilon\rangle\|_{L^2_x}^2.\end{aligned} \tag{5-25}$$

In view of the limiting equation (1-3), we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}_\beta(\rho|1) = \frac{2-\beta}{1-\beta}(\rho^{1-\beta}, \partial_t\rho) = -(2-\beta)(\rho^{-\beta}\nabla_x\rho, \rho^{-\beta}\nabla_x\rho). \tag{5-26}$$

A direct computation with the macroscopic equation (5-3) and (1-3) leads to

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}(\langle h_\epsilon\rangle - \rho, \rho^{1-\beta} - 1)\\ &\quad= (\rho^{1-\beta}, \partial_t\langle h_\epsilon\rangle) + ((1-\beta)\rho^{-\beta}\langle h_\epsilon\rangle - (2-\beta)\rho^{1-\beta}, \partial_t\rho)\\ &\quad= \frac{1-\beta}{\epsilon}(\rho^{-\beta}\nabla_x\rho, \langle v h_\epsilon\rangle) - ((1-\beta)\nabla_x(\rho^{-\beta}\langle h_\epsilon\rangle) - (2-\beta)\rho^{-\beta}\nabla_x\rho, \rho^{-\beta}\nabla_x\rho).\end{aligned} \tag{5-27}$$

The evolution of $\mathcal{H}_\beta(h_\epsilon\,|\,\rho)$ is then estimated by combining (5-25), (5-26), and (5-27):

$$\begin{aligned}\frac{1}{2-\beta}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}_\beta(h_\epsilon\,|\,\rho) &\le -\frac{1}{\epsilon^2}\|\langle vh_\epsilon\rangle\|_{L^2_x}^2 - \frac{1}{\epsilon}(\langle vh_\epsilon\rangle,\, \rho^{-\beta}\nabla_x\rho) + (\nabla_x(\rho^{-\beta}\langle h_\epsilon\rangle - \rho^{1-\beta}),\, \rho^{-\beta}\nabla_x\rho)\\ &= -\|\epsilon^{-1}\langle vh_\epsilon\rangle + \rho^{-\beta}\nabla_x\rho\|_{L^2_x}^2 + (\epsilon^{-1}\langle vh_\epsilon\rangle,\, \rho^{-\beta}\nabla_x\rho)\\ &\qquad + (\nabla_x\langle h_\epsilon\rangle + \beta(1-\rho^{-1}\langle h_\epsilon\rangle)\nabla_x\rho,\, \rho^{-2\beta}\nabla_x\rho)\end{aligned}$$

We remark that the above inequality also holds for $\beta = 1$ by a similar computation. Abbreviate

$$Q_\epsilon := \epsilon^{-1}\langle vh_\epsilon\rangle + \rho^{-\beta}\nabla_x\rho \quad\text{and}\quad R_\epsilon := -\nabla_x\cdot\langle v\otimes\nabla_v h_\epsilon\rangle - \epsilon\partial_t\langle vh_\epsilon\rangle,$$

and write the macroscopic equation (5-4) in the form $\nabla_x\langle h_\epsilon\rangle = -\epsilon^{-1}\langle h_\epsilon\rangle^\beta\langle vh_\epsilon\rangle + R_\epsilon$. We then have

$$\begin{aligned}\frac{1}{2-\beta}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}_\beta(h_\epsilon\,|\,\rho) &\le -\|Q_\epsilon\|_{L^2_x}^2 + ((1-\rho^{-\beta}\langle h_\epsilon\rangle^\beta)Q_\epsilon,\, \rho^{-\beta}\nabla_x\rho) + (R_\epsilon,\, \rho^{-2\beta}\nabla_x\rho)\\ &\qquad + (\rho^{-\beta}\langle h_\epsilon\rangle^\beta - \beta\rho^{-1}\langle h_\epsilon\rangle - 1 + \beta,\, \rho^{-2\beta}|\nabla_x\rho|^2)\\ &\le 2\|\rho^{-1-\beta}|\nabla_x\rho|\|_{L^\infty_{t,x}}^2\|\langle h_\epsilon\rangle - \rho\|_{L^2_x}^2 + (R_\epsilon,\, \rho^{-2\beta}\nabla_x\rho),\end{aligned} \tag{5-28}$$

where, for the second inequality, we used the Cauchy–Schwarz inequality

$$2((1-\rho^{-\beta}\langle h_\epsilon\rangle^\beta)Q_\epsilon,\, \rho^{-\beta}\nabla_x\rho) \le \|Q_\epsilon\|_{L^2_x}^2 + (|1-\rho^{-\beta}\langle h_\epsilon\rangle^\beta|^2,\, |\rho^{-\beta}\nabla_x\rho|^2)$$

and the following two elementary inequalities with $\beta\in[0,1]$:

$$|z^\beta - 1| \le |z-1| \quad\text{and}\quad |z^\beta - \beta z - 1 + \beta| \le |z-1|^2 \qquad \text{for any } z\in\mathbb{R}_+.$$

In view of Hölder's inequality and inequality (5-22) given in Lemma 5.4, we know that

$$\|\langle h_\epsilon\rangle - \rho\|_{L^2_x}^2 \le \|h_\epsilon - \rho\|_{L^2(\mathrm{d}m)}^2 \le 2\Lambda^\beta\mathcal{H}_\beta(h_\epsilon\,|\,\rho).$$

Combining this with (5-28) and (5-20), we derive that, for any $t\in(0,1]$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}_\beta(h_\epsilon\,|\,\rho) \le C_\rho t^{\alpha-1}\mathcal{H}_\beta(h_\epsilon\,|\,\rho) + 2(R_\epsilon,\, \rho^{-2\beta}\nabla_x\rho), \tag{5-29}$$

where the constants $\alpha\in(0,1)$ and $C_\rho>0$ are provided by Lemma 5.3.

We point out that, after integrating in time, the remainder term in (5-29) involving R_ϵ is of order $O(\epsilon)$ due to the control of the entropy production and the regularity of the limiting equation. Indeed,

$$\begin{aligned}\int_0^t (R_\epsilon,\, \rho^{-2\beta}\nabla_x\rho) &= \int_0^t (\langle v\otimes\nabla_v h_\epsilon\rangle,\, \nabla_x(\rho^{-2\beta}\nabla_x\rho)) + \epsilon\int_0^t (\langle vh_\epsilon\rangle,\, \partial_t(\rho^{-2\beta}\nabla_x\rho))\\ &\qquad - \epsilon(\langle vh_\epsilon\rangle,\, \rho^{-2\beta}\nabla_x\rho)(t) + \epsilon(\langle vh_\epsilon\rangle,\, \rho^{-2\beta}\nabla_x\rho)(0)\\ &\lesssim (\|\nabla_x(\rho^{-2\beta}\nabla_x\rho)\|_{L^\infty_{t,x}} + \epsilon\|\partial_t(\rho^{-2\beta}\nabla_x\rho)\|_{L^\infty_{t,x}})\int_0^t \|\langle vh_\epsilon\rangle\|_{L^2_x}\\ &\qquad + \epsilon\|\rho^{-2\beta}\nabla_x\rho\|_{L^\infty_{t,x}}\|\langle vh_\epsilon\rangle\|_{L^\infty([0,T];L^2_x)}.\end{aligned}$$

It then follows from (5-20), (5-25), and the global upper bound of h_ϵ (Lemma 4.1) that, for any $t \in (0, 1]$,

$$\begin{aligned}\int_0^t (R_\epsilon, \rho^{-2\beta}\nabla_x \rho) &\le C_\rho(t^{(\alpha-1)/2} + \epsilon t^{(\alpha-2)/2})\left(\int_0^t \|\langle v h_\epsilon\rangle\|_{L^2_x}^2\right)^{1/2} + C_\rho \epsilon t^{(\alpha-1)/2} \\ &\le C_\rho(\epsilon t^{(\alpha-1)/2} + \epsilon^2 t^{(\alpha-2)/2}) \sup_{s\in[0,t]} \sqrt{\mathcal{H}_\beta(h_\epsilon \,|\, 1)(s)} + C_\rho \epsilon t^{(\alpha-1)/2} \\ &\le C_\rho(\epsilon t^{(\alpha-1)/2} + \epsilon^2 t^{(\alpha-2)/2}) + C_\rho \epsilon t^{(\alpha-1)/2}.\end{aligned}$$

Combining this estimate with (5-29) as well as Grönwall's inequality, we conclude (5-23). Additionally, we arrive at (5-24) if we apply Lemma 5.3 with (5-21) instead of (5-20) in the above argument. □

We are now in a position to conclude the global-in-time diffusion asymptotics.

Proof of Theorem 1.4. We are going to combine Propositions 5.1 and 5.5 with a delicate analysis on the relative entropy around the initial time to get Theorem 1.4. The analysis is based on the barrier function method. Let us assume the constant $\alpha \in (0, 1)$ provided by Proposition 5.5.

Step 1. Pointwise estimate. Let us fix $\delta \in (0, 1)$ and $(x_1, v_1) \in \mathbb{T}^d \times B_R$ with $R > 0$ and consider the function

$$\bar{h}(t, x, v) := C_1 t + C_2(|x - x_1 - \epsilon^{-1} t v|^2 + |v - v_1|^2),$$

where the constants $C_1, C_2 > 0$ are to be determined. For any $t \le \epsilon\delta/(4(1 + R))$, we have

$$\bar{h} \ge C_2(|x - x_1|^2 + |v - v_1|^2 - 2\epsilon^{-1} t|x - x_1||v|) \ge \tfrac{1}{2} C_2 \delta^2 = \Lambda \quad \text{on } \partial B_\delta(x_1, v_1),$$

where we chose $C_2 := 2\delta^{-2}\Lambda$. For any $(x, v) \in B_\delta(x_1, v_1)$,

$$|\langle h\rangle^\beta \mathscr{L}_{\mathrm{OU}} \bar{h}| \lesssim |\Delta_v \bar{h}| + |v \cdot \nabla_v \bar{h}| \lesssim \delta^{-2}(1 + R^2)(1 + \epsilon^{-2} t^2).$$

Therefore, for any $t \le \epsilon\delta/(4(1 + R))$ and $(x, v) \in B_\delta(x_1, v_1)$,

$$(\partial_t + \epsilon^{-1} v \cdot \nabla_x - \epsilon^{-2}\langle h\rangle^\beta \mathscr{L}_{\mathrm{OU}})\bar{h} \ge C_1 - C_0 \epsilon^{-2}\delta^{-2}(1 + R^2)(1 + \epsilon^{-2} t^2) \ge 0,$$

where the constant $C_0 > 0$ is universal and we chose $C_1 := 2C_0\epsilon^{-2}\delta^{-2}(1 + R^2)$. Then, the maximum principle implies that, for any $t \le \epsilon\delta/(4(1 + R))$ and $(x, v) \in B_\delta(x_1, v_1)$,

$$|h_\epsilon(t, x, v) - h_{\epsilon,\mathrm{in}}(x_1, v_1)| \le \bar{h}(t, x, v) + \sup_{B_\delta(x_1,v_1)} |h_{\epsilon,\mathrm{in}}(x, v) - h_{\epsilon,\mathrm{in}}(x_1, v_1)|. \tag{5-30}$$

In particular, for any $t \le \epsilon\delta/(4(1 + R))$ and $(x_1, v_1) \in \mathbb{T}^d \times B_R$,

$$|h_\epsilon(t, x_1, v_1) - h_{\epsilon,\mathrm{in}}(x_1, v_1)| \lesssim \epsilon^{-2}\delta^{-2}(1 + R^2)t + \|h_{\epsilon,\mathrm{in}}\|_{\mathcal{C}^{\alpha_0}(\mathbb{T}^d\times\mathbb{R}^d)}\delta^\alpha. \tag{5-31}$$

As far as the solution ρ to the limiting equation (1-3) is concerned, using the Hölder estimate (5-19) in Lemma 5.3, we derive that, for any $t \in \mathbb{R}_+$,

$$\|\rho(t) - \rho_{\mathrm{in}}\|_{L^\infty(\mathbb{T}^d)} \lesssim (1 + \|\rho_{\mathrm{in}}\|_{\mathcal{C}^{\alpha_0}(\mathbb{T}^d)})t^\alpha. \tag{5-32}$$

Step 2. Estimate of the relative entropy around the initial time. Let us restrict our attention to the time interval $[0, \underline{T}]$ with $\underline{T} \in (0, 1)$ to be determined. Compute the relative entropy $\mathcal{H}_\beta(h_\epsilon \,|\, \rho)$ in terms of initial data as

$$\mathcal{H}_\beta(h_\epsilon \,|\, \rho) = \mathcal{H}_\beta(h_{\epsilon,\mathrm{in}} \,|\, \rho_{\mathrm{in}}) + \int_{\mathbb{T}^d \times \mathbb{R}^d} [(\varphi_\beta(h_\epsilon) - \varphi_\beta(h_{\epsilon,\mathrm{in}})) - (\varphi_\beta(\rho) - \varphi_\beta(\rho_{\mathrm{in}}))] \,\mathrm{d}m$$
$$- \int_{\mathbb{T}^d \times \mathbb{R}^d} [(\varphi'_\beta(\rho) - \varphi'_\beta(\rho_{\mathrm{in}}))(h_{\epsilon,\mathrm{in}} - \rho_{\mathrm{in}}) + \varphi'_\beta(\rho)(h_\epsilon - h_{\epsilon,\mathrm{in}}) - \varphi'_\beta(\rho)(\rho - \rho_{\mathrm{in}})] \,\mathrm{d}m. \quad (5\text{-}33)$$

Consider a truncation in v for the integrals on the right-hand side. Since h_ϵ, $h_{\epsilon,\mathrm{in}}$, ρ, and ρ_{in} are all bounded from above (Lemma 4.1), we have

$$\int_{\mathbb{T}^d \times B_R^c} [|\varphi_\beta(h_\epsilon) - \varphi_\beta(h_{\epsilon,\mathrm{in}})| + |\varphi_\beta(\rho) - \varphi_\beta(\rho_{\mathrm{in}})|] \,\mathrm{d}m \lesssim \int_{B_R^c} \mathrm{d}\mu \lesssim R^{-4}. \quad (5\text{-}34)$$

Observe that for any $a, b \in (0, \Lambda]$, there exists $\xi \in \mathbb{R}_+$ lying between a and b such that

$$\varphi_\beta(a^2) - \varphi_\beta(b^2) = 2\xi \varphi'_\beta(\xi^2)(a - b).$$

Meanwhile, $|\xi \varphi'_\beta(\xi^2)| \lesssim 1$ for any $\xi \in (0, \Lambda]$. Thus,

$$|\varphi_\beta(h_\epsilon) - \varphi_\beta(h_{\epsilon,\mathrm{in}})| \lesssim |h_\epsilon - h_{\epsilon,\mathrm{in}}|^{1/2} \quad \text{and} \quad |\varphi_\beta(\rho) - \varphi_\beta(\rho_{\mathrm{in}})| \lesssim |\rho - \rho_{\mathrm{in}}|^{1/2}.$$

Set $R := \epsilon^{-\eta/4}$, $\delta := \epsilon^{\eta/4}$, and $\underline{T} := \frac{1}{8}\epsilon^{2+2\eta}$ for some constant $\eta \in (0, 1)$ to be determined. In this setting, $\underline{T} \le \epsilon\delta/(4\langle R\rangle)$. It then follows from (5-31), (5-32), and (5-34) that, for any $t \le \underline{T}$,

$$\int_{\mathbb{T}^d \times \mathbb{R}^d} [|\varphi_\beta(h_\epsilon) - \varphi_\beta(h_{\epsilon,\mathrm{in}})| + |\varphi_\beta(\rho) - \varphi_\beta(\rho_{\mathrm{in}})|] \,\mathrm{d}m$$
$$\lesssim R^{-4} + \int_{\mathbb{T}^d \times B_R} [|h_\epsilon - h_{\epsilon,\mathrm{in}}|^{1/2} + |\rho - \rho_{\mathrm{in}}|^{1/2}] \,\mathrm{d}m$$
$$\lesssim \epsilon^\eta + \epsilon^{\eta/2} + \|h_{\epsilon,\mathrm{in}}\|^{1/2}_{\mathcal{C}^{\alpha_0}(\mathbb{T}^d \times \mathbb{R}^d)} \epsilon^{\alpha\eta/8} + (1 + \|\rho_{\mathrm{in}}\|^{1/2}_{\mathcal{C}^{\alpha_0}(\mathbb{T}^d)}) \epsilon^{\alpha + \alpha\eta}. \quad (5\text{-}35)$$

In addition,

$$\|h_\epsilon - h_{\epsilon,\mathrm{in}}\|_{L^1(\mathbb{T}^d \times B_R^c,\, \mathrm{d}m)} \lesssim R^{-4}$$

and $|\varphi'_\beta| \lesssim 1$ on $[\lambda, \Lambda]$. Therefore, combining (5-31) and (5-32) with inequality (5-22) given in Lemma 5.4 yields, for any $t \le \underline{T}$,

$$\int_{\mathbb{T}^d \times \mathbb{R}^d} [|(\varphi'_\beta(\rho) - \varphi'_\beta(\rho_{\mathrm{in}}))(h_{\epsilon,\mathrm{in}} - \rho_{\mathrm{in}})| + |\varphi'_\beta(\rho)(h_\epsilon - h_{\epsilon,\mathrm{in}})| + |\varphi'_\beta(\rho)(\rho - \rho_{\mathrm{in}})|] \,\mathrm{d}m$$
$$\lesssim \|h_{\epsilon,\mathrm{in}} - \rho_{\mathrm{in}}\|_{L^2(\mathbb{T}^d \times \mathbb{R}^d,\, \mathrm{d}m)} + \|h_\epsilon - h_{\epsilon,\mathrm{in}}\|_{L^1(\mathbb{T}^d \times \mathbb{R}^d,\, \mathrm{d}m)} + \|\rho - \rho_{\mathrm{in}}\|_{L^1(\mathbb{T}^d)}$$
$$\lesssim \mathcal{H}_\beta^{1/2}(h_{\epsilon,\mathrm{in}} \,|\, \rho_{\mathrm{in}}) + R^{-4} + \|h_\epsilon - h_{\epsilon,\mathrm{in}}\|_{L^1(\mathbb{T}^d \times B_R,\, \mathrm{d}m)} + \|\rho - \rho_{\mathrm{in}}\|_{L^1(\mathbb{T}^d)}$$
$$\lesssim \epsilon'^{1/2} + \epsilon^\eta + \|h_{\epsilon,\mathrm{in}}\|_{\mathcal{C}^{\alpha_0}(\mathbb{T}^d \times \mathbb{R}^d)} \epsilon^{\alpha\eta/4} + (1 + \|\rho_{\mathrm{in}}\|_{\mathcal{C}^{\alpha_0}(\mathbb{T}^d)}) \epsilon^{2\alpha + 2\alpha\eta}. \quad (5\text{-}36)$$

Plugging (5-35) and (5-36) into expression (5-33), we derive, for any $t \le \underline{T}$,

$$\mathcal{H}_\beta(h_\epsilon \,|\, \rho)(t) \lesssim (1 + \|h_{\epsilon,\mathrm{in}}\|_{\mathcal{C}^{\alpha_0}(\mathbb{T}^d \times \mathbb{R}^d)} + \|\rho_{\mathrm{in}}\|_{\mathcal{C}^{\alpha_0}(\mathbb{T}^d)})(\epsilon + \epsilon')^{\alpha\eta/8}. \quad (5\text{-}37)$$

Step 3. Conclusion. Recall that we have chosen $\underline{T} = \frac{1}{8}\epsilon^{2+2\eta}$. In view of (5-37) and the estimate (5-23) given in Proposition 5.5, one may optimize in η to get the result. For simplicity, we pick $\eta := \alpha/(4-2\alpha)$ so that $\underline{T} = \frac{1}{8}\epsilon^{(4-\alpha)/(2-\alpha)}$ and $\epsilon(\underline{T}^{(\alpha-1)/2} + \epsilon\underline{T}^{(\alpha-2)/2}) \lesssim \epsilon^{\alpha/2}$. It turns out that, for any $t \in [0, 1]$,

$$\mathcal{H}_\beta(h_\epsilon|\rho)(t) \le C_\rho \mathcal{H}_\beta(h_\epsilon|\rho)(\underline{T}) + C_\rho\epsilon(\underline{T}^{(\alpha-1)/2} + \epsilon\underline{T}^{(\alpha-2)/2}) \le C_*(\epsilon+\epsilon')^{\alpha\eta/8}, \tag{5-38}$$

where the constant $C_\rho > 0$ is provided in Proposition 5.5 and the constant $C_* > 0$ depends only on universal constants, $\|\rho_{\text{in}}\|_{\mathcal{C}^{\alpha_0}(\mathbb{T}^d)}$, and $\|h_{\epsilon,\text{in}}\|_{\mathcal{C}^{\alpha_0}(\mathbb{T}^d\times\mathbb{R}^d)}$. Then, using the estimate (5-24) given in Proposition 5.5 with (5-22), we arrive at point (i) of Theorem 1.4.

As for point (ii) of Theorem 1.4, applying (5-24) together with (5-22) and (5-38), for any $t \in [1, \overline{T}]$, we have

$$\begin{aligned}\|h_\epsilon(t) - \rho(t)\|^2_{L^2(\mathrm{d}m)} &\lesssim [\mathcal{H}_\beta(h_\epsilon|\rho)(1) + \epsilon(1+\overline{T}^{1/2})]e^{C\overline{T}}\\ &\lesssim [C_*(\epsilon+\epsilon')^{\alpha\eta/8} + \epsilon(1+(-\iota\log(\epsilon+\epsilon'))^{1/2})](\epsilon+\epsilon')^{-\alpha\eta/16} \lesssim C_*(\epsilon+\epsilon')^{\alpha\eta/16},\end{aligned}$$

where we picked $\overline{T} := -\iota\log(\epsilon+\epsilon')$ with $\iota := \alpha\eta/(16C)$. Finally, using Proposition 5.1 with the additional assumption that $h_{\epsilon,\text{in}} \ge \lambda$, we know from the long-time behavior that there is some universal constant $c > 0$ such that, for any $t \ge \overline{T}$,

$$\|h_\epsilon(t) - \rho(t)\|_{L^2(\mathrm{d}m)} \le \|h_\epsilon(t) - M_0\|_{L^2(\mathrm{d}m)} + \|\rho(t) - M_0\|_{L^2_x} \lesssim e^{-c\overline{T}} = (\epsilon+\epsilon')^{c\iota}. \qquad \square$$

Appendix A: Maximum principle

The following maximum principle (on a not necessarily bounded domain) is repeatedly applied throughout the article. We state it in a more suitable fashion for the Fokker–Planck equations of our concern, whose proof is in the same spirit as [Cameron et al. 2018, Lemma A.2].

Lemma A.1. *Let the domain ω be a subset of $\mathbb{R}^d \times \mathbb{R}^d$ and the parabolic cylinder $\omega_T := (0, T] \times \omega$. If $f \in \mathcal{C}^2_{\text{kin}}(\omega_T) \cap \mathcal{C}^0(\overline{\omega}_T)$ is a bounded subsolution in the sense that*

$$\mathscr{L}_1 f := (\partial_t + v\cdot\nabla_x)f - \operatorname{tr}(AD^2_v f) - B\cdot\nabla_v f \le 0 \quad \text{in } \omega_T, \tag{A-1}$$

with the coefficients $A(t, x, v), B(t, x, v) \in \mathcal{C}^0(\omega_T)$ satisfying

$$\lambda|\xi|^2 \le A(t,x,v)\xi\cdot\xi \le \Lambda|\xi|^2 \quad \text{and} \quad |B(t,x,v)\cdot\xi| \le \Lambda\langle v\rangle|\xi| \qquad \text{for any } \xi\in\mathbb{R}^d,\ (t,x,v)\in\omega_T,$$

then $\sup_{\omega_T} f \le \sup_{\partial_p\omega_T} f$, where the parabolic boundary $\partial_p\omega_T$ is defined to be $[0, T] \times \overline{\omega} - (0, T] \times \omega$.

Proof. If the domain ω is bounded, then the result is classical. For general (unbounded) ω, we consider the auxiliary functions

$$\phi_1(t, v) := e^{C_1 t}\langle v\rangle^2 \quad \text{and} \quad \phi_2(t, x) := e^{C_2 t}\langle x\rangle^2$$

with $C_1, C_2 > 0$. Since f is bounded, for any $\varepsilon_1, \varepsilon_2 > 0$, there exists $R(\varepsilon_1), R(\varepsilon_2) > 0$ (independent of C_1 and C_2) such that $f - \varepsilon_1\phi_1 - \varepsilon_2\phi_2 \le \sup_{\partial_p\omega_T} f$ in $\omega_T \cap \{|x| \ge R(\varepsilon_2) \text{ or } |v| \ge R(\varepsilon_1)\}$.

By choosing $C_1 = (d+2)\Lambda$, we have

$$\mathscr{L}_1\phi_1 = e^{C_1 t}(C_1\langle v\rangle^2 - \operatorname{tr}(A) - 2B\cdot v) \ge (C_1 - (d+2)\Lambda)\langle v\rangle^2 = 0 \quad \text{in } \omega_T.$$

For any $R_1 \geq R(\varepsilon_1)$, there exists $C_2 > 0$ depending only on R_1 such that

$$\mathscr{L}_1\phi_2 = e^{C_2 t}(C_2\langle x\rangle^2 + v\cdot x) \geq (C_2-1)\langle x\rangle^2 - |v|^2 \geq 0 \quad \text{in } \omega_T \cap \{|v| < R_1\}.$$

Therefore, for any $R_2 > R(\varepsilon_2)$, we have that $f - \varepsilon_1\phi_1 - \varepsilon_2\phi_2$ is a subsolution to (3-1) in the bounded domain $(0, T]\times(\omega\cap(B_{R_2}\times B_{R_1}))$ with the data smaller than $\sup_{\partial_p\omega_T} f$ on the boundary portion contained in $\{|x| = R_2 \text{ or } |v| = R_1\}$. Then, applying the classical maximum principle yields

$$f - \varepsilon_1\phi_1 - \varepsilon_2\phi_2 \leq \sup_{\partial_p\omega_T} f \quad \text{in } (0, T]\times(\omega\cap(B_{R_2}\times B_{R_1})).$$

Sending $R_2 \to 0$, $\varepsilon_2 \to 0$, $R_1 \to 0$, and $\varepsilon_1 \to 0$ in order, we get the conclusion. □

Appendix B: Spreading of positivity

This appendix is devoted to the proof of Proposition 4.2. The argument follows the one presented in [Henderson et al. 2020b], and it is based on the combination of Lemmas 4.5 and 4.6.

Proof of Proposition 4.2. The proof is split into four steps.

Step 1. Spreading positivity for all velocities for short times. Applying Lemma 4.5 (with $\tau = 1$) yields that there is some universal constant $c_0 > 0$ such that, for any $0 \leq t \leq \min\{1, T, c_0\langle r^{-1}\rangle^{-2}\langle v_0\rangle^{-2}\}$,

$$h(t, x, v) \geq \tfrac{1}{8}\delta\mathbb{1}_{\{|x-x_0-tv|<r/2,\, |v-v_0|<r/2\}} \geq \tfrac{1}{8}\delta\mathbb{1}_{\{|x-x_0-tv_0|<r/4,\, |v-v_0|<r/4\}}.$$

Let $r_0 := \min\{1, \frac{1}{16}r\}$ and $\underline{t} := \frac{1}{2}\underline{T}$. Then, Lemma 4.6 implies that there exists $\underline{C}_0 > 0$ depending only on universal constants, $\underline{T}$, δ, r, and v_0 such that, for any $0 < \underline{t} \leq t \leq T_0$ with

$$T_0 := \min\{1, T, c_0\langle r^{-1}\rangle^{-2}\langle v_0\rangle^{-2}, \tfrac{1}{4}r_0\langle v_0\rangle^{-1}\}$$

and $v \in \mathbb{R}^d$, we have

$$h(t, x, v) \geq \underline{C}_0^{-1}e^{-\underline{C}_0|v-v_0|^4}\mathbb{1}_{\{|x-x_0-tv_0|<2r_0\}} \geq \underline{C}_0^{-1}e^{-\underline{C}_0|v-v_0|^4}\mathbb{1}_{\{|x-x_0|<r_0\}}. \tag{B-1}$$

Step 2. Spreading positivity in space for short times. For any fixed $\bar{t} \in [\underline{t}, T_0]$ and $\bar{x} \in \mathbb{T}^d$, we set $\bar{v} := (\bar{x} - x_0)/(\bar{t} - \underline{t})$. In view of (B-1), by Lemma 4.5 (with $\tau = 2(\bar{t} - \underline{t})$, $v_0 = \bar{v}$), we deduce that, if $\bar{t} - \underline{t} \leq c_0\langle 2(\bar{t} - \underline{t})r_0^{-1}\rangle^{-2}\langle\bar{v}\rangle^{-2}$ and in particular if

$$\bar{t} \leq \underline{t} + \bar{t}_0 \quad \text{with } \bar{t}_0 := \frac{c_0 r_0^2}{4 + r_0^2}\langle\bar{x} - x_0\rangle^{-2}, \tag{B-2}$$

then there exists $\delta_0 > 0$ with the same dependence as $\underline{C}_0$ such that, for any $t \in [\underline{t}, \bar{t}]$,

$$h(t, x, v) \geq \delta_0\mathbb{1}_{\{|x-x_0-(t-\underline{t})v|<r_0/2,\, 2(\bar{t}-\underline{t})|v-\bar{v}|<r_0/2\}} \geq \delta_0\mathbb{1}_{\{|x-x_0-(t-\underline{t})\bar{v}|<r_0/4,\, |v-\bar{v}|<r_0/4\}}.$$

Then, Lemma 4.6 (with $v_0 = \bar{v}$) implies that, for any $0 < 2\underline{t} \leq t \leq \underline{t} + \bar{t}_0$ and $v \in \mathbb{R}^d$,

$$h(t, x, v) \geq \underline{C}_1^{-1}e^{-\underline{C}_1|v|^4}\mathbb{1}_{\{|x-x_0-(t-\underline{t})\bar{v}|<r_0/8\}} \tag{B-3}$$

for some constant $\underline{C}_1 > 0$ depending only on universal constants, $\underline{T}$, δ, r, v_0, and $|\bar{x} - x_0|$. In particular, for any $0 < 2\underline{t} \le \bar{t} \le \underline{t} + \bar{t}_0$ and $v \in \mathbb{R}^d$,

$$h(\bar{t}, \bar{x}, v) \ge \underline{C}_1^{-1} e^{-\underline{C}_1 |v|^4}. \tag{B-4}$$

Step 3. Spreading positivity for any finite time. We observe that the time interval above is restricted (see (B-2)), but it can be removed by applying the lemmas again. Based on the previous step, it suffices to deal with the case $\bar{t} > \bar{t}_0$. By a similar proof to (B-3), we derive

$$h(\bar{t}_0, x, v) \ge \delta_1 \mathbb{1}_{\{|x-\bar{x}|<r_0/8,\, |v|<r_0/8\}}$$

for some constant $\delta_1 > 0$ with the same dependence as $\underline{C}_1$. In view of this data, applying Lemma 4.5 to $h(\bar{t}_0 + \cdot, \cdot, \cdot)$ (with $\tau = 1$, $v_0 = 0$), we see that, for any $t \in [\bar{t}_0, \min\{T_0, \bar{t}_0 + T_1\}]$ with $T_1 := c_0 \langle 8/r_0 \rangle^{-2}$,

$$h(t, x, v) \ge \tfrac{1}{8} \delta_1 \mathbb{1}_{\{|x-\bar{x}|<r_0/16,\, |v|<r_0/16\}}.$$

It then follows from Lemma 4.6 that, for any $t \in [\bar{t}_0 + \underline{t}, \min\{T_0, \bar{t}_0 + T_1\}]$ and $v \in \mathbb{R}^d$,

$$h(t, x, v) \ge \underline{C}_2^{-1} e^{-\underline{C}_2 |v|^4} \mathbb{1}_{\{|x-\bar{x}|<r_0/32\}}$$

for some constant $\underline{C}_2 > 0$ with the same dependence as $\underline{C}_1$.

Combining this with (B-4) as well as recalling that $\underline{T} = 2\underline{t}$ and the space domain $\mathbb{T}^d$ is compact, we know that there exists $\underline{C}_3 > 0$ depending only on universal constants, $\underline{T}$, δ, r, and v_0 such that, for any $(t, x, v) \in [\underline{T}, \min\{T_0, T_1\}] \times \mathbb{T}^d \times \mathbb{R}^d$,

$$h(t, x, v) \ge \underline{C}_3^{-1} e^{-\underline{C}_3 |v|^4}.$$

Since T_0 and T_1 depend only on universal constants, r, and v_0, by applying the above arguments iteratively, we obtain the result for any finite time.

Step 4. Improving the exponential tail. We remark that this step is not necessary for the applications of the lower bound result, but it shows a more precise decay rate as $|v| \to \infty$.

By the previous step, there is some $\underline{c} > 0$ depending only on universal constants, $\underline{T}$, T, δ, r, and v_0 such that $h \ge \underline{c}$ in $[\underline{T}, T] \times \mathbb{T}^d \times B_1$. Consider the barrier function

$$\underline{h}(t, x, v) := \underline{c} e^{-C_0 (t-\underline{T})^{-1} |v|^2} \quad \text{in } [\underline{T}, T] \times \mathbb{T}^d \times B_1^c,$$

where the constant $C_0 > 1$ is to be determined. By recalling (4-2) and performing a direct computation, we have

$$\begin{aligned}(\partial_t + v \cdot \nabla_x)\underline{h} - \mathscr{R}_h \mathscr{L}_{\mathrm{OU}} \underline{h} &= \frac{C_0 \mathscr{R}_h \underline{h}}{(t-\underline{T})^2} (\mathscr{R}_h^{-1} + 2(d - |v|^2)(t - \underline{T}) - 4C_0 |v|^2) \\ &\le \frac{C_0 \mathscr{R}_h \underline{h}}{(t-\underline{T})^2} (\underline{c}^{-\beta} + 2dT - 4C_0) \qquad \text{in } (\underline{T}, T] \times \mathbb{T}^d \times B_1^c.\end{aligned}$$

In particular, by choosing C_0 sufficiently large (with the same dependence as $\underline{c}$), we have

$$(\partial_t + v \cdot \nabla_x)(\underline{h} - h) - \mathscr{R}_h \mathscr{L}_{\mathrm{OU}} (\underline{h} - h) \le 0 \quad \text{in } (\underline{T}, T] \times \mathbb{T}^d \times B_1^c.$$

In addition, by its definition, $h \geq \underline{h}$ on the boundary $\{t \in [\underline{T}, T],\ |v| = 1\} \cup \{t = 2\underline{t},\ |v| \geq 1\}$. The maximum principle (Lemma A.1) then implies that $h \geq \underline{h}$ in $[\underline{T}, T] \times \mathbb{T}^d \times B_1^c$. Therefore, we achieve the Gaussian-type lower bound for any $(t, x, v) \in [2\underline{T}, T] \times \mathbb{T}^d \times \mathbb{R}^d$. □

Appendix C: Gaining regularity of spatial increment

This appendix is devoted to the proof of two technical lemmas for spatial increments involved in the bootstrapping of higher regularity for solutions to (4-1) presented in Section 4C. For the convenience of the reader, we report a brief proof following the lines of [Imbert and Silvestre 2022, Lemma 8.1] with $s = 1$ and $\alpha_1 = \beta = 2$.

Lemma C.1. *Let $\alpha \in (0, 1)$, and let a bounded continuous function g be defined in Q_4. If there exists some constant $M > 0$ such that, for any $y \in B_1$,*

$$[\delta_y g]_{\mathcal{C}_l^0(Q_2)} \leq M \quad \text{and} \quad [\delta_y g]_{\mathcal{C}_l^{2+\alpha}(Q_2)} \leq M\|(0, y, 0)\|^2,$$

then there exists some universal constant $\eta \in (0, 1)$ such that, for any $y \in B_1$,

$$\|\delta_y g\|_{\mathcal{C}_l^{\eta}(Q_1)} \lesssim M\|(0, y, 0)\|^3.$$

Proof. Keeping in mind the assumption and Remark 2.5, for fixed $y \in B_1$, we consider the polynomial expansion p_0 of $\delta_y g$ at $z_0 \in Q_2$ with $\deg_{\text{kin}}(p_0) = 2$:

$$p_0(z) = \delta_y g(z_0) + (\partial_t + v_0 \cdot \nabla_x)\delta_y g(z_0)t + \nabla_v \delta_y g(z_0) \cdot v + \tfrac{1}{2} D_v^2 \delta_y g(z_0) v \cdot v$$

for $z := (t, x, v) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d$. For any z such that $z_0 \circ z \in Q_4$, we have

$$|\delta_y g(z_0 \circ z) - p_0(z)| \leq M\|(0, y, 0)\|^2 \|z\|^{2+\alpha}. \tag{C-1}$$

In particular, $p_0(0, y, 0) = \delta_y g(z_0)$, so that, for any $y \in B_1$,

$$\begin{aligned} |\delta_{2y} g(z_0) - 2\delta_y g(z_0)| &= |\delta_y g(z_0 \circ (0, y, 0)) - \delta_y g(z_0)| = |\delta_y g(z_0 \circ (0, y, 0)) - p_0(0, y, 0)| \\ &\leq M\|(0, y, 0)\|^{4+\alpha}. \end{aligned}$$

It then follows that, for any $z_0 \in Q_2$ and for any $k \in \mathbb{N}$ such that $z_0 \circ (0, 2^k y, 0) \in Q_4$,

$$\begin{aligned} |\delta_y g(z_0) - 2^{-k}\delta_{2^k y} g(z_0)| &\leq \sum_{j=1}^{k} 2^{-j} |\delta_{2^j y} g(z_0) - 2\delta_{2^{j-1} y} g(z_0)| \\ &\leq M\|(0, y, 0)\|^{4+\alpha} \sum_{j=1}^{k} 2^{(1+\alpha)j/3} \leq 2M\|(0, y, 0)\|^{4+\alpha} 2^{(1+\alpha)k/3}. \end{aligned} \tag{C-2}$$

Picking $k \in \mathbb{N}$ such that $\|2^{k-1}(0, y, 0)\| \leq 1 < \|2^k(0, y, 0)\|$ and using the assumption yields

$$\begin{aligned} |\delta_y g(z_0)| &\leq 2^{-k}|\delta_{2^k y} g(z_0)| + 2M\|(0, y, 0)\|^{4+\alpha} 2^{(1+\alpha)k/3} \\ &\leq \|\delta_{2^k y} g\|_{\mathcal{C}_l^0(Q_2)} \|(0, y, 0)\|^3 + 4M\|(0, y, 0)\|^3 \leq 5M\|(0, y, 0)\|^3. \end{aligned} \tag{C-3}$$

It remains to show that there exists some constant $\eta > 0$ depending only on α such that

$$|\delta_y g(z_0 \circ z) - \delta_y g(z_0)| \lesssim M\|(0, y, 0)\|^3 \|z\|^\eta. \tag{C-4}$$

By (C-1) and Lemma 2.4, we know that, for any $z_0 \in Q_1$ and $z_0 \circ z \in Q_4$,

$$\begin{aligned}|\delta_y g(z_0 \circ z) - \delta_y g(z_0)| &\le (|(\partial_t + v_0 \cdot \nabla_x)\delta_y g(z_0)| + |D_v^2 \delta_y g(z_0)|)\|z\|^2 \\ &\qquad + |\nabla_v \delta_y g(z_0)|\|z\| + M\|(0, y, 0)\|^2 \|z\|^{2+\alpha} \\ &\lesssim ([\delta_y g]_{\mathcal{C}_l^{2+\alpha}(Q_2)}\|z\| + [\delta_y g]^{1/2}_{\mathcal{C}_l^{2+\alpha}(Q_2)}[\delta_y g]^{1/2}_{\mathcal{C}_l^{0}(Q_2)} + [\delta_y g]_{\mathcal{C}_l^{0}(Q_2)})\|z\| \\ &\qquad + M\|(0, y, 0)\|^2 \|z\|^{2+\alpha}.\end{aligned}$$

If $\|z\| \le \|(0, y, 0)\|$, then combining the above expression with the assumption and (C-3) implies (C-4) with $\eta = \frac{1}{2}$. In particular, if $k \in \mathbb{N}$ such that $\|z\| < \|2^k(0, y, 0)\|$, then we have

$$2^{-k}|\delta_{2^k y} g(z_0 \circ z) - \delta_{2^k y} g_z(z_0)| \lesssim 2^{-k} M\|(0, 2^k y, 0)\|^3 \|z\|^\eta = M\|(0, y, 0)\|^3 \|z\|^\eta. \tag{C-5}$$

Now, if $\|z\| \ge \|(0, y, 0)\|$, applying (C-2) at points z_0 and $z_0 \circ z$, with $k \in \mathbb{N}$ such that $\|2^{k-1}(0, y, 0)\| \le \|z\| < \|2^k(0, y, 0)\|$, yields

$$|\delta_y g(z_0) - 2^{-k}\delta_{2^k y} g(z_0)| \le 4M\|(0, y, 0)\|^3 \|z\|^{1+\alpha}, \tag{C-6}$$

$$|\delta_y g(z_0 \circ z) - 2^{-k}\delta_{2^k y} g(z_0 \circ z)| \le 4M\|(0, y, 0)\|^3 \|z\|^{1+\alpha}. \tag{C-7}$$

Summing up (C-5), (C-6), and (C-7), we arrive at (C-4). □

Following the lines of the above proof and taking into account that $\|g\|_{\mathcal{C}_l^{2+\alpha}(Q_2)} \le M$, one is also able to prove the following result.

Lemma C.2. *If $g \in \mathcal{C}_l^{2+\alpha}(Q_2)$ with $\alpha \in (0, 1)$, then, for any $y \in B_1$, we have*

$$\|\delta_y g\|_{\mathcal{C}_l^{\alpha}(Q_1)} \lesssim \|g\|_{\mathcal{C}_l^{2+\alpha}(Q_2)}\|(0, y, 0)\|^2.$$

Acknowledgements

The authors are grateful to Cyril Imbert for suggesting the question, both François Golse and Cyril Imbert for helpful discussions, and the referees for their careful reading and comments. Zhu's research has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 754362.

References

[Anceschi and Polidoro 2020] F. Anceschi and S. Polidoro, "A survey on the classical theory for Kolmogorov equation", *Matematiche* (*Catania*) **75**:1 (2020), 221–258. MR Zbl

[Anceschi et al. 2019] F. Anceschi, M. Eleuteri, and S. Polidoro, "A geometric statement of the Harnack inequality for a degenerate Kolmogorov equation with rough coefficients", *Commun. Contemp. Math.* **21**:7 (2019), art. id. 1850057. MR Zbl

[Arnold et al. 2001] A. Arnold, P. Markowich, G. Toscani, and A. Unterreiter, "On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker–Planck type equations", *Comm. Partial Differential Equations* **26**:1-2 (2001), 43–100. MR Zbl

[Bardos et al. 1993] C. Bardos, F. Golse, and C. D. Levermore, "Fluid dynamic limits of kinetic equations, II: Convergence proofs for the Boltzmann equation", *Comm. Pure Appl. Math.* **46**:5 (1993), 667–753. MR Zbl

[Bouin et al. 2020] E. Bouin, J. Dolbeault, S. Mischler, C. Mouhot, and C. Schmeiser, "Hypocoercivity without confinement", *Pure Appl. Anal.* **2**:2 (2020), 203–232. MR Zbl

[Cameron et al. 2018] S. Cameron, L. Silvestre, and S. Snelson, "Global a priori estimates for the inhomogeneous Landau equation with moderately soft potentials", *Ann. Inst. H. Poincaré C Anal. Non Linéaire* **35**:3 (2018), 625–642. MR Zbl

[Chandrasekhar 1943] S. Chandrasekhar, "Stochastic problems in physics and astronomy", *Rev. Modern Phys.* **15** (1943), 1–89. MR Zbl

[Chavanis 2008] P. H. Chavanis, "Nonlinear mean field Fokker–Planck equations: application to the chemotaxis of biological populations", *Eur. Phys. J. B* **62**:2 (2008), 179–2008. Zbl

[Daskalopoulos and Kenig 2007] P. Daskalopoulos and C. E. Kenig, *Degenerate diffusions: initial value problems and local regularity theory*, EMS Tracts in Math. **1**, Eur. Math. Soc., Zürich, 2007. MR Zbl

[Degond and Mas-Gallic 1987] P. Degond and S. Mas-Gallic, "Existence of solutions and diffusion approximation for a model Fokker–Planck equation", *Transport Theory Statist. Phys.* **16**:4-6 (1987), 589–636. MR Zbl

[Dolbeault and Li 2018] J. Dolbeault and X. Li, "φ-entropies: convexity, coercivity and hypocoercivity for Fokker–Planck and kinetic Fokker–Planck equations", *Math. Models Methods Appl. Sci.* **28**:13 (2018), 2637–2666. MR Zbl

[Dolbeault et al. 2015] J. Dolbeault, C. Mouhot, and C. Schmeiser, "Hypocoercivity for linear kinetic equations conserving mass", *Trans. Amer. Math. Soc.* **367**:6 (2015), 3807–3828. MR Zbl

[El Ghani and Masmoudi 2010] N. El Ghani and N. Masmoudi, "Diffusion limit of the Vlasov–Poisson–Fokker–Planck system", *Commun. Math. Sci.* **8**:2 (2010), 463–479. MR Zbl

[Esposito et al. 2013] R. Esposito, Y. Guo, C. Kim, and R. Marra, "Non-isothermal boundary in the Boltzmann theory and Fourier law", *Comm. Math. Phys.* **323**:1 (2013), 177–239. MR Zbl

[Gilbarg and Trudinger 2001] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, Springer, 2001. MR Zbl

[Golse et al. 2019] F. Golse, C. Imbert, C. Mouhot, and A. F. Vasseur, "Harnack inequality for kinetic Fokker–Planck equations with rough coefficients and application to the Landau equation", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **19**:1 (2019), 253–295. MR Zbl

[Guo 2002] Y. Guo, "The Landau equation in a periodic box", *Comm. Math. Phys.* **231**:3 (2002), 391–434. MR Zbl

[Henderson and Snelson 2020] C. Henderson and S. Snelson, "C^∞ smoothing for weak solutions of the inhomogeneous Landau equation", *Arch. Ration. Mech. Anal.* **236**:1 (2020), 113–143. MR Zbl

[Henderson et al. 2019] C. Henderson, S. Snelson, and A. Tarfulea, "Local existence, lower mass bounds, and a new continuation criterion for the Landau equation", *J. Differential Equations* **266**:2-3 (2019), 1536–1577. MR Zbl

[Henderson et al. 2020a] C. Henderson, S. Snelson, and A. Tarfulea, "Local solutions of the Landau equation with rough, slowly decaying initial data", *Ann. Inst. H. Poincaré C Anal. Non Linéaire* **37**:6 (2020), 1345–1377. MR Zbl

[Henderson et al. 2020b] C. Henderson, S. Snelson, and A. Tarfulea, "Self-generating lower bounds and continuation for the Boltzmann equation", *Calc. Var. Partial Differential Equations* **59**:6 (2020), art. id. 191. MR Zbl

[Hérau 2018] F. Hérau, "Introduction to hypocoercive methods and applications for simple linear inhomogeneous kinetic models", pp. 119–147 in *Lectures on the analysis of nonlinear partial differential equations, V* (Beijing, 2014-2015), edited by J.-Y. Chemin et al., Morningside Lect. Math. **5**, Int. Press, Somerville, MA, 2018. MR Zbl

[Imbert and Mouhot 2021] C. Imbert and C. Mouhot, "The Schauder estimate in kinetic theory with application to a toy nonlinear model", *Ann. H. Lebesgue* **4** (2021), 369–405. MR Zbl

[Imbert and Silvestre 2021] C. Imbert and L. Silvestre, "The Schauder estimate for kinetic integral equations", *Anal. PDE* **14**:1 (2021), 171–204. MR Zbl

[Imbert and Silvestre 2022] C. Imbert and L. E. Silvestre, "Global regularity estimates for the Boltzmann equation without cut-off", *J. Amer. Math. Soc.* **35**:3 (2022), 625–703. MR Zbl

[Kato 1975] T. Kato, "The Cauchy problem for quasi-linear symmetric hyperbolic systems", *Arch. Ration. Mech. Anal.* **58**:3 (1975), 181–205. MR Zbl

[Kim et al. 2020] J. Kim, Y. Guo, and H. J. Hwang, "An L^2 to L^∞ framework for the Landau equation", *Peking Math. J.* **3**:2 (2020), 131–202. MR Zbl

[Kolmogoroff 1934] A. Kolmogoroff, "Zufällige Bewegungen (zur Theorie der Brownschen Bewegung)", *Ann. of Math.* (2) **35**:1 (1934), 116–117. MR Zbl

[Liao et al. 2018] J. Liao, Q. Wang, and X. Yang, "Global existence and decay rates of the solutions near Maxwellian for non-linear Fokker–Planck equations", *J. Stat. Phys.* **173**:1 (2018), 222–241. MR Zbl

[Manfredini 1997] M. Manfredini, "The Dirichlet problem for a class of ultraparabolic equations", *Adv. Differential Equations* **2**:5 (1997), 831–866. MR Zbl

[Markou 2017] I. Markou, "Hydrodynamic limit for a Fokker–Planck equation with coefficients in Sobolev spaces", *Netw. Heterog. Media* **12**:4 (2017), 683–705. MR Zbl

[Moser 1964] J. Moser, "A Harnack inequality for parabolic differential equations", *Comm. Pure Appl. Math.* **17** (1964), 101–134. MR Zbl

[Mouhot 2018] C. Mouhot, "De Giorgi–Nash–Moser and Hörmander theories: new interplays", pp. 2467–2493 in *Proceedings of the International Congress of Mathematicians*, *III* (Rio de Janeiro, 2018), edited by B. Sirakov et al., World Sci., Hackensack, NJ, 2018. MR Zbl

[Nash 1958] J. Nash, "Continuity of solutions of parabolic and elliptic equations", *Amer. J. Math.* **80** (1958), 931–954. MR Zbl

[Saint-Raymond 2009] L. Saint-Raymond, *Hydrodynamic limits of the Boltzmann equation*, Lecture Notes in Math. **1971**, Springer, 2009. MR Zbl

[Tsallis 1988] C. Tsallis, "Possible generalization of Boltzmann–Gibbs statistics", *J. Stat. Phys.* **52**:1-2 (1988), 479–487. MR Zbl

[Tsallis 2009] C. Tsallis, *Introduction to nonextensive statistical mechanics*: *approaching a complex world*, Springer, 2009. MR Zbl

[Villani 2002] C. Villani, "A review of mathematical topics in collisional kinetic theory", pp. 71–305 in *Handbook of mathematical fluid dynamics*, *I*, edited by S. Friedlander and D. Serre, North-Holland, Amsterdam, 2002. MR Zbl

[Villani 2009] C. Villani, *Hypocoercivity*, Mem. Amer. Math. Soc. **950**, Amer. Math. Soc., Providence, RI, 2009. MR Zbl

[Yau 1991] H.-T. Yau, "Relative entropy and hydrodynamics of Ginzburg–Landau models", *Lett. Math. Phys.* **22**:1 (1991), 63–80. MR Zbl

[Zhu 2021] Y. Zhu, "Velocity averaging and Hölder regularity for kinetic Fokker–Planck equations with general transport operators and rough coefficients", *SIAM J. Math. Anal.* **53**:3 (2021), 2746–2775. MR Zbl

Received 25 Feb 2021. Revised 25 May 2022. Accepted 11 Jul 2022.

FRANCESCA ANCESCHI: f.anceschi@staff.univpm.it
Dipartimento di Ingegneria Industriale e Scienze Matematiche, Università Politecnica delle Marche, Ancona, Italy

YUZHE ZHU: yuzhe.zhu@ens.fr
Département de mathématiques et applications, École normale supérieure - PSL, Paris, France

ANALYSIS AND PDE
Vol. 17 (2024), No. 2, pp. 421–454
DOI: 10.2140/apde.2024.17.421

STRICHARTZ INEQUALITIES WITH WHITE NOISE POTENTIAL ON COMPACT SURFACES

ANTOINE MOUZARD AND IMMANUEL ZACHHUBER

We prove Strichartz inequalities for the Schrödinger equation and the wave equation with multiplicative noise on a two-dimensional manifold. This relies on the Anderson Hamiltonian described using high-order paracontrolled calculus. As an application, it gives a low-regularity solution theory for the associated nonlinear equations.

Introduction

Enormous progress has been made in the last decade after Hairer [21] introduced his theory of regularity structures and the theory of paracontrolled distributions due to Gubinelli, Imkeller, and Perkowski [17] in the study of singular stochastic PDEs. A particular approach developed recently is the construction of a random stochastic operator to investigate associated PDEs. The first paper on this, by Allez and Chouk [1], dealt with the continuum Anderson Hamiltonian, hereafter simply called the Anderson Hamiltonian. They used the latter theory to make sense of the operator on the two-dimensional torus, formally

$$H = -\Delta + \xi,$$

where ξ is spatial white noise whose spatial regularity is just below -1, that is, the centered random field with formal covariance

$$\mathbb{E}[\xi(x)\xi(y)] = \delta_0(x-y).$$

In the particular case of the torus, it can be constructed as the random Fourier series

$$\xi(x) = \sum_{n\in\mathbb{Z}^2} \xi_n e^{in\cdot x},$$

with $(\xi_n)_{n\in\mathbb{Z}^2}$ independent and identically distributed standard Gaussian random variables. In general, the white noise is an isometry from $L^2(M)$ to $L^2(\Omega)$ the space of random variable with finite variance. Afterwards this approach was extended to the three-dimensional torus and somewhat reformulated by Gubinelli, Ugurcan and Zachhuber [19] and by Labbé [23] who used regularity structures and dealt with both periodic and Dirichlet boundary conditions. Finally, the construction was extended by Mouzard [27] to the case of two-dimensional manifold using high-order paracontrolled calculus.

Naturally, substantial progress was also made in the field of singular dispersive SPDEs following the paper [15] by Debussche and Weber on the cubic multiplicative stochastic Schrödinger equation and [18] by Gubinelli, Koch and Oh on the cubic additive stochastic wave equation. Since the powerful

MSC2020: 35J10, 58J05, 60H25.
Keywords: Anderson Hamiltonian, paracontrolled calculus, white noise, Schrödinger operator, Strichartz inequalities.

tools from singular SPDEs are only directly applicable to parabolic and elliptic SPDEs, these initial papers were in a not-so-singular regime, the former using an exponential transform to remove the most singular term and the latter using a "Da Prato–Debussche trick" to do the same. Gubinelli, Ugurcan and Zachhuber [19] proved some sharpened results on the multiplicative Schrödinger equation and its wave analogue by reframing it in relation to the Anderson Hamiltonian, as well as extending the results to dimension 3. Moreover, Tzvetkov and Visciglia [33] extended the results of [15] to a larger range of power nonlinearities; see also [32]. For the nonlinear wave equation with additive noise, let us mention here the follow-up paper by Gubinelli, Koch and Oh [20] in three dimensions with quadratic nonlinearity and the paper [28] by Oh, Robert and Tzvetkov which extends the results of [18] to the case of two-dimensional surfaces and is thus salient for the current paper.

Let us also mention a related area of research whose aim is to solve deterministic dispersive PDEs with random initial conditions with low regularity. The study of this, which is intimately related to the analysis of invariant measures for dispersive PDEs, goes back to the seminal work of Lebowitz, Rose and Speer [24]. A series of works by Bourgain followed; let us mention here [10] where a renormalisation procedure similar to the current case appears but for a different reason. See also the work [11] by Burq and Tzvetkov, which deals with singular random initial condition for which they obtain well-posedness results for the cubic nonlinear wave equation on a compact manifold.

In this paper, we prove Strichartz inequalities for the Schrödinger and wave equations with white noise potential on compact surfaces. In a nutshell, Strichartz inequalities leverage dispersion in order to allow us to trade integrability in time for integrability in space; see Section 2 for a more detailed introduction and [34] where this kind of approach appeared for the Anderson Hamiltonian. Moreover, we show how this provides local well-posedness for the associated nonlinear equations in a low-regularity regimes. As for the deterministic case, the Strichartz estimates obtained depend whether the manifold has a boundary or not and are improved in the flat case of the torus. By Strichartz inequalities, we generally refer to space-time bounds on the propagators of Schrödinger and wave equations where the results on integrability are strictly better than what one gets from the Sobolev embedding so — for definiteness we consider the Schrödinger case — a bound like

$$\|e^{itH}u\|_{L^p(I,L^q)} \lesssim \|u\|_{\mathcal{H}^\alpha},$$

with $p \in [1, \infty]$, $q > 2d/(d-2\alpha)$, where d denotes the dimension and $I \subset \mathbb{R}$ is an interval. The overall approach to the Schrödinger group associated to H we follow is similar to the one in [34], where such Strichartz estimates were shown for the Anderson Hamiltonian on the two and three-dimensional torus. However, one gets sharper results in the particular case of flat geometry due to the fact that one has stronger classical Strichartz inequalities available. In the more general setting of a Riemannian compact manifold, we work with a result due to Burq, Gerard and Tzvetkov [12], which has been extended to the case with boundary by Blair, Smith and Sogge [8]. These results can be thought of as quantifying the statement "finite frequencies travel at finite speeds — in (frequency-dependent) short time the evolution is morally on flat space". Let us also mention at this point the recent work by Huang and Sogge [22] which deals with a similar setting; however, their notion of singular potential refers to low integrability while in our case singular refers rather to potentials with low regularity.

For the case of Strichartz estimates for the wave equation related to H, we follow the approach introduced by Burq, Lebeau and Planchon [13] on domains with boundary. The main idea, which is why this approach is applicable, is that all that is required is that the operator driving the wave equation satisfies some growth condition on the L^q bounds on its eigenfunctions and one knows about the asymptotics of the eigenvalues, in their case the Laplace with boundary conditions. Since a Weyl law for H was obtained by Mouzard [27] and our result for the Schrödinger equation gives us a suitable L^q bound on the eigenfunctions of H, their approach turns out to be enough to prove Strichartz estimates that beat the Sobolev embedding. Overall this approach seems somewhat crude and we assume there to be sharper bounds possible, whereas in the Schrödinger case, our result is the same as the one without noise obtained in [12] worsened only by an arbitrarily small regularity loss. The state of the art of Strichartz estimates for wave equations on manifolds with boundary is the paper [7], the case of manifolds without boundary being comparable to the Strichartz estimates on Euclidean space because of the finite speed of propagation. In particular, the bounds obtained on the spectral projectors of the Anderson Hamiltonian are new and of interest themselves.

The second objective of this paper is to use the Strichartz inequalities obtained to prove local well-posedness for the associated defocussing nonlinear equations, also known as cubic multiplicative stochastic Schrödinger and wave equations. This will be done using fairly straightforward contraction arguments for which the Strichartz estimates will be crucial.

We conclude the introduction by a brief outline of the construction of the Anderson Hamiltonian; see [27] for the details. It is formally given by

$$H = -\Delta + \xi,$$

where ξ is the space white noise and belongs to $\mathcal{C}^{\alpha-2}$ for any $\alpha < 1$, where $\mathcal{C}^{\beta}$ denotes the Hölder–Besov spaces recalled in Section 1A. The noise is only a distribution, rough almost surely everywhere as opposed to potential with a localised singularity; hence Hu is well-defined for $u \in C^\infty$ but does not belong to L^2. The nature of the noise makes the naive candidate for the domain of H, that is, the closure of

$$\{u \in C^\infty;\ Hu \in L^2\},$$

with respect to the domain norm, unviable. This is precisely where the paracontrolled calculus comes into play, one can construct a random space $\mathcal{D}_\Xi \subset L^2$ such that almost surely

$$u \in \mathcal{D}_\Xi \quad \Longrightarrow \quad Hu \in L^2.$$

Here $\Xi = (\xi, \Delta^{-1}\xi \cdot \xi)$ refers to the enhanced noise; see [27] for its construction. The domain $\mathcal{D}_\Xi$ consists of functions $u \in L^2$ paracontrolled by noise-dependent functions X_1, X_2 of the form

$$u = \widetilde{\mathsf{P}}_u X_1 + \widetilde{\mathsf{P}}_u X_2 + u^\sharp,$$

with a remainder $u^\sharp \in \mathcal{H}^2$ and the $\widetilde{\mathsf{P}}_u X_i$ are terms which are dominated by X_i in terms of regularity. In particular, smooth functions do not belong to the domain in this peculiar setting. The singularity of the product is dealt with through a renormalisation procedure which corresponds to the construction of the singular term $X \cdot \xi$, where $\Delta X = \xi$. Given a regularisation ξ_ε of the noise, the product $X_\varepsilon \cdot \xi_\varepsilon$ diverges

but one gets a well-defined function after the subtraction of the diverging quantity

$$c_\varepsilon := \mathbb{E}[X_\varepsilon \cdot \xi_\varepsilon].$$

The analysis of the operator is then done with

$$\Delta^{-1}\xi \cdot \xi := \lim_{\varepsilon \to 0} (\Delta^{-1}\xi_\varepsilon \cdot \xi_\varepsilon - c_\varepsilon),$$

the Wick product, and the Anderson Hamiltonian corresponds to the limit of the family of operators

$$H_\varepsilon = -\Delta + \xi_\varepsilon - c_\varepsilon$$

as ε goes to 0. In the case of the torus, c_ε is a constant due to the invariance by translation of the noise and diverges as $|\log \varepsilon|$; for details, see Section 2.1 of [27]. Note that the operator Δ is not invertible and Δ^{-1} has to be interpreted as a parametrix, that is, an inverse up to a smooth term.

While the Anderson Hamiltonian can be interpreted as the electric Laplacian $-\Delta + V$ with electric field $V = \xi$, one can consider as an analogy the magnetic Laplacian with magnetic field $B = \xi$ space white noise. This is the content of [26], where Morin and Mouzard construct

$$H = (i\partial_1 + A_1)^2 + (i\partial_2 + A_2)^2$$

on the two-dimensional torus with magnetic potential $A = (A_1, A_2)$ the Lorentz gauge associated to the white noise magnetic field. Its study is also motivated by supraconductivity where H plays a specific role in the third critical field of Ginzburg–Landau theory. In particular, the first results, such as self-adjointness, discrete spectrum and the Weyl law, hold as for the Anderson Hamiltonian, while differences are expected to appear when one looks at finer properties. Our proof for the Strichartz inequalities for the Schrödinger group associated to the Anderson Hamiltonian, which is perturbative in nature, can directly be adapted to obtain a similar result for the random magnetic Laplacian with white noise magnetic field. For Strichartz estimates for the magnetic Schrödinger equation in the deterministic case, see for example [14].

Organisation of the paper. In Section 1, we give the context for the Strichartz inequalities on manifolds in the case of the Schrödinger equations and the heat semigroup paracontrolled calculus on manifolds; see respectively [12] and [27]. We conclude by recalling the construction of the Anderson Hamiltonian and provide new results needed in the following. In Section 2, we provide Strichartz inequalities for the Schrödinger group associated to the Anderson Hamiltonian and show how this gives local well-posedness for the stochastic cubic nonlinear Schrödinger equation with multiplicative white noise. In Section 3, we use the result on the Schrödinger group to get new bounds on the eigenvalues of the Anderson Hamiltonian and use it to prove Strichartz inequalities for the wave propagator together with the Weyl-type law. Finally, we also show how this gives local well-posedness for the stochastic cubic nonlinear wave equation with multiplicative white noise and give details for the particular case of the torus where one gets improved bounds.

1. Preliminaries

1A. ***Strichartz inequalities on manifolds.*** On the torus, regularity of distributions can be measured using the Littlewood–Paley decomposition. On a manifold, one has an analogue decomposition using

the eigenfunctions of the Laplace–Beltrami operator Δ as a generalisation of Fourier theory; see for example Section 2 in [29] by Oh, Robert, Tzvetkov and Wang. Let (M, g) be a two-dimensional compact Riemannian manifold without boundary or with boundary and Dirichlet boundary conditions. In this framework, the Laplace–Beltrami operator $-\Delta$ is a self-adjoint positive operator with discrete spectrum

$$0 \le \lambda_1 < \lambda_2 \le \lambda_3 \le \cdots,$$

with the associated normalised eigenfunctions $(\varphi_n)_{n\ge1}$ belonging to $C^\infty(M)$. In the case where M has no boundary, we have $\lambda_1 = 0$ and $\varphi_1 = \mathrm{Vol}(M)^{-1/2}$ constant. Furthermore, the Weyl law gives the asymptotics

$$\lim_{n\to\infty} \frac{\lambda_n}{n} = \frac{\mathrm{Vol}(M)}{4\pi}.$$

The basis $(\varphi_n)_{n\ge1}$ of L^2 gives the decomposition

$$u = \sum_{n\ge1} \langle u, \varphi_n\rangle \varphi_n$$

for any distribution $u \in \mathcal{D}'(M)$. On the torus, this gives the Littlewood–Paley decomposition of u, where the regularity is measured by the asymptotics behaviour of $\sum_{\lambda_k\sim 2^n} \langle u, \varphi_k\rangle$. On a manifold M, this is done with

$$\Delta_n := \psi(-2^{-2(n+1)}\Delta) - \psi(-2^{-2n}\Delta)$$

for $n \ge 0$ and

$$\Delta_{-1} := \psi(-\Delta),$$

with $\psi \in C_0^\infty(\mathbb{R})$ a nonnegative function with $\mathrm{supp}(\psi) \subset [-1, 1]$ and $\psi = 1$ on $\left[-\frac{1}{2}, \frac{1}{2}\right]$. Recall that for any function $\psi \in L^\infty(\mathbb{R})$, the operator $\psi(\Delta)$ is defined as

$$\psi(\Delta)u = \sum_{n\ge1} \psi(\lambda_n)\langle u, \varphi_n\rangle \varphi_n$$

and this yields a bounded operator from $L^2(M)$ to itself. In this setting, Besov spaces are defined for $\alpha \in \mathbb{R}$ and $p, q \in [1, \infty]$ as

$$\mathcal{B}^\alpha_{p,q} := \{u \in \mathcal{D}'(M) : \|u\|_{\mathcal{B}^\alpha_{p,q}} < \infty\},$$

where

$$\|u\|_{\mathcal{B}^\alpha_{p,q}} := \left(\|\Delta_{-1}u\|^q_{L^p(M)} + \sum_{n\ge0} 2^{\alpha q}\|\Delta_n u\|^q_{L^p(M)}\right)^{\frac{1}{q}}.$$

In the particular case $p = q = \infty$, these spaces are called *Hölder–Besov spaces* and we write

$$B^\alpha_{\infty,\infty} = \mathcal{C}^\alpha.$$

The case $p = q = 2$ corresponds to Sobolev spaces and we have

$$\|u\|^2_{\mathcal{H}^\alpha} = \|\Delta_{-1}u\|^2_{L^2(M)} + \sum_{n\ge0} 2^{2n\alpha}\|\varphi(2^{-2n}\Delta)u\|^2_{L^2(M)},$$

where $\varphi(x) := \psi(-x^2) - \psi(-x)$. Burq, Gérard and Tzvetkov [12] proved, in the case where M has no boundary, the bound

$$\|f\|_{L^q(M)} \lesssim \|\psi(-\Delta)f\|_{L^q(M)} + \left(\sum_{n\geq 0} \|\varphi(2^{-2n}\Delta)f\|^2_{L^q(M)}\right)^{\frac{1}{2}},$$

using that for $\lambda \in \mathbb{R}$ we have

$$\psi(-\lambda) + \sum_{n\geq 0} \varphi(2^{-2n}\lambda) = 1.$$

Applying this to the Schrödinger group, they obtain

$$\|e^{it\Delta}v\|_{L^p([0,1],L^q)} \lesssim \|\psi(-\Delta)v\|_{L^q(M)} + \left\|\left(\sum_{k\geq 0} \|e^{it\Delta}\varphi(2^{-2k}\Delta)v\|^2_{L^q(M)}\right)^{\frac{1}{2}}\right\|_{L^p([0,1])};$$

hence one only needs a bound for spectrally localised data. This is proved using semiclassical analysis with the use of the WKB expansion; see Proposition 2.9 from [12] which gives

$$\left(\int_J \|e^{it\Delta}\varphi(h^2\Delta)v\|^p_{L^q(M)}\,\mathrm{d}t\right)^{\frac{1}{p}} \lesssim \|v\|_{L^2(M)} \tag{1}$$

for J an interval of small enough length proportional to $h \in (0, 1)$. Moreover, a well-known trick is to slice up the time interval into small pieces; this will be useful later. The previous bounds with the Minkowski inequality lead to

$$\|e^{it\Delta}v\|_{L^p([0,1],L^q)} \lesssim \|v\|_{L^2(M)} + \left(\sum_{k\geq 0} 2^{\frac{2k}{p}} \|\varphi(2^{-2k}\Delta)v\|^2_{L^2(M)}\right)^{\frac{1}{2}} \lesssim \|v\|_{\mathcal{H}^{1/p}}.$$

This yields the following theorem.

Theorem 1.1. *Let $p \geq 2$ and $q < \infty$ such that*

$$\frac{2}{p} + \frac{2}{q} = 1.$$

Then

$$\|e^{it\Delta}u\|_{L^p([0,1],L^q)} \lesssim \|u\|_{\mathcal{H}^{1/p}}.$$

While this result is optimal on general surfaces in the case $p = 2$, this can be improved in the flat case of the torus. In fact, the first result concerning Strichartz inequalities for the Schrödinger equation on a compact manifold was obtain by Bourgain [10] on the flat torus. In the case of the Anderson Hamiltonian on a compact surface without boundary, we obtain the same result as Theorem 1.1 with an arbitrarily small loss of regularity; this is the content of Section 2. In the case of a surface with boundary, the following result was obtained by Blair, Smith and Sogge [8].

Theorem 1.2. *Let M be a surface with boundary. Let $p \in (3, \infty]$ and $q \in [2, \infty)$ such that*

$$\frac{3}{p} + \frac{2}{q} = 1.$$

Then

$$\|e^{it\Delta}u\|_{L^p([0,1],L^q)} \lesssim \|u\|_{\mathcal{H}^{2/p}}$$

and

$$\left(\int_J \|e^{it\Delta}\varphi(h^2\Delta)v\|_{L^q(M)}^p \,\mathrm{d}t\right)^{\frac{1}{p}} \lesssim h^{-\frac{1}{p}}\|\varphi(h^2\Delta)v\|_{L^2(M)} \tag{2}$$

for J an interval of small enough length proportional to $h \in (0, 1)$.

We end this section with two classical results that will be needed in this paper. First, one still has the Bernstein lemma with the Littlewood–Paley decomposition associated to the Laplace–Beltrami operator.

Lemma 1.3. *Let $g : M \to \mathbb{R}$ be a function which has spectral support in an interval $[a, b]$ with $0 < a < b < \infty$. Then for any $\alpha, \beta \in \mathbb{R}$ we have the following bounds which are the analogue of Bernstein's inequality on Euclidean space*:

$$\|g\|_{\mathcal{H}^\alpha} \lesssim \max(b^{\alpha-\beta}, a^{\alpha-\beta})\|g\|_{\mathcal{H}^\beta},$$
$$\|g\|_{\mathcal{H}^\alpha} \gtrsim \min(b^{\alpha-\beta}, a^{\alpha-\beta})\|g\|_{\mathcal{H}^\beta}.$$

The former estimate still holds in the case where $a = 0$ and $\alpha > \beta$. We will chiefly apply these bounds to Littlewood–Paley projectors where $b = 2a = 2^j$ for $j \in \mathbb{N}$.

Proof. The condition on g means that

$$g = \sum_{\lambda_k \in [a,b]} (g, \phi_k)\phi_k$$

and we have

$$\|g\|_{\mathcal{H}^\alpha}^2 = \sum_{\lambda_k \in [a,b]} (g, \phi_k)^2 \lambda_k^{2\alpha}.$$

The upper bounds follow directly with

$$\lambda_k^{2\alpha} = \lambda_k^{2\beta}\lambda_k^{2(\alpha-\beta)} \le \lambda^{2\beta}\max(b^{2(\alpha-\beta)}, a^{2(\alpha-\beta)})$$

and analogously for the lower bounds. □

The space $\mathcal{H}^\sigma$ is an algebra only for σ large enough depending on the dimension; this can be seen with the following proposition and the Sobolev embedding. These types of estimates are important for the dispersive equations with cubic nonlinearity considered here.

Lemma 1.4. *Let $\sigma \ge 0$. The space $\mathcal{H}^\sigma \cap L^\infty$ is an algebra and one has the bound*

$$\|f \cdot g\|_{\mathcal{H}^\sigma} \lesssim \|f\|_{\mathcal{H}^\sigma}\|g\|_{L^\infty} + \|g\|_{\mathcal{H}^\sigma}\|f\|_{L^\infty}.$$

1B. ***Basics on paracontrolled calculus.*** On the torus, the Littlewood–Paley decomposition can also be used to study ill-defined products. Recall that for $u \in \mathcal{D}'(\mathbb{T}^2)$, it is given by

$$u = \sum_{n \ge 0} \Delta_n u,$$

where each $\Delta_n u$ is smooth and localised in frequency in an annulus of radius 2^n for $n \ge 1$, while the Fourier transform of $\Delta_0 u$ is contained in a ball around the origin. Given two distributions $u, v \in \mathcal{D}'(\mathbb{T}^2)$,

the product is formally given by

$$\begin{aligned} u \cdot v &= \sum_{n,m \geq 0} \Delta_n u \cdot \Delta_m v \\ &= \sum_{n \lesssim m} \Delta_n u \cdot \Delta_m v + \sum_{n \sim m} \Delta_n u \cdot \Delta_m v + \sum_{m \lesssim n} \Delta_n u \cdot \Delta_m v \\ &=: P_u v + \Pi(u, v) + P_v u. \end{aligned}$$

The term $P_u v$ is called the paraproduct of v by u and is always well-defined, while the potential singularity is encoded in the resonant term $\Pi(u, v)$. Using this decomposition, Gubinelli, Imkeller and Perkowski introduced the notion of paracontrolled calculus to develop a solution theory for singular stochastic PDEs in their seminal work [17]; this correspond to Bony's paraproduct from [9] in this flat case. On a manifold, an alternative paracontrolled calculus was developed by Bailleul and Bernicot [4] based on the heat semigroup. Instead of Littlewood–Paley, which is discrete decomposition, they used the Calderón formula

$$u = \lim_{t \to 0} P_t u = \int_0^1 Q_t u \, \frac{\mathrm{d}t}{t} + P_1 u$$

for $u \in \mathcal{D}'(M)$, with P_t the heat semigroup and $Q_t = -t\partial_t P_t$. Using Gaussian upper bounds for the heat kernel and its derivatives, this defines a continuous analogue of the Littlewood–Paley decomposition where $\sqrt{t} \simeq 2^{-n}$ and yields descriptions of Besov–Hölder and Sobolev spaces for scalar fields on manifolds. This can be used to construct a paraproduct P and a resonant product Π, such that

$$u \cdot v = \mathsf{P}_u v + \Pi(u, v) + \mathsf{P}_v u,$$

that verify the same important properties of their Fourier analogue P and Π. It was later extended to a higher-order paracontrolled calculus by Bailleul, Bernicot and Frey [6] to deal with rougher noise than the initial work of Gubinelli, Imkeller and Perkowski, again in a general geometric framework. While these different works dealt with parabolic PDEs, the paracontrolled calculus can be used to study singular random operators. It was first used by Allez and Chouk [1] to study the Anderson Hamiltonian

$$H = -\Delta + \xi$$

on the two-dimensional torus. The same operator was constructed on the torus by Gubinelli, Ugurcan and Zachhuber [19] on $\mathbb{T}^d$ with $d \in \{2, 3\}$ to solve associated evolution PDEs. Labbé [23] also constructed the operator in two and three dimensions with different boundary conditions using regularity structures. Finally, Mouzard [27] used the heat semigroup paracontrolled calculus to construct the operator on a two-dimensional manifold and obtained an almost sure Weyl-type law. Note that [27] is self-contained and is a gentle introduction to the paracontrolled calculus on manifolds in the spatial framework. For another example of singular random operators, see [26], where Morin and Mouzard construct the magnetic Laplacian with white noise magnetic field on $\mathbb{T}^2$. With this work, we show that this approach is also well-suited for the study of dispersive PDEs.

The heat semigroup paracontrolled calculus is a theory to study PDEs with singular products on manifolds. Given a suitable family $(V_i)_{1 \leq i \leq d}$ of first-order differential operators, one can construct a

paraproduct P and a resonant term Π based on the heat semigroup associated to

$$L := -\sum_{i=1}^{d} V_i^2.$$

We briefly outline this construction here; see [5; 6] in the parabolic space-time setting and [27] in the space setting for the details. In particular, the Laplace–Beltrami operator on a manifold can be written in this form; see for example Stroock's book [31]. For any distribution $u \in \mathcal{D}'(M)$, the heat semigroup

$$P_t u = e^{-tL} u$$

provides a smooth approximation as t goes to 0. Introducing its derivative

$$Q_t := -t\partial_t P_t,$$

one gets an analogue of the Littlewood–Paley decomposition as explained before. While the Δ_n's enjoy proper orthogonal relation in the sense that $\Delta_n \Delta_m$ is equal to zero for $|n-m| > 1$, we only have in this continuous framework

$$Q_t Q_s = \frac{ts}{(t+s)^2}((t+s)L)^2 e^{-tL},$$

which is indeed small if $s \ll t$ or $t \ll s$. For a given integer $b \in \mathbb{N}^*$, let

$$Q_t^{(b)} := (tL)^b e^{-tL}.$$

Then

$$Q_t^{(b)} Q_s^{(b)} = \left(\frac{ts}{(t+s)^2}\right)^b Q_{t+s}^{(b)};$$

hence the parameter b encodes a cancellation property between different scales t and s. Furthermore, we have

$$\int_0^1 Q_t^{(b)} u \, \frac{\mathrm{d}t}{t} = \lim_{t\to 0} P_t^{(b)} u = u,$$

where $P_0^{(b)} = \mathrm{Id}$ and

$$-t\partial_t P_t^{(b)} = Q_t^{(b)}.$$

In particular, we have $P_t^{(b)} = p_b(tL)e^{-tL}$, with p_b a polynomial of degree $b-1$ such that $p_b(0) = 1$. Denote by StGC^a the family of operators $(Q_t)_{t\in[0,1]}$ of the form

$$Q_t = (t^{\frac{|I|}{2}} V_I)(tL)^j e^{-tL},$$

with $a = |I| + 2j$ and GC^a the operator with kernel satisfying Gaussian upper bounds with cancellation of order a; see Section 1.2 [27] for the definitions. We have

$$\begin{aligned} u\cdot v &= \lim_{t\to 0} P_t^{(b)}(P_t^{(b)} u \cdot P_t^{(b)} v) \\ &= \int_0^1 Q_t^{(b)}(P_t^{(b)} u \cdot P_t^{(b)} v)\frac{\mathrm{d}t}{t} + \int_0^1 P_t^{(b)}(Q_t^{(b)} u \cdot P_t^{(b)} v)\frac{\mathrm{d}t}{t} + \int_0^1 P_t^{(b)}(P_t^{(b)} u \cdot Q_t^{(b)} v)\frac{\mathrm{d}t}{t} + P_1^{(b)}(P_1^{(b)} u \cdot P_1^{(b)} v). \end{aligned}$$

After a number of integrations by parts, we get

$$u \cdot v = \mathsf{P}_u v + \Pi(u, v) + \mathsf{P}_v u,$$

where $\mathsf{P}_u v$ is a linear combination of terms of the form

$$\int_0^1 Q_t^{1\bullet}(P_t u \cdot Q_t^2 v)\,\frac{\mathrm{d}t}{t}$$

and $\Pi(u, v)$ of

$$\int_0^1 P_t^{\bullet}(Q_t^1 u \cdot Q_t^2 v)\,\frac{\mathrm{d}t}{t},$$

where $Q^1, Q^2 \in \mathsf{StGC}^{b/2}$ and $P \in \mathsf{StGC}^{[0,b]}$. In general, the operator V_i's do not commute, hence the need for the notation

$$Q_t^{\bullet} = \big((t^{\frac{|I|}{2}} V_I)(tL)^j e^{-tL}\big)^{\bullet} := (tL)^j e^{-tL}(t^{\frac{|I|}{2}} V_I),$$

which comes from the integration by parts. For simplicity we state most of the results of this section in Besov–Hölder spaces. The following proposition gives the continuity estimates of the paraproduct and the resonant term between Sobolev and Hölder–Besov functions, but they hold in the same way by replacing all the Sobolev spaces by Besov–Hölder spaces.

Proposition 1.5. *Let $\alpha, \beta \in (-2b, 2b)$ be regularity exponents.*

- *If $\alpha \geq 0$, then $(f, g) \mapsto \mathsf{P}_f g$ is continuous from $\mathcal{H}^\alpha \times \mathcal{C}^\beta$ to $\mathcal{H}^\beta$.*
- *If $\alpha < 0$, then $(f, g) \mapsto \mathsf{P}_f g$ is continuous from $\mathcal{H}^\alpha \times \mathcal{C}^\beta$ to $\mathcal{H}^{\alpha+\beta}$.*
- *If $\alpha + \beta > 0$, then $(f, g) \mapsto \Pi(f, g)$ is continuous from $\mathcal{H}^\alpha \times \mathcal{C}^\beta$ to $\mathcal{H}^{\alpha+\beta}$.*

While P and Π are tools to describe products, the intertwined paraproduct $\widetilde{\mathsf{P}}$ naturally appears when formulating solutions to PDEs. The intertwining relation is

$$L \circ \widetilde{\mathsf{P}} = \mathsf{P} \circ L;$$

hence $\widetilde{\mathsf{P}}_u v$ is given as a linear combination of

$$\begin{aligned}\int_0^1 L^{-1} Q_t^{1\bullet}(P_t u \cdot Q_t^2 L v)\,\frac{\mathrm{d}t}{t} &\sim \int_0^1 (tL)^{-1} Q_t^{1\bullet}(P_t u \cdot Q_t^2 (tL) v)\,\frac{\mathrm{d}t}{t} \\ &\sim \int_0^1 \widetilde{Q}_t^{1\bullet}(P_t u \cdot \widetilde{Q}_t^2 v)\,\frac{\mathrm{d}t}{t},\end{aligned}$$

with $\widetilde{Q}^1 \in \mathsf{StGC}^{b/2-2}$ and $\widetilde{Q}^2 \in \mathsf{StGC}^{b/2+2}$. The operator L is not invertible and everything here is done up to a smooth error term; see [27]. In particular, $\widetilde{\mathsf{P}}$ has the same structure as P for large b and satisfies the same continuity estimates as P. Intuitively, the intertwined operator $\widetilde{\mathsf{P}}$ describes solutions to elliptic PDEs of the form

$$Lu = u\xi = \mathsf{P}_u \xi + \mathsf{P}_\xi u + \Pi(u, \xi),$$

which can be rewritten

$$u = \widetilde{\mathsf{P}}_u(L^{-1}\xi) + u^\sharp;$$

hence this is the operator used to described the domain $\mathcal{D}_\Xi$ of the Anderson Hamiltonian. The final ingredient of paracontrolled calculus is a toolbox of correctors and commutators made to express the singular product between a paracontrolled functions u and the noise ξ in a form involving only ill-defined expressions of the noise independent of u. The first one introduced by [17] is in this framework the corrector

$$\mathsf{C}(u, X, \xi) := \Pi(\widetilde{\mathsf{P}}_u X, \xi) - u\Pi(X, \xi),$$

which translates the rough paths philosophy: the multiplication of a function that locally looks like X with ξ is possible if one is given the multiplication of X itself with ξ. This is the content of the following proposition.

Proposition 1.6. *Let $\alpha \in (0, 1)$ and $\beta, \gamma \in \mathbb{R}$. If*

$$\alpha + \beta < 0 \quad \textit{and} \quad \alpha + \beta + \gamma > 0,$$

then C extends in a unique continuous operator from $\mathcal{C}^\alpha \times \mathcal{C}^\beta \times \mathcal{C}^\gamma$ to $\mathcal{C}^{\alpha+\beta+\gamma}$.

While we do not give the proof, one has the following heuristic. For any $x \in M$, we have

$$\begin{aligned}\mathsf{C}(f, g, h)(x) &= \Pi(\widetilde{\mathsf{P}}_f g, h)(x) - f(x) \cdot \Pi(g, f)(x)\\ &= \Pi(\widetilde{\mathsf{P}}_f g - f(x) \cdot g, h)(x)\\ &\simeq \Pi(\widetilde{\mathsf{P}}_{f-f(x)} g, h)(x),\end{aligned}$$

where $\simeq$ is equal up to a smooth term since $g \simeq \widetilde{\mathsf{P}}_1 g$. Since $f \in \mathcal{C}^\alpha$ with $\alpha \in (0, 1)$, the term $f - f(x)$ allows us to gain regularity in the paraproduct using that $\alpha + \beta < 0$ ending up with a term of better regularity $\alpha + \beta + \gamma > 0$. Continuity results on a number of correctors and commutators and their iterated version are also available; we refer to [27] for further details. For example, one needs the swap operator

$$\mathsf{S}(f, g, h) = \mathsf{P}_h \widetilde{\mathsf{P}}_f g - \mathsf{P}_f \mathsf{P}_h g$$

for the study of the Anderson Hamiltonian which is continuous from $\mathcal{H}^\alpha \times \mathcal{C}^\beta \times \mathcal{C}^\gamma$ to $\mathcal{H}^{\alpha+\beta+\gamma}$ for $\alpha, \beta \in \mathbb{R}$ and $\gamma < 0$.

1C. *Construction of the Anderson Hamiltonian.* In this section, we recall the ideas behind the construction of the Anderson Hamiltonian with the heat semigroup paracontrolled calculus as done in [27] and state the important results we shall use without proofs. We also provide new straightforward results from the construction needed for our proof of Strichartz inequalities. The Anderson Hamiltonian on a two-dimensional manifold M is formally given by

$$H := L + \xi,$$

where $-L$ is the Laplace–Beltrami operator and ξ is a spatial white noise. The noise belongs almost surely to $\mathcal{C}^{\alpha-2}$ for any $\alpha < 1$; hence the product of ξ with a generic L^2-function is not defined almost surely. As explained, it was first constructed by Allez and Chouk [1] on $\mathbb{T}^2$. We work here with the construction on a two-dimensional manifolds from [27], using the high-order paracontrolled calculus since this is the setting in which we want to prove Strichartz inequalities. Following the recent development in

singular stochastic PDEs, the idea is to construct a random almost surely dense subspace $\mathcal{D}_\Xi$ of L^2 such that the operator makes sense for $u \in \mathcal{D}_\Xi \subset L^2$, with Ξ an enhancement of the noise that depends only measurably on the noise ξ. One can then prove that H is self-adjoint with discrete spectrum

$$\lambda_1(\Xi) \leq \lambda_2(\Xi) \leq \cdots \leq \lambda_n(\Xi) \leq \cdots$$

and compare it to the eigenvalues of the Laplace–Beltrami operator $(\lambda_n)_{n\geq 1}$. While the construction of the domain $\mathcal{D}_\Xi$ relied on the notion of strongly paracontrolled functions in [1; 19], the high-order paracontrolled calculus gives a finer description of the domain. In particular, it yields sharp bounds on the eigenvalues of the form

$$\lambda_n - m_\delta^1(\Xi) \leq \lambda_n(\Xi) \leq (1+\delta)\lambda_n + m_\delta^2(\Xi)$$

for any $\delta \in (0, 1)$ and $m_\delta^1(\Xi), m_\delta^2(\Xi) > 0$ random constants depending on the enhanced noise Ξ; see [27] for a precise construction. In particular, it implies the almost sure Weyl-type law

$$\lim_{\lambda\to\infty} \lambda^{-1}|\{n \geq 0 : \lambda_n(\Xi) \leq \lambda\}| = \frac{\mathrm{Vol}(M)}{4\pi}.$$

We briefly present the construction of H and refer to [27] for the details.

Coming from Lyons' rough paths [25] and Gubinelli's controlled paths [16], which were developed as a pathwise approach to stochastic integration, the method used over the last decade to solve singular stochastic PDEs is to work in random subspaces of classical function spaces built from the noise tailor-made for the problem under consideration. In the context of singular random operators, this corresponds to the construction of a random dense domain $\mathcal{D}_\Xi \subset L^2$ on which the operator almost surely makes sense. In the framework of paracontrolled calculus, one considers functions u paracontrolled by noise-dependent reference functions of the form

$$u = \widetilde{\mathsf{P}}_{u'} X + u^\sharp,$$

where the new unknown is $(u', u^\sharp)$. The function u' has to be thought as the "derivative" of u with respect to X, while the error $u^\sharp$ is a smoother remainder. The goal is to find a paracontrolled expression for $u \in L^2$ such that $Hu \in L^2$. Let us first assume that u is smooth. Then we formally get

$$Lu = Hu - u\xi = -\mathsf{P}_u\xi + Hu - \mathsf{P}_\xi u - \Pi(u, \xi) \in \mathcal{H}^{\alpha-2+\kappa}$$

for any $\kappa > 0$. Indeed, the term of lowest regularity is the paraproduct $\mathsf{P}_u\xi \in \mathcal{H}^{\alpha-2}$ since $u \in L^2$ and $\xi \in \mathcal{C}^{\alpha-2}$. Then elliptic regularity theory gives $u \in \mathcal{H}^\alpha$ and suggests for u the paracontrolled form

$$u = \widetilde{\mathsf{P}}_u X + u^\sharp,$$

with $X = -L^{-1}\xi$ and $u^\sharp \in \mathcal{H}^{2\alpha}$. Given such a function, the resonance between u and ξ can be described by

$$\Pi(u, \xi) = \Pi(\widetilde{\mathsf{P}}_u X, \xi) + \Pi(u^\sharp, \xi) = u\Pi(X, \xi) + \mathsf{C}(u, X, \xi) + \Pi(u^\sharp, \xi)$$

using the corrector C since $\Pi(\widetilde{\mathsf{P}}_u X, \xi)$ is not defined due to lack of regularity; see Propositions 1.5 and 1.6. Since $3\alpha + 2 > 0$, the only term on the right-hand side which is potentially undefined is $\Pi(X, \xi)$ and

its definition is independent of the study of H; see [27] for more details including the renormalisation. Given the enhanced data

$$\Xi := (\xi, \Pi(X, \xi)) \in \mathcal{C}^{\alpha-2} \times \mathcal{C}^{2\alpha-2} =: \mathcal{X}^{\alpha},$$

one can define the Anderson Hamiltonian H on

$$\mathcal{D} := \{u \in L^2; u - \widetilde{\mathsf{P}}_u X \in \mathcal{H}^{2\alpha}\} \subset \mathcal{H}^{\alpha},$$

with

$$Hu := Lu + \mathsf{P}_\xi u + u\Pi(X, \xi) + \mathsf{C}(u, X, \xi) + \Pi(u^\sharp, \xi).$$

However, this gives only an unbounded operator $(H, \mathcal{D})$ from $\mathcal{H}^\alpha \subset L^2$ to $\mathcal{H}^{2\alpha-2}$, which is not a subspace of L^2 and thus H will not takes value in L^2 a priori. A finer description of the domain with a second-order paracontrolled expansion allows us to construct a dense subspace $\mathcal{D}_\Xi \subset L^2$ such that $(H, \mathcal{D}_\Xi)$ is an unbounded operator on L^2. Using the classical theory for unbounded operators, it is possible to prove that H is self-adjoint with pure point spectrum. In the expression for H, the roughest term is

$$\mathsf{P}_\xi u + \mathsf{P}_u \Pi(X, \xi) \in \mathcal{H}^{2\alpha-2}.$$

To cancel it with a paracontrolled expansion, we use the commutator S to get

$$\mathsf{P}_\xi u = \mathsf{P}_\xi \widetilde{\mathsf{P}}_u X + \mathsf{P}_\xi u^\sharp = \mathsf{P}_u \mathsf{P}_\xi X + \mathsf{S}(u, X, \xi) + \mathsf{P}_\xi u^\sharp;$$

hence the roughest term is

$$\mathsf{P}_u \mathsf{P}_\xi X + \mathsf{P}_u \Pi(X, \xi) \in \mathcal{H}^{2\alpha-2}.$$

In the end, it is cancelled with the paracontrolled expansion

$$u = \widetilde{\mathsf{P}}_u X_1 + \widetilde{\mathsf{P}}_u X_2 + u^\sharp,$$

where

$$X_1 := -L^{-1}\xi \quad \text{and} \quad X_2 := -L^{-1}(\mathsf{P}_\xi X_1 + \Pi(X_1, \xi)).$$

Definition 1.7. We define the space $\mathcal{D}_\Xi$ of functions paracontrolled by Ξ as

$$\mathcal{D}_\Xi := \{u \in L^2 : u^\sharp := u - \widetilde{\mathsf{P}}_u X_1 - \widetilde{\mathsf{P}}_u X_2 \in \mathcal{H}^2\}.$$

A powerful tool to investigate the domain $\mathcal{D}_\Xi$ and H is the Γ map defined as follows. The domain is given as

$$\mathcal{D}_\Xi = \Phi^{-1}(\mathcal{H}^2),$$

with

$$\Phi(u) := u - \widetilde{\mathsf{P}}_u(X_1 + X_2).$$

The map Φ is not necessarily invertible so we introduce a parameter $s > 0$ and consider the map

$$\Phi^s(u) := u - \widetilde{\mathsf{P}}^s_u(X_1 + X_2),$$

where $\widetilde{\mathsf{P}}^s$ is a truncated paraproduct. In particular, $\widetilde{\mathsf{P}}^s$ goes to 0 as s goes to 0 and the difference $\widetilde{\mathsf{P}} - \widetilde{\mathsf{P}}^s$ is smooth for any $s > 0$. This has to be thought as a frequency cut-off where one gets rid of a number

of low frequencies in order to make a term small. Thus Φ^s is a perturbation of the identity of $\mathcal{H}^\beta$ for any $\beta \in [0, \alpha)$ and thus is invertible for $s = s(\Xi)$ small enough. We define Γ to be its inverse, which is implicitly defined by

$$\Gamma u^\sharp = \widetilde{\mathsf{P}}^s_{\Gamma u^\sharp}(X_1 + X_2) + u^\sharp$$

for any $u^\sharp \in \mathcal{H}^\beta$. It will be a crucial tool to describe the operator H since

$$\mathcal{D}_\Xi = \Phi^{-1}(\mathcal{H}^2) = (\Phi^s)^{-1}(\mathcal{H}^2) = \Gamma(\mathcal{H}^2),$$

where the equality holds because the difference $\widetilde{\mathsf{P}} - \widetilde{\mathsf{P}}^s$ is smooth. Of course the map Γ depends on the choice of s; however, the above reasoning tells us that the image of Γ does not change by changing s, so we omit this dependence in the sequel. The maps Φ^s and Γ satisfy a number of continuity estimates that we shall use throughout this work; this is the content of the following proposition. Let

$$s_\beta(\Xi) := \left(\frac{\alpha - \beta}{m \|\Xi\|_{\mathcal{X}^\alpha}(1 + \|\Xi\|_{\mathcal{X}^\alpha})} \right)^{\frac{4}{\alpha-\beta}}$$

for any $0 \le \beta < \alpha$. Note that the bounds in Sobolev and Hölder spaces are proved directly, while the bounds in L^p follow by interpolation as in [34].

Proposition 1.8. *Let $\beta \in [0, \alpha)$ and $s \in (0, 1)$. We have*

$$\|\Phi^s(u) - u\|_{\mathcal{H}^\beta} \le \frac{m}{\alpha - \beta} s^{\frac{\alpha-\beta}{4}} \|\Xi\|_{\mathcal{X}^\alpha}(1 + \|\Xi\|_{\mathcal{X}^\alpha}) \|u\|_{L^2}.$$

If moreover $s < s_\beta(\Xi)$, this implies

$$\|\Gamma u^\sharp\|_{\mathcal{H}^\beta} \le \frac{1}{1 - \frac{m}{\alpha-\beta} s^{\frac{\alpha-\beta}{4}} \|\Xi\|_{\mathcal{X}^\alpha}(1 + \|\Xi\|_{\mathcal{X}^\alpha})} \|u^\sharp\|_{\mathcal{H}^\beta},$$

as well as the same bounds in $\mathcal{C}^\beta$. The map Φ is also continuous from L^p to itself for $p \in [1, \infty]$ and $\mathcal{H}^\sigma$ to itself for $\sigma \in [0, 1)$, while the same holds for Γ provided s is small enough.

Let us insist that the norm $\mathcal{H}^\beta$ of $u^\sharp_s := \Phi^s(u)$ is always controlled by $\|u\|_{\mathcal{H}^\beta}$, while s needs to be small depending on the noise for $\|u\|_{\mathcal{H}^\beta}$ to be controlled by $\|u^\sharp_s\|_{\mathcal{H}^\beta}$. We also define the map Γ_ε associated to the regularised noise Ξ_ε as

$$\Gamma_\varepsilon u^\sharp = \widetilde{\mathsf{P}}^s_{\Gamma_\varepsilon u^\sharp} X_1^{(\varepsilon)} + \widetilde{\mathsf{P}}^s_{\Gamma_\varepsilon u^\sharp} X_2^{(\varepsilon)} + u^\sharp,$$

with

$$-LX_1^{(\varepsilon)} := \xi_\varepsilon \quad \text{and} \quad -LX_2^{(\varepsilon)} := \Pi(X_1^{(\varepsilon)}, \xi_\varepsilon) - c_\varepsilon + \mathsf{P}_{\xi_\varepsilon} X_1^{(\varepsilon)}.$$

It satisfies the same bounds as Γ with constants which depend in an increasing way on $\|\Xi_\varepsilon\|_{\mathcal{X}^\alpha} \lesssim 1 + \|\Xi\|_{\mathcal{X}^\alpha}$ and the following approximation lemma holds. Thus we may choose s independently of ε.

Lemma 1.9. *For any $0 \le \beta < \alpha$ and $0 < s < s_\beta(\Xi)$, we have*

$$\|\mathrm{Id} - \Gamma\Gamma_\varepsilon^{-1}\|_{L^2 \to \mathcal{H}^\beta} \lesssim_{\Xi,s,\beta} \|\Xi - \Xi_\varepsilon\|_{\mathcal{X}^\alpha}.$$

In particular, this implies the norm convergence of Γ_ε to Γ with the bound

$$\|\Gamma - \Gamma_\varepsilon\|_{\mathcal{H}^\beta \to \mathcal{H}^\beta} \lesssim_{\Xi,s,\beta} \|\Xi - \Xi_\varepsilon\|_{\mathcal{X}^\alpha}.$$

In particular, this allows us to prove density of the domain.

Corollary. *The domain $\mathcal{D}_\Xi$ is dense in $\mathcal{H}^\beta$ for any $\beta \in [0, \alpha)$.*

For any $u \in \mathcal{D}_\Xi$, the operator H is given by

$$Hu = Lu^\sharp + \mathsf{P}_\xi u^\sharp + \Pi(u^\sharp, \xi) + R(u),$$

with $u^\sharp = \Phi(u) \in \mathcal{H}^2$ and R an explicit operator depending on Ξ which is continuous from $\mathcal{H}^\alpha$ to $\mathcal{H}^{3\alpha-2}$. For each $s > 0$, we have a different representation of H, namely

$$Hu = H\Gamma u_s^\sharp = Lu_s^\sharp + \mathsf{P}_\xi u_s^\sharp + \Pi(u_s^\sharp, \xi) + R(\Gamma u_s^\sharp) + \Psi^s(\Gamma u_s^\sharp),$$

with $u_s^\sharp = \Phi^s(u) \in \mathcal{H}^2$ and Ψ^s an explicit operator depending on Ξ and s continuous from L^2 to C^∞, which we henceforth include in the operator R. The operator $H\Gamma$ is thus a perturbation of L; the following proposition shows that it is a continuous operator from $\mathcal{H}^2$ to L^2. In Section 2, we show that it is even a lower-order perturbation of the Laplace–Beltrami operator; this will be crucial to obtain Strichartz inequalities.

Proposition 1.10. *For any $\gamma \in (-\alpha, 3\alpha - 2)$ and s as above, we have*

$$\|Hu\|_{\mathcal{H}^\gamma} = \|H\Gamma u_s^\sharp\|_{\mathcal{H}^\gamma} \lesssim \|u_s^\sharp\|_{\mathcal{H}^{\gamma+2}},$$

with $u = \Gamma u_s^\sharp \in \mathcal{D}_\Xi$. In particular, the result holds for $\gamma \in (-1, 1)$ since the noise belongs to $\mathcal{C}^{\alpha-2}$ for any $\alpha < 1$.

Proof. We have

$$H\Gamma u_s^\sharp = Lu_s^\sharp + \mathsf{P}_\xi u_s^\sharp + \Pi(u_s^\sharp, \xi) + R(u),$$

with $u = \Gamma u_s^\sharp$. Assume first that $0 < \gamma < 3\alpha - 2$; hence

$$\begin{aligned}\|H\Gamma u_s^\sharp\|_{\mathcal{H}^\gamma} &\lesssim \|Lu_s^\sharp\|_{\mathcal{H}^\gamma} + \|\mathsf{P}_\xi u_s^\sharp + \Pi(u_s^\sharp, \xi)\|_{\mathcal{H}^\gamma} + \|R(u)\|_{\mathcal{H}^\gamma} \\ &\lesssim \|u_s^\sharp\|_{\mathcal{H}^{\gamma+2}} + \|\xi\|_{\mathcal{C}^{\alpha-2}} \|u_s^\sharp\|_{\mathcal{H}^{\gamma+2-\alpha}} + \|R(u)\|_{\mathcal{H}^{3\alpha-2}},\end{aligned}$$

where the condition $\gamma > 0$ is needed for the resonant term and $\gamma < 3\alpha - 2$ for $R(u)$. The result follows for this case since

$$\|R(u)\|_{\mathcal{H}^{3\alpha-2}} \lesssim \|u\|_{\mathcal{H}^\alpha} \lesssim \|u_s^\sharp\|_{\mathcal{H}^\alpha} \lesssim \|u_s^\sharp\|_{\mathcal{H}^{\gamma+2}}.$$

Assume now that $-\alpha < \gamma \le 0$. For any $\delta > 0$, we have

$$\begin{aligned}\|H\Gamma u_s^\sharp\|_{\mathcal{H}^\gamma} &\lesssim \|Lu_s^\sharp\|_{\mathcal{H}^\gamma} + \|\mathsf{P}_\xi u_s^\sharp + \Pi(u_s^\sharp, \xi)\|_{\mathcal{H}^\gamma} + \|R(u)\|_{\mathcal{H}^\gamma} \\ &\lesssim \|Lu_s^\sharp\|_{\mathcal{H}^\gamma} + \|\mathsf{P}_\xi u_s^\sharp + \Pi(u_s^\sharp, \xi)\|_{\mathcal{H}^\delta} + \|R(u)\|_{\mathcal{H}^\gamma} \\ &\lesssim \|u_s^\sharp\|_{\mathcal{H}^{\gamma+2}} + \|\xi\|_{\mathcal{C}^{\alpha-2}} \|u_s^\sharp\|_{\mathcal{H}^{\delta+2-\alpha}} + \|R(u)\|_{\mathcal{H}^{3\alpha-2}}\end{aligned}$$

using that $\gamma \le 0 < \delta$. The proof is complete since $\gamma > -\alpha$ and δ small enough implies $\gamma + 2 > \delta + 2 - \alpha$. □

As the parameter $s > 0$ yields different representation of H, the domain $\mathcal{D}_\Xi$ is naturally equipped with the norms

$$\|u\|_{\mathcal{D}_\Xi}^2 := \|u\|_{L^2}^2 + \|u_s^\sharp\|_{\mathcal{H}^2}^2,$$

which are equivalent to the graph norm

$$\|u\|_H^2 := \|u\|_{L^2}^2 + \|Hu\|_{L^2}^2.$$

In particular, this shows that the operator H is closed on its domain $\mathcal{D}_\Xi$.

Proposition 1.11. *Let $u \in \mathcal{D}_\Xi$ and $s > 0$. For any $\delta > 0$, we have*

$$(1-\delta)\|u_s^\sharp\|_{\mathcal{H}^2} \le \|Hu\|_{L^2} + m_\delta^2(\Xi, s)\|u\|_{L^2}$$

and

$$\|Hu\|_{L^2} \le (1+\delta)\|u_s^\sharp\|_{\mathcal{H}^2} + m_\delta^2(\Xi, s)\|u\|_{L^2},$$

with $u_s^\sharp = \Phi^s(u)$ and $m_\delta^2(\Xi, s) > 0$ an explicit constant.

In addition to this comparison between H and L in norm, one has a similar statement in the quadratic form setting.

Proposition 1.12. *Let $u \in \mathcal{D}_\Xi$ and $s > 0$. For any $\delta > 0$, we have*

$$(1-\delta)\langle \nabla u_s^\sharp, \nabla u_s^\sharp\rangle \le \langle u, Hu\rangle + m_\delta^1(\Xi, s)\|u\|_{L^2}^2$$

and

$$\langle u, Hu\rangle \le (1+\delta)\langle \nabla u_s^\sharp, \nabla u_s^\sharp\rangle + m_\delta^1(\Xi, s)\|u\|_{L^2}^2,$$

where $u_s^\sharp = \Phi^s(u)$ and $m_\delta^1(\Xi, s) > 0$ an explicit constant.

One can show that $H\Gamma$ is the limit in norm of $H_\varepsilon\Gamma_\varepsilon$ as operators from $\mathcal{H}^2$ to L^2, where

$$H_\varepsilon := L + \xi_\varepsilon - c_\varepsilon,$$

with c_ε a diverging function as ε goes to 0, again; see Section 2.1 of [27]. In particular, one can take shift c_ε by a large enough constant to ensure that H is positive. Thus the previous proposition implies that $\|\sqrt{H}u\|_{L^2}$ and $\|u_s^\sharp\|_{\mathcal{H}^1}$ are equivalent. The diverging quantity is needed to take care of the singularity as explained in the Introduction; this is the renormalisation procedure with

$$\Pi(X_1, \xi) := \lim_{\varepsilon\to 0} \Pi(X_1^{(\varepsilon)}, \xi_\varepsilon) - c_\varepsilon$$

in $\mathcal{C}^{2\alpha-2}$. In the case of the torus, the noise is invariant by translation and the function c_ε is actually a constant that diverges as $|\log\varepsilon|$; see [1]. This allows us to prove that H is a symmetric operator as the weak limit of the symmetric operators H_ε. Being closed and symmetric, it is enough to prove that

$$(H+k)u = v$$

admits a solution for some $k \in \mathbb{R}$ to get self-adjointness for H; see Theorem X.1 in [30]. This is done using the Babuška–Lax–Milgram theorem; see [3] and Proposition 1.12, which implies that H is almost surely bounded below. This implies self-adjointness and since the resolvent is a compact operator from L^2 to itself since $\mathcal{D}_\Xi \subset \mathcal{H}^\beta$ for any $\beta \in [0, \alpha)$.

Corollary 1.13. *The operator H is self-adjoint with discrete spectrum $(\lambda_n(\Xi))_{n\geq 1}$ which is a nondecreasing diverging sequence without accumulation points. Moreover, we have*

$$L^2 = \bigoplus_{n\geq 1} \mathrm{Ker}(H - \lambda_n(\Xi)),$$

with each kernel being of finite dimension. We finally have the min-max principle

$$\lambda_n(\Xi) = \inf_D \sup_{u\in D;\, \|u\|_{L^2}=1} \langle Hu, u\rangle,$$

where D is any n-dimensional subspace of $\mathcal{D}_\Xi$; this can also be written as

$$\lambda_n(\Xi) = \sup_{v_1,\dots,v_{n-1}\in L^2} \inf_{\substack{\|u\|_{L^2}=1 \\ u\in \mathrm{Vect}(v_1,\dots,v_{n-1})^\perp}} \langle Hu, u\rangle.$$

While the regularity of a function can be measured by its coefficients in the basis of the eigenfunction of the Laplacian, the same is true for the Anderson Hamiltonian and the spaces agree if the regularity one considers is below the form domain.

Proposition 1.14. *For $\beta \in (-\alpha, \alpha)$, there exist two constants $c_\Xi, C_\Xi > 0$ such that*

$$c_\Xi \|H^{\frac{\beta}{2}} u\|_{L^2} \leq \|u\|_{\mathcal{H}^\beta} \leq C_\Xi \|H^{\frac{\beta}{2}} u\|_{L^2}.$$

Proof. Observe first that the statement is clear for $\beta = 0$; we consider only the case $\beta \in (0, \alpha)$ since the case of negative β follows by duality. Again we take $(\varphi_n)_{n\geq 1}$ and $(e_n)_{n\geq 1}$ to denote the basis of eigenfunctions of $-\Delta$ and H respectively. We have for any $v \in \mathcal{D}_\Xi$

$$\begin{aligned}\|H^{\frac{\beta}{2}} v\|_{L^2} &= \left(\sum_{n\geq 1} \lambda_n^\beta \langle v, e_n\rangle^2\right)^{\frac{1}{2}} = \left(\sum_{n\geq 1} \lambda_n^\beta \langle v, e_n\rangle^{2\beta} \langle v, e_n\rangle^{2-2\beta}\right)^{\frac{1}{2}} \\ &\lesssim \left(\sum_{n\geq 1} \lambda_n \langle v, e_n\rangle^2\right)^{\frac{\beta}{2}} \left(\sum_{n\geq 1} \langle v, e_n\rangle^2\right)^{\frac{1-\beta}{2}} \lesssim \|H^{\frac{1}{2}} v\|_{L^2}^\beta \|v\|_{L^2}^{1-\beta}\end{aligned}$$

using Hölder's inequality. Thus the equivalence of $\|H^{1/2} v\|_{L^2}$ and $\|v_s^\sharp\|_{\mathcal{H}^1}$ from Proposition 1.12, together with the continuity of Φ^s from L^2 to itself, yields

$$\|H^{\frac{\beta}{2}} v\|_{L^2} \lesssim \|v_s^\sharp\|_{\mathcal{H}^1}^\beta \|v^\sharp\|_{L^2}^{1-\beta}.$$

Applying this with $v = \Gamma(\langle u_s^\sharp, \varphi_n\rangle \varphi_n)$ gives

$$\begin{aligned}\|H^{\frac{\beta}{2}} \Gamma(\langle u^\sharp, \varphi_n\rangle \varphi_n)\|_{L^2} &\lesssim \|\langle u^\sharp, \varphi_n\rangle \varphi_n\|_{\mathcal{H}^1}^\beta \|\langle u^\sharp, \varphi_n\rangle \varphi_n\|_{L^2}^{1-\beta} \\ &\lesssim |\langle u^\sharp, \varphi_n\rangle| \, \|\varphi_n\|_{\mathcal{H}^\beta}.\end{aligned}$$

Thus

$$\begin{aligned}\|H^{\frac{\beta}{2}} u\|_{L^2}^2 = \|H^{\frac{\beta}{2}} \Gamma(u_s^\sharp)\|_{L^2}^2 &\leq \sum_{n\geq 1} \|H^{\frac{\beta}{2}} \Gamma(\langle u_s^\sharp, \varphi_n\rangle \varphi_n)\|_{L^2}^2 \\ &\lesssim \sum_{n\geq 1} |\langle u_s^\sharp, \varphi_n\rangle|^2 \|\varphi_n\|_{\mathcal{H}^\beta}^2 \lesssim \|u_s^\sharp\|_{\mathcal{H}^\beta}^2.\end{aligned}$$

Since $\beta \in [0, \alpha)$, we get

$$\|H^{\frac{\beta}{2}} u\|_{L^2} \lesssim \|u\|_{\mathcal{H}^\beta}.$$

from the boundedness of Γ; see Proposition 1.8. The other inequality follows from the same reasoning with

$$\|v\|_{\mathcal{H}^\beta} \lesssim \|v_s^\sharp\|_{\mathcal{H}^\beta} \lesssim \|v_s^\sharp\|_{\mathcal{H}^1}^\beta \|v_s^\sharp\|_{L^2}^{1-\beta} \lesssim \|H^{\frac{1}{2}} v\|_{\mathcal{H}^1}^\beta \|u\|_{L^2}^{1-\beta},$$

and applying this bound to $u = \sum_{n\geq 1} \langle u, e_n\rangle e_n$ and proceeding as above we get the other direction. □

The operator H and its spectrum do not depend on $s > 0$ but the different representation of H as

$$Hu = L\Phi^s(u) + \mathsf{P}_\xi \Phi^s(u) + \Pi(\Phi^s(u), \xi) + R(u) + \Psi^s(u)$$

yields different bounds on the eigenvalues. We state the simpler form for the bounds; see [27] for the general result. It is sharp enough to obtain an almost sure Weyl-type law from the one for the Laplace–Beltrami operator.

Proposition 1.15. *Let $\delta \in (0, 1)$. Then there exist two constants $m_\delta^1(\Xi), m_\delta^2(\Xi)$ such that*

$$\lambda_n - m_\delta^1(\Xi) \leq \lambda_n(\Xi) \leq (1+\delta)\lambda_n + m_\delta^2(\Xi)$$

for any $n \in \mathbb{N}$. This implies the almost sure Weyl-type law

$$\lim_{\lambda\to\infty} \lambda^{-1} |\{n \geq 0;\ \lambda_n(\Xi) \leq \lambda\}| = \frac{\mathrm{Vol}(M)}{4\pi}.$$

2. Strichartz inequalities for the stochastic Schrödinger equation

For the rest of the work, we fix a parameter $s > 0$ small enough in order to have all the needed continuity estimates. Every constant may implicitly depend on s and on the norm of the enhanced noise; we do not explicate the dependence since it is not relevant at this stage. From now on, we will also use that α can be taken arbitrarily close to 1 since it is given by the regularity of the spatial white noise. We consider the Schrödinger operator

$$H^\sharp := \Gamma^{-1} H \Gamma,$$

which appears naturally when transforming the Schrödinger equation and the wave equation with multiplicative noise. In fact, if u solves

$$\begin{cases} i\partial_t u + Hu = 0, \\ u(0) = u_0, \end{cases}$$

then $u^\sharp := \Gamma^{-1} u$ solves the transformed equation

$$\begin{cases} i\partial_t u^\sharp + H^\sharp u^\sharp = 0, \\ u^\sharp(0) = \Gamma^{-1} u_0. \end{cases}$$

In this section, we show Strichartz inequalities for the associated Schrödinger equation with an arbitrarily small loss of regularity with respect to the deterministic case. Afterwards, in Section 2B, we detail how these can be used to get a low-regularity solution theory for the nonlinear Schrödinger equation with multiplicative noise.

2A. ***Strichartz inequalities for the Schrödinger group.*** As was hinted at in Proposition 1.11, the transformed operator $H^\sharp$ it is a lower-order perturbation of the Laplace–Beltrami operator. We obtain the following result which is somewhat similar to Theorem 6 in [12] by Burq, Gérard and Tzvetkov, where they proved that the Strichartz inequalities are stable for some lower-order perturbations. This does not cover the case of the Anderson Hamiltonian; however, our proof is very similar; see also [34].

Proposition 2.1. *Let* $0 \le \beta < 1$. *For any* $\kappa > 0$, *we have*

$$\|(H^\sharp - L)v\|_{\mathcal{H}^\beta} \lesssim \|v\|_{\mathcal{H}^{1+\beta+\kappa}}.$$

Proof. For $u = \Gamma u^\sharp \in \mathcal{D}_\Xi$, recall that

$$Hu = Lu^\sharp + \mathsf{P}_\xi u^\sharp + \Pi(u^\sharp, \xi) + R(u),$$

where

$$R(u) := \Pi(u, \Pi(X_1, \xi)) + \mathsf{P}_{\Pi(X_1,\xi)} u + \mathsf{C}(u, X_1, \xi) + \mathsf{P}_u \Pi(X_2, \xi) + \mathsf{D}(u, X_2, \xi) \\ + \mathsf{S}(u, X_2, \xi) + \mathsf{P}_\xi \widetilde{\mathsf{P}}_u X_2 - e^{-L}(\mathsf{P}_u X_1 + \mathsf{P}_u X_2).$$

Thus $H^\sharp v$ is given by

$$H^\sharp v = Lv + \mathsf{P}_\xi v + \Pi(v, \xi) + R(\Gamma v) - \widetilde{\mathsf{P}}_{H\Gamma v}(X_1 + X_2)$$

and for any $\kappa > 0$ and $\beta \in [0, \alpha]$ we have

$$\begin{aligned} \|(H^\sharp - L)v\|_{\mathcal{H}^\beta} &\lesssim \|\mathsf{P}_\xi v + \Pi(v, \xi)\|_{\mathcal{H}^\beta} + \|R(\Gamma v)\|_{\mathcal{H}^\beta} + \|\widetilde{\mathsf{P}}_{H\Gamma v}(X_1 + X_2)\|_{\mathcal{H}^\beta} \\ &\lesssim \|\xi\|_{\mathcal{C}^{-1-\kappa}} \|v\|_{\mathcal{C}^{\beta+1+\kappa}} + \|\Gamma v\|_{\mathcal{H}^\alpha} + \|H\Gamma v\|_{\mathcal{H}^{-1+\kappa+\beta}} \|X_1 + X_2\|_{\mathcal{H}^{1-\kappa}} \\ &\lesssim \|v\|_{\mathcal{H}^{1+\beta+\kappa}} + \|v\|_{\mathcal{H}^\alpha} + \|v\|_{\mathcal{H}^{1+\kappa+\beta}} \end{aligned}$$

using Proposition 1.10, and the proof is complete since $\alpha < 1$. □

Since the unitary group associated to H is bounded on L^2 and on the domain $\mathcal{D}_\Xi$ of H, this implies a similar result for the "sharpened" group associated with $H^\sharp$ in terms of classical Sobolev spaces. Recall that $H^\sharp = \Gamma^{-1} H \Gamma$, with Γ an isomorphism from L^2 to itself; thus $e^{itH^\sharp} := \Gamma^{-1} e^{itH} \Gamma$ is well-defined on L^2. We now state some of its properties.

Proposition 2.2. *For any* $0 \le \beta \le 2$ *and* $t \in \mathbb{R}$, *we have*

$$\|e^{itH^\sharp} v\|_{\mathcal{H}^\beta} \lesssim \|v\|_{\mathcal{H}^\beta}.$$

Moreover, $e^{itH^\sharp}$ *is a nonunitary strongly continuous group of* L^2 *bounded operators*, *namely*

$$e^{i(t+s)H^\sharp} v = e^{itH^\sharp} e^{isH^\sharp} v$$

for all $s, t \in \mathbb{R}$ *and* $v \in L^2$.

Proof. For $\beta = 0$, this follows from the continuity of Γ and Γ^{-1} from L^2 to itself. For $\beta = 2$, we have

$$\|e^{itH^\sharp} v\|_{\mathcal{H}^2} = \|\Gamma^{-1} e^{itH} \Gamma v\|_{\mathcal{H}^2} \lesssim \|H e^{itH} \Gamma v\|_{L^2} \lesssim \|e^{itH} H \Gamma v\|_{L^2} \lesssim \|H \Gamma v\|_{L^2} \lesssim \|v\|_{\mathcal{H}^2},$$

having used Proposition 1.11. The result for any $\beta \in (0, 2)$ is obtained by interpolation and the group property follows simply from the group property of e^{itH} by observing

$$e^{i(t+s)H^\sharp} v = \Gamma^{-1} e^{i(t+s)H} \Gamma v = \Gamma^{-1} e^{itH} \Gamma \Gamma^{-1} e^{isH} \Gamma v = e^{itH^\sharp} e^{isH^\sharp} v. \qquad \square$$

Strichartz inequalities are refinements of the estimates from the previous proposition. The following statement is such a result, which has an arbitrarily small loss of derivative coming from the irregularity of the noise in addition to the $\frac{1}{p}$ loss from the manifold setting without boundary which one sees in [12]. We refer to a pair (p, q) satisfying

$$\frac{2}{p} + \frac{2}{q} = 1$$

as a Strichartz pair from now on.

Theorem 2.3. *Let M be a two-dimensional compact manifold without boundary and let (p, q) be a Strichartz pair. Then for any $\varepsilon > 0$*

$$\|e^{itH^\sharp} v\|_{L^p([0,1],L^q)} \lesssim \|v\|_{\mathcal{H}^{1/p+\varepsilon}}.$$

This implies the bound

$$\|e^{itH} u\|_{L^p([0,1],L^q)} \lesssim \|\Gamma^{-1} u\|_{\mathcal{H}^{1/p+\varepsilon}} \lesssim \|u\|_{\mathcal{H}^{1/p+\varepsilon}}.$$

First, we need to prove the following lemma. It gives the difference between the Schrödinger groups associated to $H^\sharp$ and L from the difference between $H^\sharp$ and L itself. Moreover it quantifies that their difference is small in a short time interval if one gives up some regularity.

Lemma 2.4. *Given $v \in \mathcal{H}^2$, we have*

$$(e^{i(t-t_0)H^\sharp} - e^{i(t-t_0)L}) v = i \int_{t_0}^{t} e^{i(t-s)L} (H^\sharp - L) e^{i(s-t_0)H^\sharp} v \, \mathrm{d}s$$

for any $t, t_0 \in \mathbb{R}$.

Proof. The "sharpened" group yields the solution of the Schrödinger equation

$$(i\partial_t + H^\sharp)(e^{i(t-t_0)H^\sharp} v) = 0,$$

which is equal to v at time $t = t_0$; thus

$$(i\partial_t + L)(e^{i(t-t_0)H^\sharp} v) = (L - H^\sharp)(e^{i(t-t_0)H^\sharp} v).$$

Using the unitary group representation of the solution to the Schrödinger equation associated to L, we deduce that

$$(i\partial_t + L)(e^{i(t-t_0)L} v - e^{i(t-t_0)H^\sharp} v) = (H^\sharp - L)(e^{i(t-t_0)H^\sharp} v).$$

Since the solution is equal to 0 at time $t = t_0$, the mild formulation of this last equation yields

$$(e^{i(t-t_0)H^\sharp} - e^{i(t-t_0)L}) v = i \int_{t_0}^{t} e^{i(t-s)L} (H^\sharp - L) e^{i(s-t_0)H^\sharp} v \, \mathrm{d}s. \qquad \square$$

Proof of Theorem 2.3. For $N \in \mathbb{N}^*$ to be chosen, we have

$$\|e^{itH^\sharp} v\|^p_{L^p([0,1],L^q)} = \sum_{\ell=0}^{N} \|e^{itH^\sharp} v\|^p_{L^p([t_\ell,t_{\ell+1}],L^q)},$$

where $t_\ell := \ell/N$. For $t \in [t_\ell, t_{\ell+1})$, the previous lemma gives

$$e^{itH^\sharp} v = e^{i(t-t_\ell)H^\sharp} e^{it_\ell H^\sharp} v = e^{i(t-t_\ell)L} e^{it_\ell H^\sharp} v + i\int_{t_\ell}^{t} e^{i(t-s)L}(H^\sharp - L)e^{isH^\sharp} v \,\mathrm{d}s.$$

Applying this with $v = \Delta_k u$ gives

$$\begin{aligned}\|\Delta_j e^{itH^\sharp}\Delta_k u\|^p_{L^p([0,1],L^q)} \le \sum_{\ell=0}^{N} \|\Delta_j e^{i(t-t_\ell)L} e^{it_\ell H^\sharp}\Delta_k u\|^p_{L^p([t_\ell,t_{\ell+1}],L^q)} \\ + \sum_{\ell=0}^{N}\left\|\Delta_j \int_{t_\ell}^{t} e^{i(t-s)L}(H^\sharp - L)e^{isH^\sharp}\Delta_k u \,\mathrm{d}s\right\|^p_{L^p([t_\ell,t_{\ell+1}],L^q)}.\end{aligned}$$

Assume $N \ge 2^j$ such that $|t_{\ell+1} - t_\ell| \le 2^{-j}$. For the first term, we have

$$\begin{aligned}\|\Delta_j e^{i(t-t_\ell)L} e^{it_\ell H^\sharp}\Delta_k u\|^p_{L^p([t_\ell,t_{\ell+1}],L^q)} &= \|e^{i(t-t_\ell)L}\Delta_j e^{it_\ell H^\sharp}\Delta_k u\|^p_{L^p([t_\ell,t_{\ell+1}],L^q)} \\ &\lesssim \|\Delta_j e^{it_\ell H^\sharp}\Delta_k u\|^p_{L^2} \lesssim 2^{-j\delta p}\|\Delta_j e^{it_\ell H^\sharp}\Delta_k u\|^p_{\mathcal{H}^\delta} \\ &\lesssim 2^{-j\delta p}2^{-k\delta' p}\|\Delta_k u\|^p_{\mathcal{H}^{\delta+\delta'}}\end{aligned}$$

for any $\delta, \delta' \in \mathbb{R}$ using Proposition 2.2, the Strichartz inequality for spectrally localised data from Section 1A and Bernstein's lemma; see Lemma 1.3. For the second term, we have

$$\begin{aligned}&\left\|\Delta_j \int_{t_\ell}^{t} e^{i(t-s)L}(H^\sharp - L)e^{isH^\sharp}\Delta_k u \,\mathrm{d}s\right\|^p_{L^p([t_\ell,t_{\ell+1}],L^q)} \\ &\quad= \left\|\int_{t_\ell}^{t} e^{i(t-s)L}\Delta_j(H^\sharp - L)e^{isH^\sharp}\Delta_k u \,\mathrm{d}s\right\|^p_{L^p([t_\ell,t_{\ell+1}],L^q)} \\ &\quad\lesssim \left(\int_{t_\ell}^{t_{\ell+1}} \|e^{i(t-s)L}\Delta_j(H^\sharp - L)e^{isH^\sharp}\Delta_k u\|_{L^p([t_\ell,t_{\ell+1}],L^q)}\,\mathrm{d}s\right)^p \\ &\quad\lesssim \left(\int_{t_\ell}^{t_{\ell+1}} \|\Delta_j(H^\sharp - L)e^{isH^\sharp}\Delta_k u\|_{L^2}\,\mathrm{d}s\right)^p \lesssim 2^{-j\sigma p}\left(\int_{t_\ell}^{t_{\ell+1}} \|\Delta_j(H^\sharp - L)e^{isH^\sharp}\Delta_k u\|_{\mathcal{H}^\sigma}\,\mathrm{d}s\right)^p \\ &\quad\lesssim 2^{-j\sigma p}\left(\int_{t_\ell}^{t_{\ell+1}} \|(H^\sharp - L)e^{isH^\sharp}\Delta_k u\|_{\mathcal{H}^\sigma}\,\mathrm{d}s\right)^p \lesssim 2^{-j\sigma p}\left(\int_{t_\ell}^{t_{\ell+1}} \|e^{isH^\sharp}\Delta_k u\|_{\mathcal{H}^{\sigma+1+\kappa}}\,\mathrm{d}s\right)^p \\ &\quad\lesssim N^{-p}2^{-j\sigma p}\|\Delta_k u\|^p_{\mathcal{H}^{1+\sigma+\kappa}} \lesssim N^{-p}2^{-j\sigma p}2^{-k\sigma' p}\|\Delta_k u\|^p_{\mathcal{H}^{1+\sigma+\sigma'+\kappa}}\end{aligned}$$

for any $\sigma \in (0,1)$, $\sigma' \in \mathbb{R}$ and $0 < \kappa < 1-\alpha$, where again the dyadic factors come from Bernstein's lemma and we have used the bounds from Propositions 2.1 and 2.2 with the Strichartz inequality for spectrally localised data. Summing over the subintervals gives

$$\|\Delta_j e^{itH^\sharp}\Delta_k u\|_{L^p([0,1],L^q)} \lesssim N^{\frac{1}{p}}2^{-j\delta}2^{-k\delta'}\|\Delta_k u\|_{\mathcal{H}^{\delta+\delta'}} + N^{\frac{1-p}{p}}2^{-j\sigma}2^{-k\sigma'}\|\Delta_k u\|_{\mathcal{H}^{1+\sigma+\sigma'+\kappa}}.$$

Let $\eta > 0$ small and take

$$N = 2^j, \quad \delta = \eta + \frac{1}{p}, \quad \delta' = \sigma = \sigma' = \eta,$$

which satisfies in particular $N \geq 2^j$ and $\sigma \in [0, \alpha)$ to sum over $k \leq j$. We get

$$\begin{aligned}
\Big\| \sum_{k \leq j} \Delta_j e^{itH^\sharp} \Delta_k u \Big\|_{L^p([0,1],L^q)} &\lesssim \sum_j \sum_{k \leq j} \| \Delta_j e^{itH^\sharp} \Delta_k u \|_{L^p([0,1],L^q)} \\
&\lesssim \sum_{j \geq 0} \sum_{k \leq j} 2^{\frac{j}{p}} 2^{-j(\frac{1}{p}+\eta)} 2^{-k\eta} \| \Delta_k u \|_{\mathcal{H}^{1/p+2\eta}} + 2^{j\frac{1-p}{p}} 2^{-j\eta} 2^{-k\eta} \| \Delta_k u \|_{\mathcal{H}^{1+2\eta+\kappa}} \\
&\lesssim \sum_{j \geq 0} 2^{-j\eta} \| \Delta_{\leq j} u \|_{\mathcal{H}^{1/p+2\eta}} + 2^{j\frac{1-p}{p}} 2^{-j\eta} \| \Delta_{\leq j} u \|_{\mathcal{H}^{1+2\eta+\kappa}} \\
&\lesssim \| u \|_{\mathcal{H}^{1/p+2\eta}} + \sum_{j \geq 0} 2^{-j\eta} 2^{j\frac{1-p}{p}} 2^{-j\frac{1-p}{p}} \| \Delta_{\leq j} u \|_{\mathcal{H}^{1-1+1/p+2\eta+\kappa}} \\
&\lesssim \| u \|_{\mathcal{H}^{1/p+2\eta}} + \| u \|_{\mathcal{H}^{1/p+2\eta+\kappa}},
\end{aligned}$$

having used Bernstein's inequality, Lemma 1.3, for the projector $\Delta_{\leq j}$. For the sum over $j \leq k$, we choose instead

$$N = 2^k, \quad \delta = \delta' = \sigma = \sigma' = \eta,$$

with $\eta > 0$ small as before. Since $j \leq k$, we have $N \geq 2^j$ and thus get the bound for the other part of the double sum

$$\begin{aligned}
\Big\| \sum_{j \leq k} \Delta_j e^{itH^\sharp} \Delta_k u \Big\|_{L^p([0,1],L^q)} &\lesssim \sum_{k \geq 0} \sum_{j \leq k} \| \Delta_j e^{itH^\sharp} \Delta_k u \|_{L^p([0,1],L^q)} \\
&\lesssim \sum_{k \geq 0} \sum_{j \leq k} 2^{\frac{k}{p}} 2^{-j\eta} 2^{-k\eta} \| \Delta_k u \|_{\mathcal{H}^{2\eta}} + 2^{\frac{k(1-p)}{p}} 2^{-j\eta} 2^{-k\eta} \| \Delta_k u \|_{\mathcal{H}^{1+2\eta+\kappa}} \\
&\lesssim \| u \|_{\mathcal{H}^{1/p+2\eta}} + \| u \|_{\mathcal{H}^{1+(1-p)/p+2\eta+\kappa}} \\
&\lesssim \| u \|_{\mathcal{H}^{1/p+2\eta+\kappa}},
\end{aligned}$$

having used Bernstein's inequality again. This completes the proof since η and κ can be taken arbitrarily small. This implies the bound

$$\| e^{itH} u \|_{L^p([0,1],L^q)} \lesssim \| \Gamma^{-1} u \|_{\mathcal{H}^{1/p+\varepsilon}} \lesssim \| u \|_{\mathcal{H}^{1/p+\varepsilon}}$$

using that $\frac{1}{p} + \varepsilon < 1$ and Proposition 1.8. □

Remark. We proved that Strichartz inequalities are stable under suitable perturbation, that is, lower-order perturbation in the sense of Proposition 2.1. This is similar in spirit to Theorem 6 in [12]. One can show that the magnetic Laplacian with white noise magnetic field constructed in [26] is also a lower-order perturbation of the Laplacian on the two-dimensional torus in this sense. Thus Theorem 2.3 also gives Strichartz inequalities for the associated Schrödinger group.

As a corollary, we state the inhomogeneous inequalities needed to solve the nonlinear equation. This is straightforward; see [34].

Corollary 2.5. *In the setting of Theorem 2.3, we have in addition the bound*

$$\Big\| \int_0^t e^{i(t-s)H^\sharp} f(s) \, \mathrm{d}s \Big\|_{L^p([0,1],L^q)} \lesssim \int_0^1 \| f(s) \|_{\mathcal{H}^{1/p+\varepsilon}} \, \mathrm{d}s$$

for all $f \in L^1([0,1], \mathcal{H}^{1/p+\varepsilon})$.

The only ingredient in the proof of the theorem where the boundary appears is when we apply the result for the Laplacian. By using Theorem 1.2 and in particular (2) instead of (1), we immediately get the following analogous result which is of course weaker.

Theorem 2.6. *Let M be a compact surface with boundary. Let $p \in (3, \infty]$ and $q \in [2, \infty)$ such that*

$$\frac{3}{p} + \frac{2}{q} = 1.$$

Then for any $\varepsilon > 0$

$$\|e^{itH^\sharp} u\|_{L^p([0,1], L^q)} \lesssim \|u\|_{\mathcal{H}^{2/p+\varepsilon}}$$

and

$$\left\| \int_0^t e^{i(t-s)H^\sharp} f(s)\, \mathrm{d}s \right\|_{L^p([0,1], L^q)} \lesssim \int_0^1 \|f(s)\|_{\mathcal{H}^{2/p+\varepsilon}}\, \mathrm{d}s$$

for all $f \in L^1([0, 1], \mathcal{H}^{2/p+\varepsilon})$.

2B. ***Local well-posedness for multiplicative stochastic cubic NLS.*** We now apply our results to the local-in-time well-posedness of the cubic multiplicative stochastic NLS

$$\begin{cases} i\partial_t u + Hu = -|u|^2 u, \\ u(0) = u_0, \end{cases}$$

with $u_0 \in \mathcal{H}^\sigma$ where $\sigma \in \left(\frac{1}{2}, 1\right)$ and in the energy space, that is, $u_0 \in \mathcal{D}(\sqrt{H}) = \Gamma\mathcal{H}^1$. The latter hypothesis is natural to assume, since solutions starting in the energy space, usually called energy solutions, are intimately related to the conserved energy

$$E(u)(t) := \frac{1}{2}(u(t), Hu(t)) + \frac{1}{4}\int |u(t)|^4 = E(u_0), \tag{3}$$

introduced in [19] on the torus and [27] on general surfaces. Thus we refer to $\mathcal{D}(\sqrt{H}) = \Gamma\mathcal{H}^1$ as the energy space for the Anderson Hamiltonian; see Proposition 1.12. Note that, as is usual in these types of fixed-point arguments, the sign of the nonlinearity does not play a role for local well-posedness, but for the sake of definiteness we take the defocussing nonlinearity. We remark also that one can prove similar results for more general nonlinearities, we considered only the cubic equation in this work. See for example [33; 32], where they considered generic polynomial nonlinearity and obtained global well-posedness. As explained in Section 3.2.2 of [19], their result for the equation with white noise potential is weaker than the one for the deterministic equation since Strichartz inequalities were not known in this singular case. This was a motivation for the study of Strichartz estimates for the operator H. The well-posedness itself follows from a fairly straightforward contraction argument, similar to, e.g., Proposition 3.1 in [12]. Finally, we only consider a surface without boundary; the case with boundary is analogous using Theorem 2.6 instead of Theorem 2.3. The mild formulation is

$$u(t) = e^{itH} u_0 - i\int_0^t e^{i(t-s)H}(|u|^2 u)(s)\, \mathrm{d}s$$

and applying the Γ^{-1} map introduced in Section 1C yields the mild formulation for the transformed quantity $u^\sharp = \Gamma^{-1}u$. We get

$$u^\sharp(t) = e^{itH^\sharp}u_0^\sharp - i\int_0^t e^{i(t-s)H^\sharp}\Gamma^{-1}(|\Gamma u^\sharp|^2\Gamma u^\sharp)(s)\,\mathrm{d}s,$$

where $u_0^\sharp := \Gamma^{-1}u_0$; this is where the transformed operator $H^\sharp = \Gamma^{-1}H\Gamma$ appears naturally. Despite the seemingly complicated nonlinear expression, this new mild formulation is easier to deal with since $H^\sharp$ is a perturbation of the Laplacian and has domain $\mathcal{H}^2$; hence it is not as outlandish as H and its domain which contains no nonzero smooth functions. Now, we have to find a bound for the map

$$\Psi(v)(t) := e^{itH^\sharp}v_0 - i\int_0^t e^{i(t-s)H^\sharp}\Gamma^{-1}(|\Gamma v|^2\Gamma v)(s)\,\mathrm{d}s$$

in a suitable space which allows us to get a unique fixed point. One then recovers a solution to the original equation with $u := \Gamma v$ and choosing $v_0 := \Gamma^{-1}u_0$. Since Γ is an isomorphism on L^p for $p \in [2,\infty]$ and both Γ and $e^{itH^\sharp}$ are isomorphisms on $\mathcal{H}^\sigma$ to itself for $\sigma \in [0,1)$, it is natural to consider initial datum v_0, and thus also u_0, in $\mathcal{H}^\sigma$ for $0<\sigma<1$. Therefore we bound $\Psi(v)$ in $\mathcal{H}^\sigma$ with

$$\begin{aligned}
\|\Psi(v)(t)\|_{\mathcal{H}^\sigma} &\lesssim \|v_0\|_{\mathcal{H}^\sigma} + \int_0^t \|\Gamma v(s)^3\|_{\mathcal{H}^\sigma}\,\mathrm{d}s \\
&\lesssim \|v_0\|_{\mathcal{H}^\sigma} + \int_0^t \|\Gamma v(s)\|_{\mathcal{H}^\sigma}\|\Gamma v(s)\|_{L^\infty}^2\,\mathrm{d}s \\
&\lesssim \|v_0\|_{\mathcal{H}^\sigma} + \int_0^t \|v(s)\|_{\mathcal{H}^\sigma}\|v(s)\|_{L^\infty}^2\,\mathrm{d}s \\
&\lesssim \|v_0\|_{\mathcal{H}^\sigma} + \|v\|_{L^\infty([0,t],\mathcal{H}^\sigma)}\|v\|^2_{L^2([0,t],L^\infty)},
\end{aligned}$$

where in the first and third lines we have used the continuity of $e^{itH^\sharp}$ and Γ and Lemma 1.4 in the second line. For $\sigma<1$, the space $\mathcal{H}^\sigma$ is not an algebra and one cannot simply use its norm to bound the nonlinearity. However, one may bound it using the L^∞-norm in space by observing that one needs less integrability in time and this is precisely the point where the Strichartz estimates turn out to be useful. As for the deterministic equation, we work with the function spaces

$$\mathcal{W}^{\beta,q}(M) = \{u \in \mathcal{D}'(M);\ (1-\Delta)^{\frac{\beta}{2}}u \in L^q\},$$

with associated norm

$$\|u\|_{\mathcal{W}^{\beta,q}} := \|(1-\Delta)^{\frac{\beta}{2}}u\|_{L^q}.$$

For $\beta \in [0,1)$ and $q=2$, one recovers the Sobolev spaces and the norm is equivalent to

$$\|u\|_{\mathcal{H}^\beta} = \|(1+H)^{\frac{\beta}{2}}u\|_{L^2}$$

by Proposition 1.14. Within this framework, Strichartz inequalities from Theorem 2.3 give us the bound

$$\|e^{itH^\sharp}w\|_{L^p([0,1],\mathcal{W}^{\beta,q})} \lesssim \|w\|_{\mathcal{H}^{1/p+\beta+\kappa}}$$

for any Strichartz pair (p,q) and $\kappa>0$. Furthermore, the space $\mathcal{W}^{\beta,q}$ is continuously embedded in L^∞ for $\beta > \frac{2}{q}$. Let $\sigma \in \mathbb{R}$ such that

$$\frac{1}{p} + \frac{2}{q} + 2\kappa \le \sigma.$$

Thus for $0 < t \leq 1$, we get the bound

$$\begin{aligned}\|\Psi(v)\|_{L^p([0,t],\mathcal{W}^{2/q+\kappa,q})} &\lesssim \|v_0\|_{\mathcal{H}^{1/p+2/q+2\kappa}} + \int_0^t \|\Gamma^{-1}(|\Gamma v|^2\Gamma v)(s)\|_{\mathcal{H}^{1/p+2/q+2\kappa}}\,\mathrm{d}s \\ &\lesssim \|v_0\|_{\mathcal{H}^\sigma} + \int_0^t \|\Gamma v(s)^3\|_{\mathcal{H}^\sigma}\,\mathrm{d}s \\ &\lesssim \|v_0\|_{\mathcal{H}^\sigma} + \|v\|_{L^\infty([0,t],\mathcal{H}^\sigma)}\|v\|^2_{L^2([0,t],L^\infty)} \\ &\lesssim \|v_0\|_{\mathcal{H}^\sigma} + \|v\|_{L^\infty([0,t],\mathcal{H}^\sigma)}\|v\|^2_{L^2([0,t],\mathcal{W}^{2/q+\kappa,q})} \\ &\lesssim \|v_0\|_{\mathcal{H}^\sigma} + t^{\frac{p-2}{p}}\|v\|_{L^\infty([0,t],\mathcal{H}^\sigma)}\|v\|^2_{L^p([0,t],\mathcal{W}^{2/q+\kappa,q})}\end{aligned}$$

using Corollary 2.5 in the first line, Hölder inequality in the last line and bicontinuity of Γ from $\mathcal{H}^\sigma$ to itself. For $0 < t' \leq t$, we also have

$$\begin{aligned}\|\Psi(v)(t')\|_{\mathcal{H}^\sigma} &\lesssim \|v_0\|_{\mathcal{H}^\sigma} + \int_0^{t'} \|v(s)\|_{\mathcal{H}^\sigma}\|v(s)\|^2_{L^\infty}\,\mathrm{d}s \\ &\lesssim \|v_0\|_{\mathcal{H}^\sigma} + \|v\|_{L^\infty([0,t'],\mathcal{H}^\sigma)}\|v\|^2_{L^2([0,t'],L^\infty)}, \\ &\lesssim \|v_0\|_{\mathcal{H}^\sigma} + t'^{\frac{p-2}{p}}\|v\|_{L^\infty([0,t'],\mathcal{H}^\sigma)}\|v\|^2_{L^p([0,t'],\mathcal{W}^{2/q+\kappa,q})}.\end{aligned}$$

This gives us the combined bound

$$\|\Psi(v)\|_{L^p([0,t],\mathcal{W}^{2/q+\kappa,q})} + \|\Psi(v)\|_{L^\infty([0,t],\mathcal{H}^\sigma)} \lesssim \|v_0\|_{\mathcal{H}^\sigma} + t^{\frac{p-2}{p}}\|v\|_{L^\infty([0,t],\mathcal{H}^\sigma)}\|v\|^2_{L^p([0,t],\mathcal{W}^{2/q+\kappa,q})} \tag{4}$$

that will be the main tool for the fixed point. Note that the restrictions

$$\frac{1}{p} + \frac{2}{q} + 2\kappa \leq \sigma \quad \text{and} \quad \frac{2}{p} + \frac{2}{q} = 1$$

give

$$1 - \frac{1}{p} + 2\kappa \leq \sigma.$$

Since $p \geq 2$ and $\kappa > 0$ can be taken arbitrary small, this gives

$$\sigma > \tfrac{1}{2}$$

and leads to the following local-in-time well-posedness result. Without Strichartz estimates, even in the classical case, one could not go beyond the threshold $\sigma \geq 1$.

Theorem 2.7. *Let M be a compact surface without boundary, $\sigma > \frac{1}{2}$ and initial data $v_0 \in \mathcal{H}^\sigma$. Let $\kappa > 0$ and (p, q) a Strichartz pair such that*

$$\frac{1}{p} + \frac{2}{q} + 2\kappa \leq \sigma.$$

There exists a time $T > 0$ until which there exists a unique solution

$$v \in C([0, T], \mathcal{H}^\sigma) \cap L^p([0, T], \mathcal{W}^{\frac{2}{q}+\kappa,q})$$

to the mild formulation of the transformed PDE

$$\begin{cases} i\partial_t v + H^\sharp v = -\Gamma^{-1}(|\Gamma v|^2\Gamma v), \\ v(0) = v_0. \end{cases}$$

Moreover, the solution depends continuously on the initial data $v_0 \in \mathcal{H}^\sigma$.

Proof. This is a straightforward contraction argument where the main ingredient is the bound proved in the preceding arguments. By choosing the radius R of the ball and the final time appropriately, we can prove that

$$\Psi : B(0, R)_{C([0,T],\mathcal{H}^\sigma)\cap L^p(([0,T],\mathcal{W}^{2/q+\kappa,q})} \to B(0, R)_{C([0,T],\mathcal{H}^\sigma)\cap L^p(([0,T],\mathcal{W}^{2/q+\kappa,q})}$$

is in fact a contraction. Using the previously established bound (4) and an analogous bound for the difference, one finds that this can be achieved if one chooses

$$R = 2\widetilde{C}\|v_0\|_{\mathcal{H}^\sigma} \quad \text{and} \quad T = \left(\frac{1}{3R^2\widetilde{C}}\right)^{\frac{p}{p-2}}$$

for some constant $\widetilde{C}$ which depends on the norm of the enhanced noise Ξ and the parameters appearing. □

Finally we give the analogous result for surfaces with boundary, which is of course weaker; however, we still get a better result than one gets simply from using the algebra property of Sobolev spaces.

Theorem 2.8. *Let M be a compact surface with boundary, $\sigma > \frac{2}{3}$ and p, q, κ such that*

$$\frac{3}{p} + \frac{2}{q} = 1 \quad \textit{and} \quad \frac{2}{p} + \frac{2}{q} + 2\kappa \le \sigma.$$

For any initial datum $v_0 \in \mathcal{H}^\sigma$ there exists a unique solution

$$v \in C([0, T], \mathcal{H}^\sigma) \cap L^p([0, T], \mathcal{W}^{\frac{2}{q}+\kappa,q})$$

to the mild formulation of the transformed PDE up to a time $T > 0$ depending on the data which depends continuously on the initial condition.

3. Strichartz inequalities for the stochastic wave equation

Again, we consider the "sharpened" operator

$$H^\sharp := \Gamma^{-1} H \Gamma$$

which appears naturally when transforming the wave equation with multiplicative noise. If u solves

$$\begin{cases} \partial_t^2 u + Hu = 0, \\ (u, \partial_t u)|_{t=0} = (u_0, u_1), \end{cases}$$

then $u^\sharp := \Gamma^{-1} u$ solves the transformed equation

$$\begin{cases} \partial_t^2 u^\sharp + H^\sharp u^\sharp = 0, \\ (u^\sharp, \partial_t u^\sharp)|_{t=0} = (\Gamma^{-1} u_0, \Gamma^{-1} u_1). \end{cases}$$

In this section, we show Strichartz inequalities for the associated wave equation. We will further detail how these can be used to get a low-regularity solution theory for the nonlinear wave equation with multiplicative noise. This equation was also considered by Zachhuber [35] on the full space in two and three dimensions where global well-posedness is obtained using finite speed of propagation.

3A. ***Strichartz inequalities for the wave propagator.*** The propagator associated to the wave equation is

$$(u_0, u_1) \mapsto \cos(t\sqrt{H})u_0 + \frac{\sin(t\sqrt{H})}{\sqrt{H}}u_1,$$

with initial conditions $(u, \partial_t u)|_{t=0} = (u_0, u_1)$. As for the Schrödinger equation, the following Strichartz inequalities hold on a two-dimensional compact Riemannian manifold without boundary; see [7]. We state the result in the homogeneous case for simplicity; however, one directly obtains inhomogeneous bounds as in Corollary 2.5. We cite the following Strichartz estimates, which hold on compact surfaces respectively without and with boundary; see [7].

Theorem 3.1. *Let (M, g) be a compact two-dimensional Riemannian manifold without boundary. Let $p, q \in [2, \infty]$ such that*

$$\frac{2}{p} + \frac{1}{q} \leq \frac{1}{2}$$

and consider

$$\frac{1}{p} + \frac{2}{q} := 1 - \sigma.$$

Then the solution to

$$\begin{cases} (\partial_t^2 - \Delta_g)u = 0, \\ (u, \partial_t u)|_{t=0} = (u_0, u_1) \in \mathcal{H}^\sigma \times \mathcal{H}^{\sigma-1} \end{cases}$$

satisfies the bound

$$\|u\|_{L^p([0,T],L^q)} \lesssim \|u_0\|_{\mathcal{H}^\sigma} + \|u_1\|_{\mathcal{H}^{\sigma-1}}.$$

In the case where the surface M has a boundary, there is this slightly weaker result.

Theorem 3.2. *Let (M, g) be a compact two-dimensional Riemannian manifold with boundary. Let $p \in (2, \infty]$ and $q \in [2, \infty)$ such that*

$$\frac{3}{p} + \frac{1}{q} \leq \frac{1}{2}$$

and consider σ given by

$$\frac{1}{p} + \frac{2}{q} = 1 - \sigma.$$

Then the solution to

$$\begin{cases} (\partial_t^2 - \Delta_g)u = 0, \\ (u, \partial_t u)|_{t=0} = (u_0, u_1) \in \mathcal{H}^\sigma \times \mathcal{H}^{\sigma-1} \end{cases}$$

satisfies the bound

$$\|u\|_{L^p([0,T],L^q)} \lesssim \|u_0\|_{\mathcal{H}^\sigma} + \|u_1\|_{\mathcal{H}^{\sigma-1}}.$$

3B. ***Strichartz inequalities for wave equations with rough potentials.*** While our proof of the Strichartz inequalities for the Schrödinger equation with white noise potential strongly relies on the deterministic result, this is not the case for the wave equation. In this case, we follow the approach from [13] for which one has two mains ingredients, firstly a Weyl law for the Laplace–Beltrami operator and secondly L^q bounds on its eigenfunctions. In particular, we treat at the same time the case with and without boundary here, the only difference being that one has weaker L^q bounds on the eigenfunctions.

An analogous Weyl law for the Anderson Hamiltonian was obtained in [27]; see Proposition 1.15 in Section 1C, and the analogue of the second part follows from the Strichartz inequalities for the Schrödinger

group obtained in Section 2. Let $(e_n)_{n\geq 1}$ be an orthonormal family of eigenfunctions of H associated to $(\lambda_n(\Xi))_{n\geq 1}$. Since the eigenfunctions belong to the domain $\mathcal{D}_\Xi$, they belong in particular to L^∞ and we have the following bounds on its L^q-norm for $q \in (2, \infty)$. Recall that $(\lambda_n)_{n\geq 1}$ are the eigenvalues of the Laplacian.

Proposition 3.3. *Let $q \in (2, \infty)$ and M a compact surface without boundary. We have*

$$\|e_n\|_{L^q} \lesssim \sqrt{\lambda_n(\Xi)}^{\frac{1}{2}-\frac{1}{q}+\kappa}$$

for any $\kappa > 0$. In particular, this implies

$$\|e_n\|_{L^q} \lesssim (1+\sqrt{\lambda_n})^{\frac{1}{2}-\frac{1}{q}+\kappa} \lesssim (1+\sqrt{n})^{\frac{1}{2}-\frac{1}{q}+\kappa}.$$

Proof. We have

$$\|e_n\|_{L^q} = \|e^{it\lambda_n} e_n\|_{L^p([0,1],L^q)} = \|e^{itH} e_n\|_{L^p([0,1],L^q)},$$

with (p, q) a Strichartz pair. For any $\kappa > 0$, this gives

$$\|e_n\|_{L^q} \lesssim \|e_n\|_{\mathcal{H}^{1/p+\kappa}} \lesssim \|\sqrt{H}^{\frac{1}{p}+\kappa} e_n\|_{L^2} \lesssim \sqrt{\lambda_n(\Xi)}^{\frac{1}{p}+\kappa}$$

using Proposition 1.14 and

$$\frac{1}{p} = \frac{1}{2} - \frac{1}{q}.$$

Finally, Proposition 1.15 gives the bound

$$\lambda_n(\Xi) \lesssim 1 + \lambda_n$$

and completes the proof. □

Another important operator is the projection onto the eigenspaces of H. Let

$$\Pi_\lambda u := \sum_{\lambda_n(\Xi)\in[\lambda,\lambda+1)} \langle u, e_n\rangle e_n$$

for any $\lambda \geq 0$. These spectral projectors satisfy the following bounds.

Proposition 3.4. *Let $\lambda \geq 0$ and $q \in (2, \infty)$. We have*

$$\|\Pi_\lambda u\|_{L^q} \lesssim \sqrt{\lambda+1}^{\frac{1}{2}-\frac{1}{q}+\varepsilon} \|u\|_{L^2}$$

for any $\varepsilon > 0$.

Proof. Consider $\lfloor H \rfloor$ the "integer part" of H, which is the self-adjoint operator defined by

$$\lfloor H \rfloor e_n := \lfloor \lambda_n(\Xi) \rfloor e_n$$

for $n \geq 1$. Then we have for any $\varepsilon > 0$ the bound

$$\|e^{it\lfloor H \rfloor} v\|_{L^p([0,1],L^q)} \lesssim \|v\|_{\mathcal{H}^{1/p+\varepsilon}},$$

which follows from the one for H, namely Theorem 2.3. Indeed, we have

$$e^{it\lfloor H \rfloor} v - e^{itH} v = -i \int_0^t e^{i(t-s)H} (H - \lfloor H \rfloor) e^{is\lfloor H \rfloor} v \, \mathrm{d}s,$$

and using Theorem 2.3 and Corollary 2.5, this gives

$$\begin{aligned}\|e^{it\lfloor H\rfloor}v\|_{L^p([0,1],L^q)} &\lesssim \|e^{itH}v\|_{L^p([0,1],L^q)} + \int_0^1 \|e^{i(t-s)\lfloor H\rfloor}(H-\lfloor H\rfloor)e^{isH}\|_{L^p([0,1],L^q)}\,\mathrm{d}s \\ &\lesssim \|v\|_{\mathcal{H}^{1/p+\varepsilon}} + \int_0^1 \|(H-\lfloor H\rfloor)e^{i(t-s)\lfloor H\rfloor}v\|_{\mathcal{H}^{1/p+\varepsilon}}\,\mathrm{d}s \\ &\lesssim \|v\|_{\mathcal{H}^{1/p+\varepsilon}}\end{aligned}$$

for any $\varepsilon > 0$ using that $\|H - \lfloor H\rfloor\|_{\mathcal{H}^\beta \to \mathcal{H}^\beta}$ is bounded by 1 for $\beta < 1$, which is true basically by construction together with Proposition 1.14; see also the proof of Proposition 3.6. Assuming that $\lambda \in \mathbb{N}$, however, the result follows directly in the same way by shifting $\lfloor H\rfloor$. For any $\lambda \geq 0$, we have

$$\|e^{it\lfloor H\rfloor}\Pi_\lambda u\|_{L^p([0,1],L^q)} = \|e^{it\lambda}\Pi_\lambda u\|_{L^p([0,1],L^q)} = \|\Pi_\lambda u\|_{L^q}$$

since the Weyl law guarantees that the number of eigenvalues in $[\lambda, \lambda+1)$ is finite. Thus we get using the Strichartz inequalities from Theorem 2.3

$$\|\Pi_\lambda u\|_{L^q} \lesssim \|\Pi_\lambda u\|_{\mathcal{H}^{1/p+\varepsilon}} \lesssim \sqrt{\lambda+1}^{\frac{1}{p}+\varepsilon}\|u\|_{L^2} \lesssim \sqrt{\lambda+1}^{\frac{1}{2}-\frac{1}{q}+\varepsilon}\|u\|_{L^2}$$

using again Proposition 1.14. □

As mentioned before, this is the point where there are slightly weaker results in the case of a surface with boundary. We use Theorem 2.6 instead.

Proposition 3.5. *Let $q \in (2,\infty)$ and M be a compact surface with boundary. We have*

$$\|e_n\|_{L^q} \lesssim \sqrt{\lambda_n(\Xi)}^{\frac{2}{3}-\frac{4}{3q}+\kappa}$$

for any $\kappa > 0$. In particular, this implies

$$\|e_n\|_{L^q} \lesssim (1+\sqrt{\lambda_n})^{\frac{2}{3}-\frac{4}{3q}+\kappa} \lesssim (1+\sqrt{n})^{\frac{2}{3}-\frac{4}{3q}+\kappa}.$$

Moreover, for the operator Π_λ we have

$$\|\Pi_\lambda u\|_{L^q} \lesssim \sqrt{\lambda+1}^{\frac{2}{3}-\frac{4}{3q}+\kappa}\|u\|_{L^2}$$

for any $\kappa > 0$.

Let B be the operator defined by

$$Be_n := \lfloor\sqrt{\lambda_n(\Xi)}\rfloor e_n$$

for any $n \geq 1$. The following proposition gives continuity estimates for the unitary groups associated to $\sqrt{H}$ and B and bound the difference between the two operators.

Proposition 3.6. *For any $\beta \in [0,1)$ and $t \in \mathbb{R}$, we have*

$$\|e^{it\sqrt{H}}u\|_{\mathcal{H}^\beta} \lesssim \|u\|_{\mathcal{H}^\beta},$$
$$\|e^{itB}u\|_{\mathcal{H}^\beta} \lesssim \|u\|_{\mathcal{H}^\beta}.$$

Moreover, the difference $B - \sqrt{H}$ is bounded on $\mathcal{H}^\beta$ for any $\beta \in [0,1)$ and the difference between the groups is given by

$$e^{itB}u - e^{it\sqrt{H}} = -i\int_0^t e^{i(t-s)B}(\sqrt{H}-B)e^{is\sqrt{H}}\,\mathrm{d}s.$$

Proof. We have

$$\|e^{it\sqrt{H}}v\|_{L^2} \lesssim \|v\|_{L^2}.$$

Thus

$$\|H^{\frac{\beta}{2}}e^{it\sqrt{H}}v\|_{L^2} = \|e^{it\sqrt{H}}H^{\frac{\beta}{2}}v\|_{L^2} \lesssim \|H^{\frac{\beta}{2}}v\|_{L^2}$$

for any $\beta \in (0, \alpha)$. Using Proposition 1.14, this gives

$$\|e^{it\sqrt{H}}v\|_{\mathcal{H}^\beta} \lesssim \|v\|_{\mathcal{H}^\beta}$$

and the result for e^{itB} follows from this. For the difference, $\|B-\sqrt{H}\|_{L^2\to L^2}$ is bounded by 1 and we have

$$\|H^{\frac{\beta}{2}}(B-\sqrt{H})u\|_{L^2} = \|(B-\sqrt{H})H^{\frac{\beta}{2}}u\|_{L^2} \le \|H^{\frac{\beta}{2}}u\|_{L^2}$$

and hence the boundedness of $B-\sqrt{H}$ on $\mathcal{H}^\beta$. The result on the difference of the groups

$$e^{itB}u - e^{it\sqrt{H}} = -i\int_0^t e^{i(t-s)B}(\sqrt{H}-B)e^{is\sqrt{H}}\,\mathrm{d}s$$

follows with the same reasoning as in Lemma 2.4. □

We now have all the ingredients to prove the Strichartz inequalities for the wave propagator associated to the Anderson Hamiltonian.

Theorem 3.7. *Let M be a compact surface without boundary $(p, q) \in [2, \infty)^2$ and $0 < \sigma < \alpha$ such that $p \le q$ and*

$$\sigma = \frac{3}{2} - \frac{2}{p} + \frac{1}{q}.$$

Then, for any $\kappa > 0$, we have the bound

$$\left\|\cos(t\sqrt{H})u_0 + \frac{\sin(t\sqrt{H})}{\sqrt{H}}u_1\right\|_{L^p([0,1],L^q)} \lesssim \|(u_0, u_1)\|_{\mathcal{H}^{\sigma+\kappa}\times\mathcal{H}^{\sigma-1+\kappa}}.$$

Proof. We start by proving the bound for e^{itB} using the spectral decomposition

$$e^{itB}u = \sum_{n\ge 0} e^{itn}\Pi_n u$$

and then bound the difference of the two groups. First, the condition $p \le q$ implies

$$\|e^{itB}u\|_{L^p([0,1],L^q(M))} \le \|e^{itB}u\|_{L^q(M,L^p([0,1]))};$$

hence it is enough to bound the right-hand side. Using the Sobolev embedding in the time variable and the L^q bound on the eigenvalues from Proposition 3.4, we have

$$\begin{aligned}
\|e^{itB}u\|^2_{L^q(M,L^p([0,1]))} &= \left\|\|e^{itB}u\|^2_{L^p([0,1])}\right\|_{L^{q/2}(M)} \lesssim \left\|\|e^{itB}u\|^2_{\mathcal{H}^{1/2-1/p}([0,1])}\right\|_{L^{q/2}(M)} \\
&\lesssim \sum_{n\ge 0}\left\|\|e^{itn}\Pi_n u\|^2_{\mathcal{H}^{1/2-1/p}([0,1])}\right\|_{L^{q/2}(M)} \lesssim \sum_{n\ge 0}\|e^{itn}\|^2_{\mathcal{H}^{1/2-1/p}([0,1])}\|\Pi_n u\|^2_{L^q(M)} \\
&\lesssim \sum_{n\ge 0}(n+1)^{1-\frac{2}{p}}(\sqrt{n}+1)^{1-\frac{2}{q}+2\kappa}\|\Pi_n u\|^2_{L^2} \\
&\lesssim \|\sqrt{H}^{\frac{3}{2}-\frac{2}{p}-\frac{1}{q}+\kappa}u\|^2_{L^2} \lesssim \|u\|^2_{\mathcal{H}^{3/2-2/p-1/q+\kappa}},
\end{aligned}$$

which gives the result for B. To obtain the proof for $\sqrt{H}$, we use

$$e^{itB}u - e^{it\sqrt{H}} = -i\int_0^t e^{i(t-s)B}(\sqrt{H}-B)e^{is\sqrt{H}}\,\mathrm{d}s.$$

Indeed, this gives

$$\begin{aligned}\|e^{it\sqrt{H}}u\|_{L^p([0,1],L^q)} &\lesssim \|e^{itB}u\|_{L^p([0,1],L^q)} + \int_0^1 \|e^{i(t-s)B}(\sqrt{H}-B)e^{is\sqrt{H}}\|_{L^p([0,1],L^q)}\,\mathrm{d}s \\ &\lesssim \|u\|_{\mathcal{H}^{\sigma+\kappa}} + \int_0^1 \|(\sqrt{H}-B)e^{i(t-s)B}u\|_{\mathcal{H}^{\sigma+\kappa}}\,\mathrm{d}s \lesssim \|u\|_{\mathcal{H}^{\sigma+\kappa}}\end{aligned}$$

for any $\kappa > 0$. The proof is directly completed from

$$\cos(t\sqrt{H}) = \frac{e^{it\sqrt{H}} + e^{-it\sqrt{H}}}{2} \quad \text{and} \quad \frac{\sin(\sqrt{H})}{\sqrt{H}} = \frac{e^{it\sqrt{H}} - e^{-it\sqrt{H}}}{2i\sqrt{H}}. \qquad \square$$

Again, the inhomogeneous inequalities follow directly and we omit the proof.

Corollary 3.8. *Let p, q, σ be as in Theorem 3.7. Then we have the bound*

$$\left\|\int_0^t \frac{\sin((t-s)\sqrt{H})}{\sqrt{H}} f(s)\right\|_{L^p([0,1],L^q)} \lesssim \int_0^1 \|f(s)\|_{\mathcal{H}^{\sigma-1+\kappa}}\,\mathrm{d}s$$

for $f \in L^1([0,1], \mathcal{H}^{\sigma-1+\kappa})$.

Moreover, we have the analogous result for surfaces with boundary, which is proved analogously by using Proposition 3.5 instead of Proposition 3.4.

Theorem 3.9. *Let M be a compact surface with boundary and $p, q \in [2, \infty)$ such that $p \le q$ and*

$$\sigma = \frac{5}{3} - \frac{2}{p} - \frac{4}{3q} \in (0, \alpha).$$

Then for any $\kappa > 0$, we have the bound

$$\left\|\cos(t\sqrt{H})u_0 + \frac{\sin(t\sqrt{H})}{\sqrt{H}}u_1 + \int_0^t \frac{\sin((t-s)\sqrt{H})}{\sqrt{H}} v\right\|_{L^p([0,1],L^q)} \lesssim \|(u_0, u_1)\|_{\mathcal{H}^{\sigma+\kappa}\times\mathcal{H}^{\sigma-1+\kappa}} + \|v\|_{L^1([0,1],\mathcal{H}^{\sigma-1+\kappa})}$$

for initial data $(u_0, u_1) \in \mathcal{H}^\sigma \times \mathcal{H}^{\sigma-1}$ and inhomogeneity $v \in L^1([0,1], \mathcal{H}^{\sigma-1+\kappa})$.

3C. *Local well-posedness for the multiplicative cubic stochastic wave equation.* Now we use the results from the previous section to prove local well-posedness of stochastic multiplicative wave equations of the form

$$\begin{cases} \partial_t^2 u + Hu = -u|u|^2, \\ (u, \partial_t u)|_{t=0} = (u_0, u_1) \end{cases}$$

in a low-regularity regime on general two-dimensional surfaces with or without boundary. While we have the classical Sobolev embedding

$$\mathcal{H}^\nu \hookrightarrow L^{\frac{2}{1-\nu}}$$

for $\nu \in [0, 1)$, we also make use of the dual Sobolev bound,

$$\text{for all } \sigma \in (0, 1], \quad L^{\frac{2}{2-\sigma}} \hookrightarrow \mathcal{H}^{\sigma-1}$$

which is true on general manifolds, see for example the book by Aubin [2]. Using this, we make a preliminary computation meant to show how far we get by using *only the Sobolev embedding result*. Then we will see how the bounds in Theorem 3.7 give better results on general manifolds. We first rewrite the equation under the mild formulation

$$u(t) = \cos(t\sqrt{H})u_0 + \frac{\sin(t\sqrt{H})}{\sqrt{H}}u_1 + \int_0^t \frac{\sin((t-s)\sqrt{H})}{\sqrt{H}}u(s)^3 \, \mathrm{d}s.$$

Then apply the dual Sobolev bound for $\sigma \in (0, 1]$ and $p = \frac{2}{2-\sigma} \in (1, 2]$ to get

$$\begin{aligned}\|u(t)\|_{\mathcal{H}^\sigma} &\lesssim \|u_0\|_{\mathcal{H}^\sigma} + \|u_1\|_{\mathcal{H}^{\sigma-1}} + \|u^3\|_{L^1([0,t],\mathcal{H}^{\sigma-1})} \\ &\lesssim \|u_0\|_{\mathcal{H}^\sigma} + \|u_1\|_{\mathcal{H}^{\sigma-1}} + \|u^3\|_{L^1([0,t],L^p)} \\ &\lesssim \|u_0\|_{\mathcal{H}^\sigma} + \|u_1\|_{\mathcal{H}^{\sigma-1}} + \|u\|_{L^\infty([0,t],L^{2/(1-\sigma)})}\|u\|^2_{L^2([0,t],L^4)},\end{aligned}$$

having applied Hölder with $\frac{1}{2} + \frac{1-\sigma}{2} = \frac{2-\sigma}{2}$. Finally, the Sobolev embedding gives

$$\|u(t)\|_{\mathcal{H}^\sigma} \lesssim \|u_0\|_{\mathcal{H}^\sigma} + \|u_1\|_{\mathcal{H}^{\sigma-1}} + \|u\|_{L^\infty([0,t],\mathcal{H}^\sigma)}\|u\|^2_{L^2([0,t],\mathcal{H}^{1/2})}.$$

This can then lead to a solution by fixed point by choosing $\sigma \geq \frac{1}{2}$. Clearly this is can be improved by using more subtle bounds than the Sobolev embedding. The Strichartz inequalities from the previous section allow us to get local well-posedness below; this is the content of the following theorems. As before we separately state the cases of surfaces without boundary and with boundary, which are proved in precisely the same way, just using Theorems 3.7 and 3.9 respectively.

Theorem 3.10. *Let M be a compact surface without boundary and $\sigma \in \left(\frac{1}{4}, \frac{1}{2}\right)$ and $\delta > 0$ sufficiently small. Then for any initial data $(u_0, u_1) \in \mathcal{H}^\sigma \times \mathcal{H}^{\sigma-1}$ there exists a time $T > 0$ depending on the data such that there exists a unique solution*

$$u \in C([0, T], \mathcal{H}^\sigma) \cap L^{\frac{2}{1-\delta}}([0, T], L^4)$$

to the mild formulation of the multiplicative cubic stochastic wave equation. Moreover, the solution depends continuously on the initial data (u_0, u_1).

Proof. As usual, this is proved in a standard way using the Banach fixed-point theorem. Define the map

$$\Psi(u)(t) := \cos(t\sqrt{H})u_0 + \frac{\sin(t\sqrt{H})}{\sqrt{H}}u_1 + \int_0^t \frac{\sin((t-s)\sqrt{H})}{\sqrt{H}}u(s)^3 \, \mathrm{d}s.$$

For $t > 0$, we have as above

$$\begin{aligned}\|u(t)\|_{\mathcal{H}^\sigma} &\lesssim \|u_0\|_{\mathcal{H}^\sigma} + \|u_1\|_{\mathcal{H}^{\sigma-1}} + \|u\|_{L^\infty([0,t],\mathcal{H}^\sigma)}\|u\|^2_{L^2([0,t],L^4)} \\ &\lesssim \|u_0\|_{\mathcal{H}^\sigma} + \|u_1\|_{\mathcal{H}^{\sigma-1}} + t^\delta\|u\|_{L^\infty([0,t],\mathcal{H}^\sigma)}\|u\|^2_{L^{2/(1-\delta)}([0,t],L^4)}\end{aligned}$$

using Hölder inequality in the last line for $\delta \in (0, 1)$. We then apply Theorem 3.7 with $p = \frac{2}{1-\delta}$ and $q = 4$ and obtain

$$\begin{aligned}\|\Psi(u)\|_{L^{2/(1-\delta)}([0,T],L^4)} &\lesssim \|u_0\|_{\mathcal{H}^{3/2-(1-\delta)-1/4+\kappa}} + \|u_1\|_{\mathcal{H}^{1/2-(1-\delta)-1/4+\kappa}} + \|u^3\|_{L^1([0,T],\mathcal{H}^{1/2-(1-\delta)-1/4+\kappa})} \\ &\lesssim \|u_0\|_{\mathcal{H}^\sigma} + \|u_1\|_{\mathcal{H}^{\sigma-1}} + \|u^3\|_{L^1([0,T],\mathcal{H}^{\sigma-1})}\end{aligned}$$

using that $\sigma > \frac{1}{4}$ and $\delta < \sigma - \frac{1}{4}$ gives $\frac{3}{2} - (1-\delta) - \frac{1}{4} + \kappa \le \sigma$ for $\kappa > 0$ small enough. Finally, we get

$$\|\Psi(u)\|_{L^{2/(1-\delta)}([0,T],L^4)} \lesssim \|u_0\|_{\mathcal{H}^\sigma} + \|u_1\|_{\mathcal{H}^{\sigma-1}} + T^\delta \|u\|_{L^\infty([0,T],\mathcal{H}^\sigma)} \|u\|^2_{L^{2/(1-\delta)}([0,T],L^4)}$$

as above. Thus we can get a fixed point in

$$C([0,T],\mathcal{H}^\sigma) \cap L^{\frac{2}{1-\delta}}([0,T],L^4)$$

in the usual way for $T > 0$ small enough. □

In a completely analogous way we get the following result for the case of surfaces with boundary using the Strichartz estimates from Theorem 3.9.

Theorem 3.11. *Let M be a compact surface with boundary and $\sigma \in \left(\frac{1}{3}, \frac{1}{2}\right)$ and $\delta > 0$ sufficiently small. Then for any initial data $(u_0, u_1) \in \mathcal{H}^\sigma \times \mathcal{H}^{\sigma-1}$ there exists a time $T > 0$ depending on the data such that there exists a unique solution*

$$u \in C([0,T],\mathcal{H}^\sigma) \cap L^{\frac{2}{1-\delta}}([0,T],L^4)$$

to the mild formulation of the multiplicative cubic stochastic wave equation. Moreover, the solution depends continuously on the initial data (u_0, u_1).

References

[1] R. Allez and K. Chouk, "The continuous Anderson Hamiltonian in dimension two", preprint, 2015. arXiv 1511.02718

[2] T. Aubin, *Some nonlinear problems in Riemannian geometry*, Springer, 1998. MR Zbl

[3] I. Babuška, "Error-bounds for finite element method", *Numer. Math.* **16**:4 (1971), 322–333. MR Zbl

[4] I. Bailleul and F. Bernicot, "Heat semigroup and singular PDEs", *J. Funct. Anal.* **270**:9 (2016), 3344–3452. MR Zbl

[5] I. Bailleul and F. Bernicot, "High order paracontrolled calculus", *Forum Math. Sigma* **7** (2019), art. id. e44. MR Zbl

[6] I. Bailleul, F. Bernicot, and D. Frey, "Space-time paraproducts for paracontrolled calculus, 3D-PAM and multiplicative Burgers equations", *Ann. Sci. École Norm. Sup.* (4) **51**:6 (2018), 1399–1456. MR Zbl

[7] M. D. Blair, H. F. Smith, and C. D. Sogge, "Strichartz estimates for the wave equation on manifolds with boundary", *Ann. Inst. H. Poincaré C Anal. Non Linéaire* **26**:5 (2009), 1817–1829. MR Zbl

[8] M. D. Blair, H. F. Smith, and C. D. Sogge, "Strichartz estimates and the nonlinear Schrödinger equation on manifolds with boundary", *Math. Ann.* **354**:4 (2012), 1397–1430. MR Zbl

[9] J.-M. Bony, "Calcul symbolique et propagation des singularités pour les équations aux dérivées partielles non linéaires", *Ann. Sci. École Norm. Sup.* (4) **14**:2 (1981), 209–246. MR Zbl

[10] J. Bourgain, "Fourier transform restriction phenomena for certain lattice subsets and applications to nonlinear evolution equations, I: Schrödinger equations", *Geom. Funct. Anal.* **3**:2 (1993), 107–156. MR Zbl

[11] N. Burq and N. Tzvetkov, "Random data Cauchy theory for supercritical wave equations, I: Local theory", *Invent. Math.* **173**:3 (2008), 449–475. MR Zbl

[12] N. Burq, P. Gérard, and N. Tzvetkov, "Strichartz inequalities and the nonlinear Schrödinger equation on compact manifolds", *Amer. J. Math.* **126**:3 (2004), 569–605. MR Zbl

[13] N. Burq, G. Lebeau, and F. Planchon, "Global existence for energy critical waves in 3-D domains", *J. Amer. Math. Soc.* **21**:3 (2008), 831–845. MR Zbl

[14] P. D'Ancona, L. Fanelli, L. Vega, and N. Visciglia, "Endpoint Strichartz estimates for the magnetic Schrödinger equation", *J. Funct. Anal.* **258**:10 (2010), 3227–3240. MR Zbl

[15] A. Debussche and H. Weber, "The Schrödinger equation with spatial white noise potential", *Electron. J. Probab.* **23** (2018), art. id. 28. MR Zbl

[16] M. Gubinelli, "Controlling rough paths", *J. Funct. Anal.* **216**:1 (2004), 86–140. MR Zbl

[17] M. Gubinelli, P. Imkeller, and N. Perkowski, "Paracontrolled distributions and singular PDEs", *Forum Math. Pi* **3** (2015), art. id. e6. MR Zbl

[18] M. Gubinelli, H. Koch, and T. Oh, "Renormalization of the two-dimensional stochastic nonlinear wave equations", *Trans. Amer. Math. Soc.* **370**:10 (2018), 7335–7359. MR Zbl

[19] M. Gubinelli, B. Ugurcan, and I. Zachhuber, "Semilinear evolution equations for the Anderson Hamiltonian in two and three dimensions", *Stoch. Partial Differ. Equ. Anal. Comput.* **8**:1 (2020), 82–149. MR Zbl

[20] M. Gubinelli, H. Koch, and T. Oh, "Paracontrolled approach to the three-dimensional stochastic nonlinear wave equation with quadratic nonlinearity", *J. Eur. Math. Soc.* (online publication January 2023).

[21] M. Hairer, "A theory of regularity structures", *Invent. Math.* **198**:2 (2014), 269–504. MR Zbl

[22] X. Huang and C. D. Sogge, "Quasimode and Strichartz estimates for time-dependent Schrödinger equations with singular potentials", *Math. Res. Lett.* **29**:3 (2022), 727–761. MR Zbl

[23] C. Labbé, "The continuous Anderson Hamiltonian in $d \leq 3$", *J. Funct. Anal.* **277**:9 (2019), 3187–3235. MR Zbl

[24] J. L. Lebowitz, H. A. Rose, and E. R. Speer, "Statistical mechanics of the nonlinear Schrödinger equation", *J. Stat. Phys.* **50**:3-4 (1988), 657–687. MR Zbl

[25] T. J. Lyons, "Differential equations driven by rough signals", *Rev. Mat. Iberoam.* **14**:2 (1998), 215–310. MR Zbl

[26] L. Morin and A. Mouzard, "2D random magnetic Laplacian with white noise magnetic field", *Stochastic Process. Appl.* **143** (2022), 160–184. MR Zbl

[27] A. Mouzard, "Weyl law for the Anderson Hamiltonian on a two-dimensional manifold", *Ann. Inst. Henri Poincaré Probab. Stat.* **58**:3 (2022), 1385–1425. MR Zbl

[28] T. Oh, T. Robert, and N. Tzvetkov, "Stochastic nonlinear wave dynamics on compact surfaces", 2019. To appear in *Ann. H. Lebesgue*. arXiv 1904.05277

[29] T. Oh, T. Robert, N. Tzvetkov, and Y. Wang, "Stochastic quantization of Liouville conformal field theory", preprint, 2020. arXiv 2004.04194

[30] M. Reed and B. Simon, *Methods of modern mathematical physics*, *II*: *Fourier analysis*, *self-adjointness*, Academic Press, New York, 1975. MR Zbl

[31] D. W. Stroock, *An introduction to the analysis of paths on a Riemannian manifold*, Math. Surv. Monogr. **74**, Amer. Math. Soc., Providence, RI, 2000. MR Zbl

[32] N. Tzvetkov and N. Visciglia, "Global dynamics of the 2D NLS with white noise potential and generic polynomial nonlinearity", preprint, 2022. arXiv 2204.03280

[33] N. Tzvetkov and N. Visciglia, "Two dimensional nonlinear Schrödinger equation with spatial white noise potential and fourth order nonlinearity", *Stoch. Partial Differ. Equ. Anal. Comput.* (online publication April 2022).

[34] I. Zachhuber, "Strichartz estimates and low-regularity solutions to multiplicative stochastic NLS", preprint, 2019. arXiv 1911.01982

[35] I. Zachhuber, "Finite speed of propagation for the 2- and 3-dimensional multiplicative stochastic wave equation", preprint, 2021. arXiv 2110.08086

Received 26 Apr 2021. Revised 13 Apr 2022. Accepted 15 Jul 2022.

ANTOINE MOUZARD: antoine.mouzard@ens-rennes.fr
Université de Rennes, CNRS, IRMAR-UMR 6625, Rennes, France

IMMANUEL ZACHHUBER: immanuel.zachhuber@fu-berlin.de
Freie Universität Berlin, Berlin, Germany

ANALYSIS AND PDE
Vol. 17 (2024), No. 2, pp. 455–498
DOI: 10.2140/apde.2024.17.455

CURVEWISE CHARACTERIZATIONS OF MINIMAL UPPER GRADIENTS AND THE CONSTRUCTION OF A SOBOLEV DIFFERENTIAL

SYLVESTER ERIKSSON-BIQUE AND ELEFTERIOS SOULTANIS

We represent minimal upper gradients of Newtonian functions, in the range $1 \leq p < \infty$, by maximal directional derivatives along "generic" curves passing through a given point, using plan-modulus duality and disintegration techniques. As an application we introduce the notion of p-weak charts and prove that every Newtonian function admits a differential with respect to such charts, yielding a linear approximation along p-almost every curve. The differential can be computed curvewise, is linear, and satisfies the usual Leibniz and chain rules.

The arising p-weak differentiable structure exists for spaces with finite Hausdorff dimension and agrees with Cheeger's structure in the presence of a Poincaré inequality. In particular, it exists whenever the space is metrically doubling. It is moreover compatible with, and gives a geometric interpretation of, Gigli's abstract differentiable structure, whenever it exists. The p-weak charts give rise to a finite-dimensional p-weak cotangent bundle and pointwise norm, which recovers the minimal upper gradient of Newtonian functions and can be computed by a maximization process over generic curves. As a result we obtain new proofs of reflexivity and density of Lipschitz functions in Newtonian spaces, as well as a characterization of infinitesimal Hilbertianity in terms of the pointwise norm.

1. Introduction

1A. ***Overview.*** Minimal weak upper gradients of Sobolev-type functions on metric measure spaces were first introduced by Cheeger [1999], building on the notion of upper gradients from [Heinonen and Koskela 1998]. Shanmugalingam [2000] developed *Newtonian spaces* $N^{1,p}(X)$ using the modulus perspective of [Heinonen and Koskela 1998] and proved that they coincide with the Sobolev space defined by Cheeger up to modification of its elements on a set of measure zero. Further notions of Sobolev spaces, based on test plans, were developed by Ambrosio, Gigli and Savaré [Ambrosio et al. 2014], with a corresponding

MSC2020: primary 46E36, 49J52; secondary 26B05, 30L99, 53C23.
Keywords: Sobolev, test plan, minimal upper gradient, differential structure, differential, chart.

notion of minimal gradient. Earlier, Hajłasz [1996] had introduced a Sobolev space whose associated minimal gradient, however, lacks suitable locality properties. While the various Sobolev spaces (with the exception of Hajłasz's definition) are equivalent for generic metric measure spaces, Newtonian spaces consist of representatives which are absolutely continuous along generic curves, a property central to the results in this paper.

The minimal p-weak upper gradient $g_f \in L^p(X)$ of a Newtonian function $f \in N^{1,p}(X)$ on a metric measure space X is a Borel function characterized (up to a null-set) as the minimal function satisfying

$$|(f \circ \gamma)'_t| \leq g_f(\gamma_t)|\gamma'_t| \quad \text{for a.e. } t \in I, \tag{1-1}$$

for all absolutely continuous $\gamma : I \to X$ outside a curve family of zero p-modulus. Here $|\gamma'_t|$ denotes the metric derivative of γ for a.e. t; see Section 2. When $X = \mathbb{R}^n$ and $f \in C_c^\infty(\mathbb{R}^n)$, g_f is given by $g_f = \|\nabla f\|$; in this case, for each $x \in X$, there exists a (smooth) gradient curve $\gamma : (-\varepsilon, \varepsilon) \to X$, with $\gamma_0 = x$, satisfying

$$(f \circ \gamma)'_0 = g_f(x)|\gamma'_0|. \tag{1-2}$$

In general, however, despite the minimality of g_f, the equality in (1-2) is not always attained. For example the fat Sierpiński carpet (with the Hausdorff 2-measure and Euclidean metric) constructed in [Mackay et al. 2013] with a sequence in $\ell^2 \setminus \ell^1$, as pointed out in the introduction of that work, gives zero p-modulus ($p > 1$) to the family of curves parallel to the x-axis, and thus to the family of gradient curves of the function $f(x, y) = x$. We remark that the example above is measure doubling and supports a Poincaré inequality; in this context an approximate form of (1-2) for Lipschitz functions was proven in [Cheeger and Kleiner 2009, Theorem 4.2].

Towards a positive answer for generic spaces, an "integral formulation" of (1-2) given by [Gigli 2015, Theorem 3.14] states that, when $p > 1$ and $f \in N^{1,p}(X)$, there exist probability measures $\boldsymbol{\eta}$ on $C(I; X)$ (known as test plans representing the gradient of f) such that

$$\lim_{t\to 0} \int \frac{f(\gamma_t) - f(\gamma_0)}{t} \, \mathrm{d}\boldsymbol{\eta} = \lim_{t\to 0} \int \frac{1}{t} \int_0^t g_f(\gamma_s)|\gamma'_s| \, \mathrm{d}s \, \mathrm{d}\boldsymbol{\eta}.$$

In this paper we obtain a "pointwise" variant of (1-2) for general metric measure spaces using a combination of plan-modulus duality, developed in [Ambrosio et al. 2015b; Durand-Cartagena et al. 2021; Honzlová Exnerová et al. 2021], and disintegration techniques. (For $p > 1$, a pointwise variant also follows from Gigli's integral formulation; see Section 3C.) Theorem 1.1 below expresses the minimal weak upper gradient of a Newtonian function as the supremum of directional derivatives along generic curves passing through a given point. Here, it is crucial to use Newtonian functions, which are absolutely continuous along almost every curve.

This curvewise characterization of minimal upper gradients yields the existence of an abundance of curves in a given region of the space, provided the region supports nontrivial Newtonian functions. The idea of constructing an abundance of curves goes back to [Semmes 1996] in the presence of a Poincaré inequality. Under this assumption Cheeger showed that $g_f = \operatorname{Lip} f$, where $\operatorname{Lip} f$ denotes the pointwise Lipschitz constant of a Lipschitz function f. Note that inequality $\operatorname{Lip} f \leq g$ for *continuous*

upper gradients g of a Lipschitz function f on a geodesic space is a direct, but central, observation made in [Cheeger 1999, pp. 432–433].

The work of Cheeger lead to many developments, including [Cheeger and Kleiner 2009] pioneering the idea of using directional derivatives along curves (and the early version of Theorem 1.1, appearing as Theorem 4.2 in that work) as well as the development and detailed analysis of *Lipschitz differentiability spaces*; see [Keith 2004a; Bate 2015; Bate and Speight 2013; Cheeger et al. 2016; Schioppa 2016a; 2016b]. In the latter, curves are replaced by *curve fragments* whose abundance is expressed using *Alberti representations*. Alberti representations are similar to *plans* used in this paper. The connection between such representations and the ideas in [Cheeger and Kleiner 2009] was first observed by Preiss, see [Bate 2015, p. 2], and can be used to prove the self-improvement of the Lip-lip inequality to the Lip-lip equality, see [Bate 2015; Schioppa 2016a; Cheeger et al. 2016].

Similarly the abundance of curves, obtained here using duality, yields geometric information on *Sobolev* functions on *general* metric measure spaces. (Indeed, duality, in the disguise of a minimax principle, was previously used to find Alberti representations in Lipschitz differentiability spaces; see [Bate 2015, Theorem 5.1] which uses [Rudin 1980, Lemma 9.4.3].) As an important first application, we use curvewise directional derivatives to define *p-weak charts* and a differential for Newtonian functions with respect to such charts. The arising *p-weak differentiable structure*, i.e., a covering by p-weak charts, exists far more generally than for Lipschitz differentiability spaces — indeed metric doubling and finite Hausdorff measure suffice; see Proposition 5.4. This existence result involves a new and crucial dimension bound for the charts and the induced differential structure; see Theorem 1.11(c) or Proposition 4.13. With the aid of Theorem 1.1 we adapt Cheeger's construction to produce a measurable L^∞-bundle, called the *p-weak cotangent bundle*, over spaces admitting a p-weak differentiable structure; differentials of Newtonian functions are sections over this bundle. While the Cheeger differential $\mathrm{d}_C f$ yields a linearization of a Lipschitz function f, our p-weak differential $\mathrm{d} f$ is given by a linearization along p-almost every curve, and the pointwise norm of $\mathrm{d} f$ recovers the minimal weak upper gradient; see Theorem 1.7.

This definition of a weak differentiable structure seems to be the natural extension of the seminal work [Cheeger 1999] to settings without a Poincaré inequality and yields a "partial differentiable structure", which has been the aim of many authors previously; see [Lučić et al. 2021; Alberti and Marchese 2016; Schioppa 2016a; Cheeger et al. 2016]. Namely, the p-weak cotangent bundle measures and makes precise the set of accessible directions (for positive modulus) in the space. By constructing the differential using directional derivatives along curves, we give it a concrete geometric meaning. A sequence of recent work has sought such concrete descriptions; see, e.g., [Ikonen et al. 2022; Lučić et al. 2021]. Our approach yields a new unification of the concrete and abstract cotangent modules of [Cheeger 1999] and [Gigli 2018], respectively; the p-weak cotangent bundle is compatible with Gigli's cotangent module when the latter is locally finitely generated, and with Cheeger's cotangent bundle when the space satisfies a Poincaré inequality; see Theorems 1.11 and 1.8.

The geometric approach in this paper has many natural applications. We mention here the *tensorization of Cheeger energy*, pursued in [Ambrosio et al. 2014; 2015c; Lučić et al. 2021], and the identification of abstract tangent bundles with geometric tangent cones; see [Alberti and Marchese 2016; Lučić et al.

2021]. Our methods give tools to generalize and refine the results mentioned above, and moreover enable a blow-up analysis to study analogues of *generalized linearity* considered for example in [Cheeger 1999; Cheeger et al. 2016]. Indeed, blow-ups of plans that define the pointwise norm on a p-weak chart (see Lemmas 4.1, 4.3 and 4.2) along a sequence of rescaled spaces yield curves in the limiting space along which limiting maps of rescaled Newtonian maps behave linearly. In this context we highlight [Schioppa 2016a], which gives a similar geometric and blow-up analysis in the context of abstract Weaver derivations. We leave the detailed exploration and development of these ideas for future work.

1B. *Curvewise characterization of minimal upper gradients.* Throughout the paper, we fix a *metric measure space* $X = (X, d, \mu)$, that is, a complete separable metric space (X, d) together with a Radon measure μ which is finite on bounded sets. A *plan* is a finite measure $\boldsymbol{\eta}$ on $C(I; X)$ that is concentrated on the set $\mathrm{AC}(I; X)$ of absolutely continuous curves. The natural evaluation map $e : C(I; X) \times I \to X$, $(\gamma, t) \mapsto \gamma_t$, gives rise to the *barycenter* $\mathrm{d}\boldsymbol{\eta}^{\#} := e_*(|\gamma_t'|\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\eta})$ of $\boldsymbol{\eta}$. If $\mathrm{d}\boldsymbol{\eta}^{\#} = \rho\,\mathrm{d}\mu$ for some $\rho \in L^q(\mu)$, we say that $\boldsymbol{\eta}$ is a q-plan (not to be confused with q-test plans, see, e.g., Section 2 or [Ambrosio et al. 2015b]). Every finite measure $\boldsymbol{\pi}$ on $C(I; X) \times I$ admits a disintegration with respect to e: for $e_*\boldsymbol{\pi}$-almost every $x \in X$, there exists a (unique) measure $\boldsymbol{\pi}_x \in \mathcal{P}(C(I; X) \times I)$, concentrated on $e^{-1}(x)$, such that the collection $\{\boldsymbol{\pi}_x\}$ satisfies

$$\boldsymbol{\pi}(B) = \int \boldsymbol{\pi}_x(B)\,\mathrm{d}(e_*\boldsymbol{\pi})(x)$$

for all Borel sets $B \subset C(I; X) \times I$. We refer to [Ambrosio et al. 2008; Bogachev 2007] for more details.

We use these notions to define a "generic curve": if $\boldsymbol{\eta}$ is a q-plan on X and $\{\boldsymbol{\pi}_x\}$ the disintegration of $\mathrm{d}\boldsymbol{\pi} := |\gamma_t'|\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\eta}$ with respect to e, then $\boldsymbol{\pi}_x$-a.e.-curve passes through x, for $e_*\boldsymbol{\pi}$-a.e. $x \in X$. In the forthcoming discussion, we omit the reference to e in the disintegration. We now formulate our first result, in which the equality in (1-2) is obtained as an essential supremum with respect to the disintegration for almost every point. In the statement below we write

$$\mathrm{Diff}(f) = \{(\gamma, t) \in \mathrm{AC}(I; X) \times I : f \circ \gamma \in \mathrm{AC}(I; \mathbb{R}),\ (f \circ \gamma)_t' \text{ and } |\gamma_t'| > 0 \text{ exist}\}$$

for a μ-measurable function $f : X \to [-\infty, \infty]$.

Theorem 1.1. *Let $1 \le p < \infty$, and let $1 < q \le \infty$ satisfy $1/p + 1/q = 1$. Suppose $f \in N^{1,p}(X)$, g_f is a Borel representative of the minimal p-weak upper gradient of f, and $D := \{g_f > 0\}$. There exists a q-plan $\boldsymbol{\eta}$ with $\mu|_D \ll \boldsymbol{\eta}^{\#}$ so that the disintegration $\{\boldsymbol{\pi}_x\}$ of $\mathrm{d}\boldsymbol{\pi} := |\gamma_t'|\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\eta}$ is concentrated on $e^{-1}(x) \cap \mathrm{Diff}(f)$ and*

$$g_f(x) = \left\| \frac{(f \circ \gamma)_t'}{|\gamma_t'|} \right\|_{L^\infty(\boldsymbol{\pi}_x)} \tag{1-3}$$

for μ-almost every $x \in D$.

Remark 1.2. The statement also holds when $f \in N^{1,p}_{\mathrm{loc}}(X)$, that is, $f|_{B(x,r)} \in N^{1,p}(B(x, r))$ for each ball $B(x, r) \subset X$. Indeed, a localization argument, replacing f by $f\eta_n$ with η_n a sequence of Lipschitz functions with bounded support and $\eta_n|_{B(x_0,n-1)} = 1$ for some x_0, reduces the statement for $f \in N^{1,p}_{\mathrm{loc}}(X)$ to Theorem 1.1. Similarly, other notions in this paper, such as charts, could use a local Sobolev space, but to avoid technicalities we do not discuss this point further. A reader can see Lemma 4.5 and its proof for a prototypical form of such a localization argument.

In particular, we have the following corollary.

Corollary 1.3. *Let p, q and f, g_f, D be as in Theorem 1.1. There exists a q-plan $\boldsymbol{\eta}$ and, for every $\varepsilon > 0$, a Borel set $B = B_\varepsilon \subset \mathrm{Diff}(f)$ with the following property: if $\{\boldsymbol{\pi}_x\}$ denotes the disintegration of $\mathrm{d}\boldsymbol{\pi} := |\gamma_t'|\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\eta}$, then $\boldsymbol{\pi}_x(B) > 0$ and*

$$(1-\varepsilon)g_f(x)|\gamma_t'| \le (f\circ\gamma)_t' \le g_f(x)|\gamma_t'| \quad \textit{for every } (\gamma, t) \in e^{-1}(x)\cap B,$$

for μ-a.e. $x \in D$.

Theorem 1.1 notably covers the case $p = 1$. In Section 3C we also prove a variant (Theorem 3.6) when $p > 1$, using test plans representing a gradient instead of plan-modulus duality.

1C. *Application: p-weak differentiable structure.* Cheeger [1999] showed that PI-spaces (metric measure spaces with a doubling measure supporting some Poincaré inequality) admit a countable cover by *Cheeger charts*, also called a *Lipschitz differentiable structure* (see [Keith 2004b]). Let $\mathrm{LIP}(X)$ denote the collection of Lipschitz functions on X, and let $\mathrm{LIP}_b(X)$ consist those Lipschitz functions with bounded support. A Cheeger chart (U, φ) of dimension n consists of a Borel set U with $\mu(U) > 0$, and a Lipschitz function $\varphi : X \to \mathbb{R}^n$ such that, for every $f \in \mathrm{LIP}(X)$ and μ-a.e. $x \in U$, there exists a unique linear map $\mathrm{d}_{C,x} f : \mathbb{R}^n \to \mathbb{R}$, called the Cheeger differential of f, such that

$$f(y) - f(x) = \mathrm{d}_{C,x} f(\varphi(y) - \varphi(x)) + o(d(x, y)) \quad \text{as } y \to x. \tag{1-4}$$

Not every space admits Lipschitz differentiable structure, as shown by the so called *Rickman's rug* $X := [0, 1]^2$ equipped with the metric $d((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + |y_1 - y_2|^\alpha$, where $\alpha \in (0, 1)$ and $\mu = \mathcal{L}^2|_X$. Indeed, a Weierstrass-type function in the y-variable combined with [Schioppa 2016a, Theorem 1.14] would yield nonhorizontal rectifiable curves if the space were a differentiability space, contradicting the fact that all rectifiable curves in X are horizontal.

Here, we introduce *p-weak differentiable structures*, which exist in much more generality (including Rickman's rug, see the discussion after Definition 1.4), adapting Cheeger's construction by substituting (1-4) for a weaker curvewise control. To accomplish this, we replace the pointwise Lipschitz constant by the minimal p-weak upper gradient in the definition of "infinitesimal linear independence" (1-5) and use Theorem 1.1 to circumvent the difficulties arising from the fact that the latter is defined only up to a null-set.

In the remainder of the introduction, we use the notation $|Df|_p$ for the minimal p-weak upper gradient of $f \in N^{1,p}_{\mathrm{loc}}(X)$ and refer to Section 2 for more discussion on this notation. Given $p \ge 1$ and $N \in \mathbb{N}$, we say that a Sobolev map $\varphi \in N^{1,p}_{\mathrm{loc}}(X; \mathbb{R}^N)$ is *p-independent* in $U \subset X$ if

$$\operatorname*{ess\,inf}_{v\in S^{N-1}} |D(v\cdot\varphi)|_p > 0 \quad \mu\text{-a.e. in } U, \tag{1-5}$$

and *p-maximal* in U if no *Lipschitz* map into a higher-dimensional target is p-independent in a positive measure subset of U. Here, we use the essential infimum of an uncountable collection, which agrees μ-a.e. with the pointwise infimum over any countable dense collection of S^{N-1}; see Section A2. Note that p-maximality does not depend on the particular map φ but rather the dimension of its target space.

Definition 1.4. An N-dimensional p-weak chart (U, φ) of X consists of a Borel set $U \subset X$ with positive measure and a Lipschitz function $\varphi : X \to \mathbb{R}^N$ which is p-independent and p-maximal in U. We say that X admits a p-weak differentiable structure if it can be covered up to a null set by countably many p-weak charts.

By convention, zero-dimensional p-weak charts satisfy $\varphi \equiv 0$ and (1-5) is a vacuous condition, while maximality means that $|Df|_p = 0$ μ-a.e. on U for every $f \in \mathrm{LIP}_b(X)$ (see also Proposition 4.4). In Section 4F we briefly discuss a lower-regularity requirement in Definition 1.4 and the fact the resulting notion yields essentially the same p-weak differentiable structure. We also show that an N-dimensional p-weak chart (U, φ) satisfies $N \le \dim_H U$, where $\dim_H U$ denotes the Hausdorff dimension of U; see Proposition 4.13. In particular, we have the following theorem.

Theorem 1.5. *A metric measure space of finite Hausdorff dimension admits a p-weak differentiable structure for any $p \ge 1$. In particular, this holds if the space is metrically doubling.*

We refer to Proposition 5.4 for a more technical statement, which immediately implies the theorem. Next, we give an analogue of the Cheeger differential (1-4) using p-weak charts.

Definition 1.6. Given an N-dimensional p-weak chart (U, φ) of X, a *p-weak differential* of a Newtonian function $f \in N^{1,p}(X)$ with respect to φ is a map $\mathrm{d}f : U \to (\mathbb{R}^N)^*$ (whose value at $x \in U$ is denoted by $\mathrm{d}_x f$) which satisfies

$$f(\gamma_s) - f(\gamma_t) = \mathrm{d}_{\gamma_t} f(\varphi(\gamma_s) - \varphi(\gamma_t)) + o(|t-s|) \quad \text{for a.e. } t \in \gamma^{-1}(U), \text{ as } s \to t, \tag{1-6}$$

for p-a.e. absolutely continuous curve γ in X. We say that a function $f \in N^{1,p}(X)$ has a p-weak differential with respect to φ, if such a $\mathrm{d}f$ exists.

If the curve γ does not enter U, or only spends zero length in the set, then condition (1-6) becomes vacuously satisfied with both sides vanishing. The p-weak differential is uniquely determined up to almost everywhere equivalence by (1-6); see Lemma 4.3. Further, it is also local, i.e., if $g \in N^{1,p}(X)$ and $f|_A = g|_A$ on a positive measure subset $A \subset U$, then $\mathrm{d}f|_A = \mathrm{d}g|_A$. The differential satisfies various natural computation rules; see Propositions 4.10 and 5.7 for the most important ones.

Theorem 1.7. *Suppose $p \ge 1$, and $\varphi : X \to \mathbb{R}^N$ is a p-weak chart on U. Then any $f \in N^{1,p}(X)$ has a p-weak differential $\mathrm{d}f : U \to (\mathbb{R}^N)^*$ with respect to φ, which is μ-a.e. unique, and the map $f \mapsto \mathrm{d}f$ is linear.*

Moreover, for μ-a.e. $x \in U$, there is a norm $|\cdot|_x$ on $(\mathbb{R}^N)^*$ such that $x \mapsto |\xi|_x$ is Borel for every $\xi \in (\mathbb{R}^N)^*$ and

$$|\mathrm{d}f|_x = |Df|_p(x) \quad \textit{for } \mu\textit{-a.e. } x \in X,$$

for every $f \in N^{1,p}(X)$.

Whereas Lipschitz functions are differentiable with respect to Cheeger charts, (1-5) yields only the curvewise control (1-6). Indeed, if there are very few or no rectifiable curves, or if the curves only point into certain directions, then the p-weak differential vanishes, or measures only these directions, respectively. For example, given a fat Cantor set $K \subset \mathbb{R}^n$ with $\mathcal{L}^n(K) > 0$, $X := (K, d_{\mathrm{Eucl}}, \mathcal{L}^n|_K)$ is a

Lipschitz differentiability space but the minimal weak upper gradient of every Lipschitz function is zero. On the other hand, Rickman's rug admits nontrivial p-weak charts $\varphi(x, y) = x$. The p-weak differential in this case can be identified with the x-derivative, $\mathrm{d}f \equiv \partial_x f$, and the only curves with positive p-modulus are those which are horizontal. These examples demonstrate that p-weak differentiable structures might exist for spaces not admitting a Cheeger structure, but the two need not coincide even if both exist. However, if a Poincaré inequality is present, the two structures coincide.

Theorem 1.8. *Suppose X is a p-PI space for $p \geq 1$. Then any p-weak chart (U, φ) of X is a Cheeger chart.*

It follows from the discussion after Definition 1.4 that a p-PI space admits p-weak charts. In Section 1D, we obtain a precise statement on the relationship between the p-weak and Lipschitz differentiable structure, as well as a characterization of the existence of p-weak differentiable structures in terms of Gigli's cotangent module [2018]. Here we mention a noteworthy corollary of the existence of a p-weak differentiable structure.

Theorem 1.9. *Let $p \geq 1$. If X admits a p-weak differential structure, then $\mathrm{LIP}_b(X)$ is norm-dense in $N^{1,p}(X)$.*

Theorem 1.9 has been obtained by other methods for $p > 1$ in [Ambrosio et al. 2013] but is new in the case $p = 1$. In particular, we highlight that the density holds if X has finite Hausdorff dimension.

1D. *Connections to Cheeger's and Gigli's differentiable structures.* Together with the pointwise norm from Theorem 1.7, a p-weak differentiable structure gives rise to a *p-weak cotangent bundle* T_p^*X over X, analogous to the measurable L^∞-cotangent bundle T_C^*X arising from the Lipschitz differentiable structure [Cheeger 1999; Keith 2004b], which is equipped with the pointwise norm

$$|\xi|_{C,x} := \mathrm{Lip}(\xi \circ \varphi)(x), \quad \xi \in (\mathbb{R}^N)^*,$$

for μ-a.e. $x \in U$, where (U, φ) is an N-dimensional Cheeger chart. For any $f \in \mathrm{LIP}_b(X)$, the differentials $\mathrm{d}f$ and $\mathrm{d}_C f$ are sections of the cotangent bundles T_p^*X and T_C^*X, respectively. We refer to Section 5 for the precise definition of measurable L^∞-bundles and their sections.

In the next theorem we show that there is a submetric bundle map $T_C^*X \to T_p^*X$ and give a condition under which the bundle map is an isometric isomorphism. See Section 5 for the definition of bundle maps. In the statement, a modulus of continuity is an increasing continuous function $\omega : [0, \infty) \to [0, \infty)$, with $\omega(0) = 0$, and a linear submetry between normed spaces V and W is a surjective linear map $L : V \to W$, with $L(B_V(r)) = B_W(r)$.

Theorem 1.10. *Suppose X admits a Cheeger structure and let $p \geq 1$. There is a bundle map $\pi = \pi_{C,p} : T_C^*X \to T_p^*X$ which is a linear submetry μ-a.e. and satisfies*

$$\pi_x(\mathrm{d}_{C,x} f) = \mathrm{d}_x f \quad \text{for } \mu\text{-a.e. } x \in X, \tag{1-7}$$

for every $f \in \mathrm{LIP}_b(X)$. If there exists a collection $\{\omega_x\}_{x \in X}$ of moduli of continuity satisfying

$$\mathrm{Lip}\, f(x) \leq \omega_x(|Df|_p(x)) \quad \text{for } \mu\text{-a.e. on } X,$$

for every $f \in \mathrm{LIP}_b(X)$, then $\pi_{C,p}$ is an isometric bijection μ-a.e.

Theorem 1.10 follows from [Ikonen et al. 2022, Theorem 1.1] and the following theorem, which identifies the space $\Gamma_p(T_p^*X)$ of p-integrable sections of the p-weak cotangent bundle T_p^*X with Gigli's cotangent module $L^p(T^*X)$. We refer to Section 6 for the relevant definitions, and remark here that Gigli's construction is the most general in the sense that $L^p(T^*X)$ can be defined for any metric measure space. It is a priori defined only as an abstract L^p-normed L^∞-module in the sense of [Gigli 2015; 2018].

We say that $L^p(T^*X)$ is *locally finitely generated* if X has a countable Borel partition $\mathscr{B}$ so that each $B \in \mathscr{B}$ admits a finite generating set in B. Here, a collection $\mathscr{V} \subset L^p(T^*X)$ is a generating set in B (or generates $L^p(T^*X)$ in B) if $\chi_B L^p(T^*X)$ is the smallest closed submodule of $L^p(T^*X)$ containing $\chi_B v$ for every $v \in \mathscr{V}$. Gigli's cotangent modules admit a *dimensional decomposition*, i.e., a Borel partition $\{A_N\}_{N\in\mathbb{N}\cup\{\infty\}}$ of X so that $L^p(T^*X)$ admits a generating set of cardinality N (and no smaller) in A_N for each N. For $N=\infty$, no finite set generates $L^p(T^*X)$ in A_N. The dimensional decomposition is uniquely determined up to μ-negligible sets.

Below we denote by $\mathrm{d}_G f$ and $|\cdot|_G$ the abstract differential and pointwise norm in the sense of Gigli; see Theorem 6.1. A morphism between L^p-normed L^∞-modules (i.e., a continuous L^∞-linear map) is said to be an isometric isomorphism if it preserves the pointwise norm and has an inverse that is a morphism.

Theorem 1.11. *Let X be a metric measure space and $p \geq 1$. Then X admits a p-weak differentiable structure if and only if $L^p(T^*X)$ is locally finitely generated. In this case,*

(a) *there exists an isometric isomorphism $\iota : \Gamma_p(T_p^*X) \to L^p(T^*X)$ of normed modules satisfying $\iota(\mathrm{d}f) = \mathrm{d}_G f$ for every $f \in N^{1,p}(X)$ and uniquely determined by this property,*

(b) *each set A_N in the dimensional decomposition of X can be covered up to a null-set by N-dimensional p-weak charts,*

(c) *$N \leq \dim_H(A_N)$ for each $N \in \mathbb{N}$.*

Theorem 1.11 gives a concrete interpretation of Gigli's cotangent module, and bounds the Hausdorff dimension of the sets in the dimensional decomposition. As corollaries we obtain the reflexivity of $N^{1,p}(X)$ when $p > 1$, and a characterization of infinitesimal Hilbertianity in terms of the pointwise norm of Theorem 1.7 when $p = 2$, for spaces admitting a p-weak differentiable structure; see Corollary 6.7. Reflexivity could also be obtained directly from Theorem 1.7 following the argument in [Cheeger 1999, Section 4].

2. Preliminaries

Throughout this paper $X = (X, d, \mu)$ will be a complete separable metric measure space equipped with a Radon measure μ finite on balls. We denote by $C(I; X)$ the space of continuous curves $\gamma : I \to X$ equipped with the metric of uniform convergence and by $\mathrm{AC}(I; X)$ the subset of absolutely continuous curves in X, where $I \subset \mathbb{R}$ is an interval. Mostly, we will be concerned with statements independent of parametrization; thus the choice of the interval I is immaterial. However, when we need to refer to the end points of the curve, then we will take $I = [0, 1]$.

If γ is a curve, its value at $t \in I$ is denoted by $\gamma_t := \gamma(t)$. If $f : X \to \mathbb{R}^N$ is a function, we also use this notation as $(f \circ \gamma)_t = f(\gamma_t)$. The derivative of f in the direction of γ at γ_t, when it exists, is denoted

by $(f \circ \gamma)'_t = (f \circ \gamma)'(t)$. The metric derivative of the curve, in the sense of say [Ambrosio et al. 2008, Section I.1], is defined as $|\gamma'_t| = \lim_{h\to 0} d(\gamma_{t+h}, \gamma_t)/h$, when it exists. The metric derivative is defined almost everywhere on I for $\gamma \in \mathrm{AC}(I; X)$.

2A. ***Plans and modulus.*** A finite measure $\boldsymbol{\eta}$ on $C(I; X)$ is called a *plan* if it is concentrated on $\mathrm{AC}(I; X)$, and a *q-plan* if the *barycenter* $\mathrm{d}\boldsymbol{\eta}^{\#} := e_*(|\gamma'_t|\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\eta})$ satisfies $\mathrm{d}\boldsymbol{\eta}^{\#} = \rho\,\mathrm{d}\mu$ for some $\rho \in L^q(\mu)$. We denote by $AC_q(I; X)$ the space of curves $\gamma \in \mathrm{AC}(I; X)$ satisfying $\int_0^1 |\gamma'_t|^q\,\mathrm{d}t < \infty$, and say that a q-plan $\boldsymbol{\eta} \in \mathcal{P}(C(I; X))$ is a *q-test plan*, if it is concentrated on $AC_q(I; X)$ and

$$e_{t*}\boldsymbol{\eta} \le C\mu \quad \text{for every } t \in I, \quad \text{and} \quad \iint_0^1 |\gamma'_t|^q\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\eta} < \infty$$

for some constant $C > 0$. Here $e_t : C(I; X) \to X$ is the map $e_t(\gamma) = \gamma_t$.

Remark 2.1. Every q-test plan is also a q-plan. However, the converse can fail for two reasons. A q-test plan fixes a given parametrization for curves (with an integrability condition on the speed) and insists on a compression bound $e_{t*}(\boldsymbol{\eta}) \le C\mu$. However, for each q-plan supported on $\Gamma \subset \mathrm{AC}(I; X)$, one can construct associated q-test plans supported on reparametrized curves, which are subcurves of curves in Γ.

The argument for this is a combination of two observations in [Ambrosio et al. 2015b]. First, for each q-plan one can reparametrize curves to get a plan with a good "parametric barycenter" [loc. cit., Definition 8.1 and Theorem 8.3]. The parametric barycenter depends on the parametrization, while the barycenter $\boldsymbol{\eta}^{\#}$ does not. The second point concerns the compression bound, where given the previous plan, one can take subsegments of curves and average these over shifts to get a compression bound, which is explained as part of the proof of [loc. cit., Theorem 9.4].

This remark would allow, for example, to phrase Theorem 1.1 with test plans instead of plans, if one were so inclined.

If $\Gamma \subset C(I; X)$ is a family of curves, then a Borel function $\rho\colon X \to [0, \infty]$ is called admissible if $\int_\gamma \rho\,\mathrm{d}s \ge 1$ for each rectifiable $\gamma \in \Gamma$. In particular, if there are no rectifiable curves, then this condition is vacuous. We define, for $p \in [1, \infty)$,

$$\mathrm{Mod}_p(\Gamma) = \inf_\rho \int_X \rho^p\,\mathrm{d}\mu,$$

where the infimum is over all admissible ρ. We remark, that due to Vitali–Carathéodory, such an infimum can always be taken with respect to lower semicontinuous functions. Notice that the modulus is supported on rectifiable curves and is independent of the parametrization of such curves. We say that a property holds for *p-almost every curve* if there is a family of curves Γ_B so that $\mathrm{Mod}_p(\Gamma_B) = 0$ and the property holds for all $\gamma \in C(I; X) \setminus \Gamma_B$. Modulus is invariant of the parametrization of curves, but some of our statements depend on a parametrization. In those cases, we will say that the property holds for *p-almost every absolutely continuous curve in X* (or p-a.e. $\gamma \in \mathrm{AC}(I; X)$) to emphasize that the property holds for each $\gamma \in \mathrm{AC}(I; X) \setminus \Gamma_B$ with $\mathrm{Mod}_p(\Gamma_B) = 0$. The reader may consult [Heinonen et al. 2015, Sections 4–7] for a more in-depth treatment of modulus, upper gradients and Vitali–Carathéodory.

Remark 2.2. A crucial fact we will use is that if Γ satisfies $\mathrm{Mod}_p(\Gamma) = 0$, then for any q-plan $\boldsymbol{\eta}$ we have $\boldsymbol{\eta}(\Gamma) = 0$ (which holds for $p \in [1, \infty)$ and q its dual exponent). The converse is also true for $p \in (1, \infty)$.

See the arguments and discussion in [Ambrosio et al. 2015b, Sections 4 and 9]. One point here is that if we used q-test plans, this relationship would be more complex, and we would need to consider "stable" families of curves; see [loc. cit., Theorem 9.4]. The case of $p = 1$ is also somewhat subtle, and we will deal with a special case of this issue in Section 3. The argument of Proposition 2.3 would give the converse for compact families of curves and $p = 1$. See also, [Honzlová Exnerová et al. 2021] for a much more detailed exploration of this borderline case.

The previous remark concerns an inequality relating modulus and q-plans. However, there is a closer connection, and in a sense these are dual to each other. Previously, this has been explored in [Ambrosio et al. 2015b, Theorem 5.1] for $p > 1$, and in [Honzlová Exnerová et al. 2021, Theorem 6.3] for $p = 1$. Due to its importance for us, we summarize one main consequence of these results. We further briefly describe the main steps of a direct proof from [David and Eriksson-Bique 2020, Proposition 4.5]. A similar argument appeared previously in a more specific context in [Durand-Cartagena et al. 2021, Theorem 3.7].

Proposition 2.3. *Let $p \in [1, \infty)$ and q its dual exponent with $p^{-1} + q^{-1} = 1$. If $K \subset C(I; X)$ is a compact family of curves, and $\mathrm{Mod}_p(K) \in (0, \infty)$, then there exists a q-plan $\boldsymbol{\eta}$ with $\mathrm{spt}(\boldsymbol{\eta}) \subset K$.*

Proof. A power of the modulus $\mathrm{Mod}_p(K)^{1/p}$ arises from a convex optimization problem on ρ with a constraint for every curve $\gamma \in K$. A dual formulation of this corresponds to a variable for each constraint, i.e., a measure ν supported on K. Thus, it is reasonable to consider a modified Lagrangian defined by

$$\Phi(\rho, \nu) = \|\rho\|_{L^p} - \mathrm{Mod}_p(K)^{1/p} \int_K \int_\gamma \rho \, \mathrm{d}s \, \mathrm{d}\nu_\gamma,$$

where $\rho : X \to [0, \infty]$ is a function and ν is a probability measure supported on K. Let $P(K)$ be the collection of these probability measures supported on K equipped with the topology of weak* convergence. In order to obtain the required continuity, we will restrict to $\rho \in G$, with

$$G := \{\rho : X \to [0, 1] : \rho \text{ compactly supported and continuous in } X\}.$$

The set G is equipped with the topology of uniform convergence. Then $\Phi : G \times P(K) \to \mathbb{R}$ is a functional with two properties: $\Phi(\cdot, \nu)$ is convex and continuous for each $\nu \in P(K)$, and $\Phi(\rho, \cdot)$ is concave and upper semicontinuous for each $\rho : X \to [0, 1]$. Further $P(K)$ is compact and convex in the weak* topology and G is a convex subset.

By Sion's minimax theorem, see, e.g., statement in [David and Eriksson-Bique 2020, Theorem 4.7], we have

$$\sup_{\nu \in P(K)} \inf_{\rho \in G} \Phi(\rho, \nu) = \inf_{\rho \in G} \sup_{\nu \in P(K)} \Phi(\rho, \nu).$$

We can compute $\inf_{\rho \in G} \sup_{\nu \in P(K)} \Phi(\rho, \nu) \geq 0$. Indeed, given any $\rho \in G$, we can use the definition of modulus to find a $\gamma \in K$ with $\int_\gamma \rho \, \mathrm{d}s \leq \|g\|_p / \mathrm{Mod}_p(K)^{-1/p}$. If we choose $\nu = \delta_\gamma$, a Dirac measure on γ, the bound immediately follows.

Therefore, we have also $\sup_{\nu \in P(K)} \inf_{\rho \in G} \Phi(\rho, \nu) \geq 0$. But, up to showing that this supremum is attained, there must be some $\boldsymbol{\eta} \in P(K)$ for which we get $\inf_{\rho \in G} \Phi(\rho, \boldsymbol{\eta}) \geq 0$. After unwinding the definition of a q-plan, and an application of Radon–Nikodym on X, the measure $\boldsymbol{\eta}$ is our desired q-plan. □

2B. *Sobolev spaces and functions.* A function $f : (X, d_X) \to (Y, d_Y)$ between two metric spaces is called Lipschitz if $\mathrm{LIP}(f) := \sup_{x,y\in X, x\neq y} d_Y(f(x), f(y))/d_X(x, y) < \infty$. A bijection $f : X \to Y$ is called bi-Lipschitz if f and f^{-1} are Lipschitz. Further, if $x \in X$, we define the local Lipschitz constant as

$$\mathrm{Lip}\, f(x) := \limsup_{y\to x, y\neq x} \frac{d_Y(f(x), f(y))}{d_X(x, y)}.$$

Let $\mathrm{LIP}_b(X)$ be the collection of Lipschitz maps $f : X \to \mathbb{R}$ with bounded support.

Definition 2.4. Let $f : X \to \mathbb{R} \cup \{\pm\infty\}$ be measurable, $g : X \to [0, \infty]$ a Borel function, and $\gamma : I \to X$ a rectifiable path. We say that g is an upper gradient of f along γ, if $\int_\gamma g\, \mathrm{d}s < \infty$ and

$$|f(\gamma_t) - f(\gamma_s)| \le \int_{\gamma|_{[s,t]}} g$$

for each $s < t$ with $s, t \in I$ with the convention $\infty - \infty = \infty$. We say that g is an upper gradient of f if it is an upper gradient along every rectifiable curve, and a p-weak upper gradient if g is an upper gradient of f along p-a.e. rectifiable curve.

The space $N^{1,p}(X)$ is defined as all μ-measurable functions $f \in L^p(X)$ which have an upper gradient g in $L^p(X)$. The (semi-)norm on this space is defined as

$$\|f\|_{N^{1,p}} = (\|f\|_{L^p}^p + \inf \|g\|_{L^p}^p)^{1/p},$$

where the infimum is taken over all L^p-integrable upper gradients g of f. The theory of these spaces was largely developed in [Shanmugalingam 2000]; see also [Heinonen et al. 2015] for most of the classical theory. By the results there combined with an observation of [Hajłasz 2003] in the case of $p = 1$, one can show that there always exists a unique minimal g_f, which is an upper gradient along p-almost every path, and for which $\|f\|_{N^{1,p}} = (\|f\|_{L^p}^p + \|g_f\|_{L^p}^p)^{1/p}$. We call g_f the *minimal p-upper gradient*. Similarly, we can define $f \in N^{1,p}_{\mathrm{loc}}(X)$ if $f\eta \in N^{1,p}$ whenever $\eta \in \mathrm{LIP}_b(X)$. In these cases we also can define a minimal p-upper gradient g_f, so that $\eta g_f \in L^p(X)$ for every $\eta \in \mathrm{LIP}_b(X)$. In other words, $g_f \in L^p_{\mathrm{loc}}(X)$.

We denote by $N^{1,p}(X; \mathbb{R}^N) \simeq N^{1,p}(X)^N$ the space of functions $\varphi\colon X \to \mathbb{R}^N$ so that each component is in $N^{1,p}$. Similarly, we define $\mathrm{LIP}_b(X; \mathbb{R}^N) \simeq \mathrm{LIP}_b(X)^N$.

Another notion of Sobolev space can be defined using *q-test plans* and we denote it by $W^{1,p}(X)$, with $|Df|_p$ denoting the minimal gradient of $f \in W^{1,p}(X)$. Namely, a function $f \in L^p(\mu)$ belongs to the Sobolev space $W^{1,p}(X)$ if there exists $g \in L^p(\mu)$ such that

$$\int |f(\gamma_1) - f(\gamma_0)|\, \mathrm{d}\boldsymbol{\eta} \le \iint_0^1 g(\gamma_t)|\gamma_t'|\, \mathrm{d}t\, \mathrm{d}\boldsymbol{\eta}$$

for every q-test plan $\boldsymbol{\eta}$ on X. The space has a norm $\|f\|_{W^{1,p}} = (\|f\|_{L^p}^p + \inf_g \|g\|_{L^p}^p)^{1/p}$, where the infimum is over all such functions g. We refer to [Di Marino and Squassina 2019] for details.

Note that any representative of an element of $W^{1,p}(X)$ still belongs to $W^{1,p}(X)$, whilst a representative of an element in $N^{1,p}(X)$ belongs to $N^{1,p}(X)$ if and only if they agree outside a p-exceptional set. The next theorem says that up to this ambiguity of representatives, the two approaches produce the same

object. The measurability conclusion is also a corollary of [Eriksson-Bique 2023]. We refer to [Ambrosio et al. 2015a] for a proof.

Theorem 2.5. *Let $p \in (1, \infty)$. If $f \in N^{1,p}(X)$, then $f \in W^{1,p}(X)$ and $g_f = |Df|_p$ μ-a.e. Conversely, if $f \in W^{1,p}(X)$, then f has a Borel representative $\bar{f} \in N^{1,p}(X)$ with $g_{\bar{f}} = |Df|_p$ μ-a.e.*

3. Curvewise (almost) optimality of minimal upper gradients

3A. *Upper gradients with respect to plans.* Given a plan $\boldsymbol{\eta}$, we can speak of a gradient along its curves.

Definition 3.1. If $\boldsymbol{\eta}$ is a q-plan and $f \in N^{1,p}(X)$, then a Borel function g is an $\boldsymbol{\eta}$-upper gradient if g is an upper gradient of f along γ for $\boldsymbol{\eta}$-almost every γ.

The following lemma gives a notion of a minimal $\boldsymbol{\eta}$-upper gradient and shows how to compute it by using derivatives along curves.

Lemma 3.2. *Suppose g_f is a minimal upper gradient and $\boldsymbol{\eta}$ is any q-plan and $\mathrm{d}\boldsymbol{\pi} = \mathrm{d}\boldsymbol{\eta}|\gamma_t'|\,\mathrm{d}t$, with disintegration $\boldsymbol{\pi}_x$. Then:*

(1) *$g_{\boldsymbol{\eta}} = \|(f \circ \gamma)_t'/\gamma_t'\|_{L^\infty(\boldsymbol{\pi}_x)}$ is a $\boldsymbol{\eta}$-upper gradient.*

(2) *$g_{\boldsymbol{\eta}} \le g$ for any other η-upper gradient for almost every $x \in X$.*

(3) *$g_{\boldsymbol{\eta}} \le g_f$ for almost every $x \in X$.*

(4) *Suppose $\boldsymbol{\eta}'$ is another q-plan and $\boldsymbol{\eta} \ll \boldsymbol{\eta}'$. Then $g_{\boldsymbol{\eta}} \le g_{\boldsymbol{\eta}'}$.*

Proof. Let g_f be the minimal p-upper gradient for f. By Lemma A.2 there is a Borel family $\Gamma_0 \subset C(I; X)$, so that f is absolutely continuous on each curve $\gamma \notin \Gamma_0^c$ with upper gradient g_f and so that $\boldsymbol{\eta}(\Gamma_0) = 0$. By Corollary A.3 and Lemma A.1 there is a set $N \subset C(I; X) \times I$ so that for $\boldsymbol{\pi}(N) = 0$, and for each $(\gamma, t) \notin N$, both $(f \circ \gamma)'(t)$ and $|\gamma_t'|$ are defined and measurable. Let $M_0 = \Gamma_0 \times I \cup N$. We get $\boldsymbol{\pi}(M_0) = 0$. For each curve $\gamma \notin \Gamma_0$ the function f is absolutely continuous with upper gradient $(f \circ \gamma)_t'/|\gamma_t'|$. Since $g_{\boldsymbol{\eta}}(\gamma_t) \ge (f \circ \gamma)_t'/|\gamma_t'|$ for $\boldsymbol{\pi}$-almost every $(\gamma, t) \in M_0$, we have that $g_{\boldsymbol{\eta}}$ is an $\boldsymbol{\eta}$ upper gradient.

If g is any other Borel $\boldsymbol{\eta}$-upper gradient, then the set of $(\gamma, t) \in \mathrm{Diff}(f) \setminus M_0$ with $(f \circ \gamma)_t'/|\gamma_t'| > g(\gamma_t)$ must have null measure, and thus the claim follows by Fubini and the definition in (1).

The function g_f is an upper gradient for f on curves in Γ_0^c, and thus the claim follows again from curvewise absolute continuity and by showing that the set of (γ, t) with $(f \circ \gamma)_t'/|\gamma_t'| > g_f(\gamma_t)$ must have null $\boldsymbol{\pi}$-measure. The final claim follows since $g_{\boldsymbol{\eta}'}$ must be a $\boldsymbol{\eta}$-upper gradient for f. □

3B. *Proof of Theorem 1.1.* In this subsection we prove Theorem 1.1. The idea is that for each q-plan $\boldsymbol{\eta}$ we can associate a gradient "along" the curves of such a plan. Each such gradient must be less than the minimal upper gradient, and thus the task is to show that by varying over different plans $\boldsymbol{\eta}$ we can obtain the minimal upper gradient through maximization. In order to show equality of the result of this maximization, we argue by contradiction, that if it were not a minimal upper gradient, then we could witness this by a given plan. This is the core of the following result. It should be compared to [Ambrosio et al. 2015b, Sections 9–11], where a similar analysis is done, but with different terminology and only for $p > 1$. In the following statement we will need to refer to end points of curves, and thus choose the domain of curves as $I = [0, 1]$.

Lemma 3.3. *Let $p \in [1, \infty)$, and q be its dual exponent. Let $f \in N^{1,p}(X)$. Suppose g is any nonnegative Borel function so that $A = \{g < g_f\}$ has positive measure. Then there exists a q-plan $\boldsymbol{\eta}$, so that for $\boldsymbol{\eta}$-almost every curve $\gamma : [0, 1] \to X$ we have*

$$|f(\gamma_1) - f(\gamma_0)| > \int_\gamma g \, \mathrm{d}s. \tag{3-1}$$

Proof. By Vitali–Carathéodory we may find a lower semicontinuous $\tilde{g} \geq g$ which is integrable and so that $\tilde{A} = \{\tilde{g} < g_f\}$ has positive measure. We will suppress the tildes below to simplify notation and thus only consider the case of g lower semicontinuous. Since $g < g_f$ on a positive measure subset, g cannot be a minimal upper gradient, and thus there must exist a family $\Gamma \subset C(I; X)$ of curves with $\mathrm{Mod}_p(\Gamma) > 0$, so that (3-1) holds for each $\gamma \in \Gamma$. Modulus is invariant under reparametrization of curves and so we may consider the subset of those $\gamma \in \Gamma$ which are Lipschitz. We want to find a plan supported on Γ. However, the issue with this is that since $p = 1$ is allowed the family Γ may not be compact, the duality of modulus and q-plans may fail. So, we seek to "cover" Γ, up to a null modulus family by compact families. This covering is done in an iterative way.

Fix an R so that the modulus of Γ_R of those curves in Γ, which are contained in a ball $B(x_0, R)$ for some fixed $x_0 \in X$, is positive. Since f is measurable and X is complete and separable, Egorov's theorem implies the existence of an increasing sequence of compact sets K_n satisfying $\mu(B(x_0, R) \setminus \bigcup K_n) = 0$ for which $f|_{K_n}$ is continuous for each n. Define $\mu(B(x_0, R) \setminus K_n) = \varepsilon_n$. By passing to a subsequence of n, we may assume that $\sum_n \sqrt{\varepsilon_n} < 1$.

Define $\bar{\Gamma}$ as the collection of $\gamma \in \Gamma_R$ so that f is absolutely continuous on γ and $\mathcal{H}^1\big(\gamma \setminus \big(\bigcup_{n=1}^\infty K_n\big)\big) = 0$. This holds for Mod_p-almost every curve, since $f \in N^{1,p}(X)$ and since p-almost every curve spends measure zero in the null set $X \setminus \bigcup_{n=1}^\infty K_n$. Thus, $\mathrm{Mod}_p(\bar{\Gamma}) > 0$.

Next, let Γ^m be those curves $\gamma : I \to X$, which are m-Lipschitz, so that $\mathrm{Len}(\gamma) \leq m|b - a|$, $\mathrm{diam}(\gamma) \geq 1/m$, $\gamma_0, \gamma_1 \in K_m$ and (3-1) holds. We will show that every $\gamma \in \bar{\Gamma}$ contains a subcurve, up to reparametrization, in $\bigcup_{m=1}^\infty \Gamma^m$. From this, and [Björn and Björn 2011, Lemma 1.34], it follows that $\mathrm{Mod}_p\big(\bigcup_{m=1}^\infty \Gamma^m\big) > 0$, and thus there is some $M > 0$ so that $\mathrm{Mod}_p(\Gamma^M) > 0$. It is easy to show that Γ^m is a closed family of curves in $C(I; X)$ with respect to uniform convergence, since g is taken to be lower semicontinuous (see, e.g., [Keith 2003, Proposition 4]).

To obtain the previous fact, consider a nonconstant curve $\gamma \in \bar{\Gamma}$. We have

$$|f(\gamma_1) - f(\gamma_0)| > \int_\gamma g \, \mathrm{d}s.$$

We may also parametrize γ by constant speed as the claim is invariant under reparametrizations.

Since γ has constant speed, we know $\big|I \setminus \bigcup_{n=1}^\infty \gamma^{-1}(K_n)\big| = 0$ and $f \circ \gamma$ is continuous. Since $\int_\gamma g \, \mathrm{d}s < \infty$ and $f \circ \gamma$ is continuous, we can find (for all $n \geq N$ for some $N \in \mathbb{N}$) sequences $a_n, b_n \in [0, 1]$ so that $\lim_{n\to\infty} a_n = a$, $\gamma_{a_n} \in K_n$, $\gamma_{b_n} \in K_n$ and $\lim_{n\to\infty} b_n = b$. Then, for sufficiently large n

$$|f(\gamma_{b_n}) - f(\gamma_{a_n})| > \int_{\gamma|_{[a_n, b_n]}} g \, \mathrm{d}s.$$

For n large enough we also have $\mathrm{Len}(\gamma_{[a_n,b_n]}) \le n|b-a|$, $\mathrm{diam}(\gamma_{[a_n,b_n]})) \ge 1/n$. Since the curves are parametrized by constant speed, they are n-Lipschitz. So $\gamma' = \gamma_{[a_n,b_n]}$ is, up to a reparametrization, in Γ^n for n large enough, and the claim follows.

Fix $M > 0$ so that $\mathrm{Mod}_p(\Gamma^M) > 0$. Next, choose $\delta < \min(\mathrm{Mod}_p(\Gamma^M), 1)$. Define $\delta_n = \varepsilon_n^{1/2p}$. Choose N so that $\sum_{n=N}^{\infty} \sqrt{\varepsilon_n} < \delta^{1+p}/2$. Let Γ_t^M be the family of curves $\gamma \in \Gamma^M$ so that $\int_\gamma 1_{X\setminus K_n}\,\mathrm{d}s \le \delta\delta_n$ for each $n \ge N$. Since $\left(\sum_{n\ge N}(1_{X\setminus K_n}/(\delta\delta_n))^p\right)^{1/p}$ is a function admissible for $\Gamma^M \setminus \Gamma_t^M$, we have

$$\mathrm{Mod}_p(\Gamma^M \setminus \Gamma_t^M) \le \sum_{n\ge N} \frac{\varepsilon_n}{\delta^p \delta_n^p} < \delta/2.$$

Thus, by subadditivity of modulus, see, e.g., [Fuglede 1957, Theorem 1],

$$\mathrm{Mod}_p(\Gamma_t^M) \ge \mathrm{Mod}_p(\Gamma^M) - \mathrm{Mod}_p(\Gamma^M \setminus \Gamma_t^M) > \delta/2.$$

By Lemma 3.4, since Γ^M is closed, the family $\Gamma_t^M \subset \Gamma^M$ is a compact family of curves in a complete space. Then, by Proposition 2.3 there exists a q-plan $\boldsymbol{\eta}$ supported on Γ_t^M. Each curve $\gamma \in \Gamma_t^M$ satisfies (3-1), and thus the claim follows. □

For the following proof, recall that if $A, B \subset X$, then $d(A, B) := \inf_{a\in A} \inf_{b\in B} d(a, b)$, and $N_\varepsilon(A) := \bigcup_{a\in A} B(a, \varepsilon)$ for $\varepsilon > 0$.

Lemma 3.4. *Suppose that K_n are compact sets, $\eta_n > 0$ constants with $\lim_{n\to\infty} \eta_n = 0$, $L > 0$ and let $\Gamma \subset C(I; X)$ be a closed family of curves in a complete space X. Let $\Gamma^{n,L}$ be the family of curves $\gamma \in \Gamma$ for which $\mathrm{Len}(\gamma) \le L$, $\mathrm{diam}(\gamma) \ge 1/L$ and which are L-Lipschitz, with $\int_\gamma 1_{X\setminus K_n}\,\mathrm{d}s \le \eta_n$ for each $n \in \mathbb{N}$. Then $\Gamma^{n,L}$ is compact.*

Proof. Let $I = [a, b]$. Since Γ and $\Gamma^{n,L}$ are closed, it suffices to show precompactness.

Let $\gamma \in \Gamma^{n,L}$. We may suppose that $\eta_n < 1/(2L)$ by restricting to large enough n. Then, we have for each n

$$\int_\gamma 1_{K_n}\,\mathrm{d}s = \int_\gamma 1\,\mathrm{d}s - \int_\gamma 1_{X\setminus K_n}\,\mathrm{d}s \ge \mathrm{diam}(\gamma) - \eta_n > \frac{1}{L} - \eta_n.$$

Thus $\gamma \cap K_n \ne \varnothing$. Moreover, if $t \in I$, and $d(\gamma_t, K_n) = s$, then there will be a subsegment of length at least $\min(s, \mathrm{diam}(\gamma_k)/2)$ in $X \setminus K_n$. This gives $\min(s, \mathrm{diam}(\gamma)/2) \le \eta_n < 1/(2L)$. This is only possible if $s \le \eta_n$, since $\mathrm{diam}(\gamma)/2 \ge 1/L$. Indeed, we have $d(\gamma, K_n) \le \eta_n$.

To run the usual proof of Arzelà–Ascoli, since we have equicontinuity with the Lipschitz bound, we only need to show that for each fixed $t \in I$ the set $A_t = \{\gamma_t : \gamma \in \Gamma^{n,L}\}$ is precompact. However, since X is complete, it suffices to show that A_t is totally bounded. Fix $\varepsilon > 0$. We concluded that $d(\gamma, K_n) \le \eta_n$ for all $n \in \mathbb{N}$. Thus, we have for some large n that $\eta_n \le \varepsilon/4$ and that $A_t \subset N_{\eta_n}(K_n) \subset N_{\varepsilon/4}(K_n)$. Since K_n is compact, it is totally bounded, and the claim follows by covering K_n by finitely many $\varepsilon/4$ balls and noting that $\varepsilon > 0$ is arbitrary. □

Proof of Theorem 1.1. Let Π_q be the set of all q-plans, and for each $\boldsymbol{\eta} \in \Pi_q$, with its disintegration being given by $\boldsymbol{\pi}_x$, define

$$g_{\boldsymbol{\eta}}(x) = \left\| \frac{(f\circ\gamma)'_t}{|\gamma'_t|} \right\|_{L^\infty(\boldsymbol{\pi}_x)}.$$

Finally, define

$$|D_\pi f| = \operatorname*{ess\,sup}_{\boldsymbol{\eta}\in\Pi_\infty} g_\pi(x).$$

Claim 1. *There is a q-plan $\tilde{\boldsymbol{\eta}}$ so that $|D_\pi f| = g_{\tilde{\boldsymbol{\eta}}}$.*

By Lemma A.5, we can find a sequence $\boldsymbol{\eta}_n$ so that

$$g_{\boldsymbol{\eta}_n} \to |D_\pi f|$$

almost everywhere. Consider the measures $\mathrm{d}\boldsymbol{\pi}^n := |\gamma_t'|\,\mathrm{d}\boldsymbol{\eta}_n\,\mathrm{d}t$ on $\mathrm{AC}(I;X)\times I$. Set

$$a_n = 1 + \boldsymbol{\eta}_n(C(I;X)) + \left\|\frac{d\boldsymbol{\eta}^\#}{d\mu}\right\|_{L^q} + \boldsymbol{\pi}^n(\mathrm{AC}(I;X)\times I),$$

where $\boldsymbol{\eta}^\#$ is the barycenter of $\boldsymbol{\eta}$, which is absolutely continuous with respect to μ. Let $\tilde{\boldsymbol{\eta}} = \sum_{n=1}^\infty a_n^{-1}2^{-n}\boldsymbol{\eta}_n$. This will be a plan with $g_{\tilde{\boldsymbol{\eta}}} \geq g_{\boldsymbol{\eta}_n}$ for each n by Lemma 3.2. For μ-almost every x, we have $g_{\tilde{\boldsymbol{\eta}}} \geq |D_\pi f|$. Then, by Lemma 3.3 we have $\|(f\circ\gamma)_t'/|\gamma_t'|\|_{L^\infty(\boldsymbol{\pi}_x)} = |D_\pi f|$, as stated.

Claim 2. *We have $|D_\pi f| = g_f$ almost everywhere.*

Since g_f is a p-weak upper gradient, Lemma 3.2 gives $|D_\pi f| \leq g_f$. Suppose for the sake of contradiction then that $|D_\pi f| < g_f$ on a positive measure subset. Then, by Lemma 3.2, there exists a plan $\boldsymbol{\eta}'$ so that

$$|f(\gamma_1) - f(\gamma_0)| > \int_\gamma |D_\pi f|\,\mathrm{d}s$$

for $\boldsymbol{\eta}'$-almost every γ.

However, by the definition of a plan upper gradient, we have for $\boldsymbol{\eta}'$ almost every curve that

$$|f(\gamma_1) - f(\gamma_0)| \leq \int_\gamma g_{\boldsymbol{\eta}'}\,\mathrm{d}s.$$

Now, as $g_{\boldsymbol{\eta}'} \leq |D_\pi f|$ almost everywhere and as $\boldsymbol{\eta}'$ is a q-plan, we have for $\boldsymbol{\eta}'$-almost every curve γ that

$$\int_\gamma g_{\boldsymbol{\eta}'}\,\mathrm{d}s \leq \int_\gamma |D_\pi f|\,\mathrm{d}s,$$

which contradicts the above inequalities.

Finally, since $|D_\pi f| = g_{\tilde{\boldsymbol{\eta}}} = g_f$, we must have $\mu|_D \ll \tilde{\boldsymbol{\eta}}^\#$. Indeed, otherwise there would be a non-null Borel set $E \subset D$ for which $\mu(E) > 0$ and $\tilde{\boldsymbol{\eta}}^\#(E) = 0$. However, then $g_{\tilde{\boldsymbol{\eta}}}|_E = 0$, contradicting the equality μ-almost everywhere. □

We now prove Corollary 1.3.

Proof. Let $f \in N^{1,p}$ and consider the plan $\boldsymbol{\eta}'$ obtained from Theorem 1.1. Let $\boldsymbol{\eta}'' = r_*(\boldsymbol{\eta}')$, where $r : C(I;X) \to C(I;X)$ is the reversal-map which reverses the orientation of every path. Define $\boldsymbol{\eta} = \boldsymbol{\eta}'' + \boldsymbol{\eta}'$. Fix $\varepsilon > 0$, and define $B = \{(\gamma,t) \in \mathrm{Diff}(f) : g_f(x) \geq (f\circ\gamma)_t'/|\gamma_t'| \geq (1-\varepsilon)g_f(x)\}$. Since $\|(f\circ\gamma)_t'/|\gamma_t'|\|_{L^\infty(\boldsymbol{\pi}_x')} = g_f(x)$ for μ-almost every $x \in D$, where $\boldsymbol{\pi}_x'$ is the disintegration for $\boldsymbol{\eta}'$, we have $\boldsymbol{\pi}_x(B) > 0$ for μ-almost every $x \in D$ where $\boldsymbol{\pi}_x$ is the disintegration corresponding to $\boldsymbol{\eta}$. Note that, we can remove the absolute values from the supremum norm since for each path γ in the support of $\boldsymbol{\eta}'$ we include also its reversal, and r preserves $\boldsymbol{\eta}$. □

3C. *Alternative curvewise characterizations of upper gradients when $p > 1$.* In this subsection we assume that $p, q \in (1, \infty)$ satisfy $1/p + 1/q = 1$ and prove a variant Theorem 1.1 using test plans representing gradients, introduced by Gigli.

Given $f \in N^{1,p}(X)$, a q-test plan $\boldsymbol{\eta}$ *represents* g_f if

$$\frac{f \circ e_t - f \circ e_0}{t \widetilde{E}_t^{1/q}} \to g_f \circ e_0 \qquad \text{and} \qquad \widetilde{E}_t^{1/p} \to g_f \circ e_0 \quad \text{in } L^p(\boldsymbol{\eta}),$$

where

$$\widetilde{E}_t(\gamma) = \frac{1}{t} \int_0^t |\gamma_s'|^q \, \mathrm{d}s, \quad \gamma \in \mathrm{AC}(I; X), \ \widetilde{E}_t(\gamma) = +\infty \text{ otherwise.}$$

A test plan $\boldsymbol{\eta}$ representing the gradient of a Sobolev map $f \in N^{1,p}(X)$ is concentrated on "gradient curves" of f in an asymptotic and integrated sense. We refer to [Gigli 2015; Pasqualetto 2022] for discussion of the definition we are using here. The following result of Gigli states that Sobolev functions always possess test plans representing their gradient. In the statement, $\mathcal{P}_q(X)$ denotes probability measures ν on X with $\int d(x_0, x)^q \, \mathrm{d}\nu(x) < \infty$ for some and thus any $x_0 \in X$.

Theorem 3.5 [Gigli 2015, Theorem 3.14]. *If $f \in N^{1,p}(X)$ and $\nu \in \mathcal{P}_q(X)$ satisfies $\nu \le C\mu$ for some $C > 0$, there exists a q-test plan $\boldsymbol{\eta}$* ***representing*** *g_f, with $e_{0*}\boldsymbol{\eta} = \nu$.*

We now state the main result of this subsection.

Theorem 3.6. *Let $f \in N^{1,p}(X)$ and g_f be a Borel representative of the minimal p-weak upper gradient of f, with $D := \{g_f > 0\}$ of positive μ-measure. Let $\boldsymbol{\eta}$ be a q-test plan representing g_f with $\mu|_D \ll e_{0*}\boldsymbol{\eta} \ll \mu|_D$.*

For every $\varepsilon > 0$ there exists a Borel set $B \subset \mathrm{Diff}(f)$ such that $\mathrm{d}\boldsymbol{\pi} := \chi_B |\gamma_t'| \, \mathrm{d}t \, \mathrm{d}\boldsymbol{\eta}$ is a positive (finite) measure with $\mu|_D \ll e_*\boldsymbol{\pi} \ll \mu|_D$, whose disintegration $\{\boldsymbol{\pi}_x\}$ with respect to e satisfies

$$(1-\varepsilon) g_f(x) \le \frac{(f \circ \gamma)_t'}{|\gamma_t'|} \le g_f(x) \quad \textit{and} \quad (1-\varepsilon) g_f(x)^{p/q} \le |\gamma_t'| \le (1+\varepsilon) g_f(x)^{p/q} \quad \textit{for } \boldsymbol{\pi}_x\textit{-a.e. } (\gamma, t),$$

for μ-almost every $x \in D$.

For the proof, we present the following three elementary lemmas. Define

$$D_t(\gamma) = \frac{f(\gamma_t) - f(\gamma_0)}{t} \qquad \text{and} \qquad G_t(\gamma) = \frac{1}{t} \int_0^t g_f(\gamma_s)^p \, \mathrm{d}s, \quad \gamma \in \mathrm{AC}(I; X),$$

and $+\infty$ otherwise. The following observation is essentially made in [Pasqualetto 2022, Lemma 1.19] (we are using different notation for our purposes). See Lemma A.1(3) for the Borel measurability of the functionals in the claim.

Lemma 3.7. *Suppose $f \in N^{1,p}(X)$ and suppose $\boldsymbol{\eta}$ is a q-test plan representing g_f. Then*

$$D_t, G_t, \widetilde{E}_t \to g_f^p \circ e_0 \quad \textit{in } L^1(\boldsymbol{\eta}).$$

Proof. Since $\widetilde{E}_t^{1/p} \to g_f^p \circ e_0$ in $L^p(\boldsymbol{\eta})$, it follows that $\widetilde{E}_t \to g_f^p \circ e_0$ in $L^1(\boldsymbol{\eta})$. The convergence $D_t \to g_f^p \circ e_0$ is proven in [Pasqualetto 2022, Lemma 1.19], while $G_t \to g_f^p \circ e_0$ in $L^1(\boldsymbol{\eta})$ follows from [Gigli 2015, Proposition 2.11]. □

Lemma 3.8. *For every $\varepsilon > 0$ there exists $\delta > 0$ with the following property: if $a, b > 0$ and $a^p/p + b^q/q \le ab/(1-\delta)$, then $|a^{p/q}/b - 1| < \varepsilon$.*

Proof. The function $h : (0, \infty) \to (0, \infty)$, given by $h(t) = t/p + t^{-q/p}/q$, has a global minimum at $t = 1$, with $h(1) = 1$. Thus $h|_{(0,1]}$ and $h|_{[1,\infty)}$ have continuous inverses and it follows that for every $\varepsilon > 0$ there exists $\delta > 0$ such that if $|1 - h(t)| < \delta$ then $|1 - t| < \varepsilon$ (expressing the fact that both inverses are continuous at 1). The claim follows from this by noting that if $a^p/p + b^q/q \le ab/(1-\delta)$ then $0 \le h(t) - 1 < \delta$, where $t := a^{p/q}/b$. □

Lemma 3.9. *Let $h \le g$ be two integrable functions on an interval $I = [0, T]$, with*

$$\liminf_{n\to\infty} \frac{1}{T_n} \int_0^{T_n} g \,\mathrm{d}s =: A > 0 \quad \textit{and} \quad \lim_{n\to\infty} \frac{1}{T_n} \int_0^{T_n} [g - h] \,\mathrm{d}s = 0$$

for some sequence $T_n \to 0$. Then, for every $\varepsilon > 0$ and n, the set $\{(1-\varepsilon)g < h\} \cap [0, T_n]$ has positive $\mathcal{L}^1$-measure.

Proof. For large enough n we have $0 < A/2 < (1/T_n) \int_0^{T_n} g \,\mathrm{d}s$ and $0 \le (1/T_n) \int_0^{T_n} [g - h] \,\mathrm{d}s < \varepsilon A/2$. Thus, we may find some n_0 for which $(1/T_n) \int_0^{T_n} [g - h] \,\mathrm{d}s < (\varepsilon/T_n) \int_0^{T_n} g \,\mathrm{d}s$ for each $n > n_0$. It follows that $\int_0^{T_n} [(1-\varepsilon)g - h] \,\mathrm{d}s < 0$ for $n > n_0$, and the claim follows from this. □

We will also need the following technical result; compare Lemma 3.7.

Lemma 3.10. *Let $E \subset X$ be a Borel set, $t > 0$, and let*

$$D_{E,t}(\gamma) := \frac{1}{t} \int_0^t \chi_E(\gamma_s)(f \circ \gamma)'_s \,\mathrm{d}s, \quad \gamma \in \Gamma(f).$$

Then $D_{E,t} \to (\chi_E g_f^p) \circ e_0$ in $L^1(\boldsymbol{\eta})$.

Proof. Define

$$F_t(\gamma) := \frac{1}{t} \int_0^t g_f(\gamma_s)|\gamma'_s| \,\mathrm{d}s.$$

Since $D_t \le F_t \le (1/p)G_t + (1/q)\widetilde{E}_t$ $\boldsymbol{\eta}$-almost everywhere, Lemma 3.7 implies $F_t \to g_f^p \circ e_0$ and thus $(\chi_E \circ e_0)F_t \to (\chi_E g_f^p) \circ e_0$ in $L^1(\boldsymbol{\eta})$. We show that $(\chi_E \circ e_0)F_t - D_{E,t} \to 0$ in $L^1(\boldsymbol{\eta})$.

For $\boldsymbol{\eta}$-almost every γ we have

$$\begin{aligned}
&|\chi_E(\gamma_0)F_t(\gamma) - D_{E,t}(\gamma)| \\
&\quad = \left| \frac{1}{t} \int_0^t [\chi_E(\gamma_0) g_f(\gamma_s)|\gamma'_s| - \chi_E(\gamma_s)(f \circ \gamma)'_s] \,\mathrm{d}s \right| \\
&\quad \le \frac{1}{t} \int_0^t \big(|(\chi_E g_f)(\gamma_s) - (\chi_E g_f)(\gamma_0)||\gamma'_s| + \chi_E(\gamma_0)|g_f(\gamma_s) - g_f(\gamma_0)||\gamma'_s| \\
&\qquad\qquad + \chi_E(\gamma_s)[g_f(\gamma_s)|\gamma'_s| - (f \circ \gamma)'_s] \big) \,\mathrm{d}s \\
&\quad \le \left[\left(\frac{1}{t} \int_0^t |(\chi_E g_f)(\gamma_s) - (\chi_E g_f)(\gamma_0)|^p \,\mathrm{d}s \right)^{1/p} + \left(\frac{1}{t} \int_0^t |g_f(\gamma_s) - g_f(\gamma_0)|^p \,\mathrm{d}s \right)^{1/p} \right] \left(\frac{1}{t} \int_0^t |\gamma'_s|^q \,\mathrm{d}s \right)^{1/q} \\
&\qquad\qquad + F_t(\gamma) - D_t(\gamma).
\end{aligned}$$

This estimate, together with the Hölder inequality and Lemma 3.7, yields

$$\begin{aligned}
&\limsup_{t\to 0}\int |(\chi_E\circ e_0)F_t - D_{E,t}|\,\mathrm{d}\boldsymbol{\eta}\\
&\quad\le \limsup_{t\to 0}\Bigg[\Bigg(\int \frac{1}{t}\int_0^t |g_f(\gamma_s)-g_f(\gamma_0)|^p\,\mathrm{d}s\,\mathrm{d}\boldsymbol{\eta}\Bigg)^{1/p}\\
&\qquad\qquad + \Bigg(\int \frac{1}{t}\int_0^t |(\chi_E g_f)(\gamma_s)-(\chi_E g_f)(\gamma_0)|^p\,\mathrm{d}s\,\mathrm{d}\boldsymbol{\eta}\Bigg)^{1/p}\Bigg]\times\Bigg(\int g_f^p\circ e_0\,\mathrm{d}\boldsymbol{\eta}\Bigg)^{1/q}\\
&\quad= \limsup_{t\to 0}\Bigg[\Bigg(\frac{1}{t}\int_0^t \|g_f\circ e_s - g_f\circ e_0\|^p_{L^p(\boldsymbol{\eta})}\,\mathrm{d}s\Bigg)^{1/p}\\
&\qquad\qquad + \Bigg(\frac{1}{t}\int_0^t \|(\chi_E g_f)\circ e_s - (\chi_E g_f)\circ e_0\|^p_{L^p(\boldsymbol{\eta})}\,\mathrm{d}s\Bigg)^{1/p}\Bigg]\times\Bigg(\int g_f^p\circ e_0\,\mathrm{d}\boldsymbol{\eta}\Bigg)^{1/q}.
\end{aligned}$$

Since $s\mapsto h\circ e_s$ is continuous in $L^p(\boldsymbol{\eta})$ whenever $h\in L^p(\mu)$ (see [Gigli and Pasqualetto 2020, Proposition 2.1.4]) all terms above tend to zero, proving the claimed convergence. □

Proof of Theorem 3.6. Let N be the negligible set is as in Corollary A.3. The function

$$A(\gamma,t)=\frac{1}{p}g_f(\gamma_t)^p+\frac{1}{q}|\gamma_t'|^q,\quad (\gamma,t)\notin N,\ A(\gamma,t)=+\infty,\ (\gamma,t)\in N,$$

is Borel. Let $\boldsymbol{\eta}$ represent g_f and satisfy $\mu|_D \ll e_{0*}\boldsymbol{\eta}\ll \mu|_D$. Fix $\varepsilon>0$, let $\delta>0$ be as in Lemma 3.8, and set $\delta_0=\min\{\varepsilon,\delta\}$. We define the Borel function

$$H(\gamma,t)=(1-\delta_0)A(\gamma,t)-(f\circ\gamma)_t',\quad (\gamma,t)\notin N,\ H=+\infty \text{ otherwise};$$

see Corollary A.3. The set $B:=\{H\le 0\}$ is Borel and, for $(\gamma,t)\notin N$, we have

$$(f\circ\gamma)_t'\le g_f(\gamma_t)|\gamma_t'|\le A(\gamma,t). \tag{3-2}$$

Note that

$$H(\gamma,t)\le 0 \quad\text{implies}\quad (1-\varepsilon)g_f(\gamma_t)|\gamma_t'|\le (f\circ\gamma)_t' \text{ and } \left|1-\frac{g_f(\gamma_t)^{p/q}}{|\gamma_t'|}\right|<\varepsilon; \tag{3-3}$$

see (3-2) and Lemma 3.8. Once we show that $\mathrm{d}\boldsymbol{\pi}:=\chi_B|\gamma_t'|\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\eta}$ satisfies

$$\mu|_D\ll e_*\boldsymbol{\pi}\ll\mu|_D,$$

it follows from (3-2) and (3-3) that $\boldsymbol{\pi}':=\boldsymbol{\pi}/\boldsymbol{\pi}(C(I;X)\times I)\in\mathcal{P}(C(I;X)\times I)$ satisfies

$$(1-\varepsilon)g_f(\gamma_t)|\gamma_t'|\le (f\circ\gamma)_t'\le g_f(\gamma_t)\quad\text{and}\quad \frac{g(\gamma_t)^{p/q}}{1+\varepsilon}\le|\gamma_t'|\le\frac{g(\gamma_t)^{p/q}}{1-\varepsilon}$$

for $\boldsymbol{\pi}'$-almost every (γ,t), which readily implies the inequalities in the theorem.

To prove $e_*\boldsymbol{\pi}\ll\mu|_D$ observe that (3-3) implies $\chi_B|\gamma_t'|\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\eta}\le(1+\varepsilon)g(\gamma_t)^{p/q}\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\eta}$ and thus

$$\iint_0^1 \chi_B(\gamma,t)\chi_E(\gamma_t)|\gamma_t'|\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\eta}\le(1+\varepsilon)\int_0^1\int_X \chi_E g_f^{p/q}e_{t*}(\mathrm{d}\boldsymbol{\eta})\,\mathrm{d}t\le C\int_E g_f^{p/q}\,\mathrm{d}\mu$$

for any Borel set $E\subset X$.

It remains to prove that $\mu|_D \ll e_*\boldsymbol{\pi}$. Let $E \subset D$ be a Borel set with $\mu(E) > 0$. Then $e_{0*}\boldsymbol{\eta}(E) = \boldsymbol{\eta}(\{\gamma : \gamma_0 \in E\}) > 0$. Since

$$0 \le \frac{1}{t}\int_0^t \chi_E(\gamma_s)A(\gamma, s)\,\mathrm{d}s - D_{E,t}(\gamma) \le \frac{1}{p}G_t(\gamma) + \frac{1}{q}\widetilde{E}_t(\gamma) - D_t(\gamma) \xrightarrow{t\to 0} 0,$$
$$D_{E,t} \xrightarrow{t\to 0} \chi_E g_f^p \circ e_0$$

in $L^1(\boldsymbol{\eta})$, see Lemmas 3.7 and 3.10 respectively, there exists a sequence $T_n \to 0$ such that for $\boldsymbol{\eta}$-almost every $\gamma \in e_0^{-1}(E)$ the functions

$$h_\gamma(s) := \chi_E(\gamma_s)(f\circ\gamma)'_s, \quad g_\gamma(s) := \chi_E(\gamma_s)A(\gamma, s)$$

satisfy the hypotheses of Lemma 3.9. It follows that for $\boldsymbol{\eta}$-almost every $\gamma \in e_0^{-1}(E)$ the sets

$$I_\gamma^n := \{s \in [0, T_n] : (1-\delta_0)g_\gamma(s) < h_\gamma(s)\} = \{s \in [0, T_n] : \gamma_s \in E,\ H(\gamma, s) \le 0\}$$

have positive measure for all n. Notice that, for $\boldsymbol{\eta}$-almost every γ, if $s \in I_\gamma^n$ then $\gamma_s \in E$ and $|\gamma'_s| > 0$, $g_f(\gamma_s) > 0$ (since $0 < (f\circ\gamma)'_s \le g_f(\gamma_t)|\gamma'_s|$). Consequently

$$\int_0^1 \chi_B(\gamma, s)\chi_E(\gamma_s)|\gamma'_s|\,\mathrm{d}s \ge \int_{I_\gamma^n} |\gamma'_s|\,\mathrm{d}s > 0$$

for $\boldsymbol{\eta}$-almost every $\gamma \in e_0^{-1}(E)$, which in turn implies $e_*\boldsymbol{\pi}(E) > 0$. Since $E \subset D$ is an arbitrary Borel set with positive μ-measure, this completes the proof. □

4. Charts and differentials

4A. ***Notational remarks.*** In what follows, define for any set $U \subset X$ the set of curves which spend positive *length* in U:

$$\Gamma_U^+ = \left\{\gamma \in \mathrm{AC}(I; X) : \int_\gamma \chi_U\,\mathrm{d}s > 0\right\}.$$

Having positive length in U is more restrictive than assuming that $\gamma^{-1}(U)$ has positive measure. We will also discuss p-weak differentials and covector fields of the form $\mathrm{d}f : U \to (\mathbb{R}^N)^*$ or $\boldsymbol{\xi} : U \to (\mathbb{R}^N)^*$ for measurable subsets $U \subset X$. The values of such a map at $x \in U$ are denoted by $\mathrm{d}_x f$, $\boldsymbol{\xi}_x$, respectively.

4B. ***Canonical minimal gradients.*** Let $p \ge 1$ and $N \ge 0$ be given. For the next three lemmas we fix $\varphi \in N^{1,p}_{\mathrm{loc}}(X; \mathbb{R}^N) \simeq N^{1,p}_{\mathrm{loc}}(X)^N$, with the convention $N^{1,p}_{\mathrm{loc}}(X; \mathbb{R}^N) = N^{1,p}_{\mathrm{loc}}(X)^N = \{0\}$ when $N = 0$. Our aim is to construct a "canonical" representative of the minimal weak upper gradients $|D(\xi\circ\varphi)|_p$ of the functions $\xi\circ\varphi$. We will use a plan to represent it.

Lemma 4.1. *There exists a q-plan $\boldsymbol{\eta}$ and a Borel set D with $\mu|_D \ll \boldsymbol{\eta}^\#$ such that*

$$\Phi_\xi(x) := \chi_D(x)\left\|\frac{\xi((\varphi\circ\gamma)'_t)}{|\gamma'_t|}\right\|_{L^\infty(\boldsymbol{\pi}_x)} \tag{4-1}$$

is a representative of $|D(\xi\circ\varphi)|_p$ for every $\xi \in (\mathbb{R}^N)^*$. Here $\{\boldsymbol{\pi}_x\}$ is the disintegration of $\mathrm{d}\boldsymbol{\pi} := |\gamma'_t|\,\mathrm{d}\boldsymbol{\eta}\,\mathrm{d}t$ with respect to the evaluation map e.

Proof. Let $\{\xi_0, \xi_1, \dots\} \subset (\mathbb{R}^N)^*$ be a countable dense set and, for each $n \in \mathbb{N}$, choose Borel representatives ρ_n of $|D(\xi_n \circ \varphi)|_p$ and define $D_n := \{\rho_n > 0\}$. By Theorem 1.1 and the Borel regularity of μ, for each $n \in \mathbb{N}$ there exists a q-plan $\boldsymbol{\eta}_n$ and a Borel set $B_n \subset D_n$ with $\mu(D_n \setminus B_n) = 0$ such that the disintegration $\{\boldsymbol{\pi}^n_x\}$ of $\mathrm{d}\boldsymbol{\pi}^n := |\gamma_t'|\,\mathrm{d}\boldsymbol{\eta}_n\,\mathrm{d}t$ satisfies

$$\left\| \frac{\xi_n((\varphi \circ \gamma)_t')}{|\gamma_t'|} \right\|_{L^\infty(\boldsymbol{\pi}^n_x)} = \rho_n(x)$$

for every $x \in B_\xi$.

Define $D := \bigcup_{n\in\mathbb{N}} B_n$ and $\boldsymbol{\eta} = \sum_n 2^{-n} a_n^{-1} \boldsymbol{\eta}_n$, where $a_n = 1 + \boldsymbol{\eta}_n(C(I; X)) + \|d\boldsymbol{\eta}_n^{\#}/d\mu\|_{L^q} + \boldsymbol{\pi}^n(\mathrm{AC}(I; X) \times I)$. Then $\mu|_D \ll \boldsymbol{\eta}^{\#}$. Define $\Phi_\xi(x)$ as in (4-1). By Lemma 3.2 we have $\rho_n = \Phi_{\xi_n}$ μ-a.e. on X and thus the claim holds for every $\xi_n \in A$.

We prove the claim in the statement for arbitrary $\xi \in (\mathbb{R}^N)^*$. Let $(\xi_{n_l})_l \subset A$ be a sequence with $|\xi_{n_l} - \xi| < 2^{-l}$ and denote by $\varphi_1, \dots, \varphi_N \in N^{1,p}(X)$ the component functions of φ. Since

$$||D(\xi_{n_l} \circ \varphi)|_p - |D(\xi \circ \varphi)|_p| \le |D((\xi_{n_l} - \xi) \circ \varphi)|_p \le |\xi_{n_l} - \xi| \sum_k^N |D\varphi_k|_p$$

μ-a.e., we have $|D(\xi \circ \varphi)|_p = \lim_{l\to\infty} \Phi_{\xi_{n_l}}$ μ-a.e. on X. In particular, $|D(\xi \circ \varphi)|_p = 0$ μ-a.e. on $X \setminus D$. On the other hand, for p-a.e. curve γ, we have

$$|\xi_{n_l}((\varphi \circ \gamma)_t') - \xi((\varphi \circ \gamma)_t')| \le |\xi_{n_l} - \xi| \sum_k^N |D\varphi_k|_p(\gamma_t)|\gamma_t'| \quad \text{for a.e. } t.$$

Since $\boldsymbol{\eta}$ is a q-plan with $\mu|_D \ll \boldsymbol{\eta}^{\#}$, this implies

$$\limsup_{l\to\infty} \left| \frac{\xi_{n_l}((\varphi \circ \gamma)_t')}{|\gamma_t'|} - \frac{\xi((\varphi \circ \gamma)_t')}{|\gamma_t'|} \right| \le \limsup_{l\to\infty} |\xi_{n_l} - \xi| \sum_k^N |D\varphi_k|_p(x) = 0 \quad \text{for } \boldsymbol{\pi}_x\text{-a.e. } (\gamma, t),$$

for μ-a.e. $x \in D$. Thus $\Phi_\xi(x) = \lim_{l\to\infty} \Phi_{\xi_{n_l}}(x)$ for μ-a.e. $x \in D$. Since $\Phi_\xi = 0 = |D(\xi \circ \varphi)|_p$ μ-a.e. on $X \setminus D$, the proof is completed. □

In the next two lemmas we collect the properties of the Borel function constructed above.

Lemma 4.2. *The map $\Phi : (\mathbb{R}^N)^* \times X \to \mathbb{R}$ given by (4-1) is Borel and satisfies the following*:

(1) *For every $\xi \in (\mathbb{R}^N)^*$, $\Phi_\xi := \Phi(\xi, \cdot)$ is a representative of $|D(\xi \circ \varphi)|_p$.*

(2) *For every $x \in X$, $\Phi^x := \Phi(\cdot, x)$ is a seminorm in $(\mathbb{R}^N)^*$.*

Moreover, there exists a path family Γ_B with $\mathrm{Mod}_p(\Gamma_B) = 0$ and for each $\gamma \in \mathrm{AC}(I; X) \setminus \Gamma_B$ a null-set $E_\gamma \subset I$ so that, for ***every*** *$\xi \in (\mathbb{R}^N)^*$, we have*:

(3) *Φ_ξ is an upper gradient of $\xi \circ \varphi$ along γ.*

(4) *$|(\xi \circ \varphi \circ \gamma)_t'| \le \Phi_\xi(\gamma_t)|\gamma_t'|$ for $t \notin E_\gamma$.*

Proof of Lemma 4.2. Borel measurability follows from Lemma A.1 and Corollary A.3, and property (1) follows from Lemma 4.1, while (2) follows from (4-1).

Fix a countable dense set $A \subset (\mathbb{R}^N)^*$ and one $\xi \in A$. We have that $\Phi(\xi, x)$ is a weak upper gradient for $\xi \circ \varphi$, so there is family of curves Γ_ξ so that $\xi \circ \varphi$ is absolutely continuous with upper gradient $|D(\xi \circ \varphi)|_p$ on each $\gamma \in \Gamma_i$, and so that $\mathrm{Mod}_p(\Gamma \setminus \Gamma_i) = 0$. Let $\Gamma' = \bigcap_{\xi \in A} \Gamma_\xi$, whose complement $\Gamma_B = \mathrm{AC}(I; X) \setminus \Gamma'$ has null p-modulus.

Since $\zeta \circ \varphi$ has as upper gradient $\Phi_\zeta(x)$ on γ for each $\zeta \in A$, by considering a sequence ξ_l in A converging to $\xi \notin A$ we obtain the same conclusion.

Finally, fixing an absolutely continuous curve $\gamma \notin \Gamma_B$ there is a full measure set F^1_γ, where the components of $\varphi \circ \gamma_t$ are differentiable at $t \in F^1_\gamma$. Both sides of (4) are continuous and defined in ξ on the set F^1_γ. Since $\Phi_\xi(x)$ is an upper gradient for $\xi \circ \varphi$ along γ, there is a full measure subset $F_\gamma \subset F^1_\gamma$, where the inequality holds for $\xi \in A$. Continuity then extends it for all $\xi \in (\mathbb{R}^N)^*$ and $t \in F_\gamma$ and the claim follows by setting $E_\gamma = I \setminus F_\gamma$. □

Next, we collect some basic properties of the canonical minimal gradient. Let Φ be the map given by (4-1).

Lemma 4.3. *Set $I(\varphi)(x) := \inf_{\|\xi\|_*=1} \Phi^x(\xi)$ for μ-a.e. $x \in X$. Then:*

(1) *$I(\varphi) = \operatorname{ess\,inf}_{\|\xi\|_*=1} |D(\xi \circ \varphi)|_p$ μ-a.e. in X.*

(2) *If $U \subset X$ and $\boldsymbol{\xi} : U \to (\mathbb{R}^N)^*$ are Borel, then $\Phi^x(\boldsymbol{\xi}_x) = 0$ μ-a.e. $x \in U$ if and only if $\boldsymbol{\xi}_{\gamma_t}((\varphi \circ \gamma)'_t) = 0$ a.e. $t \in \gamma^{-1}(U)$ for p-a.e. absolutely continuous γ in X.*

(3) *If φ is p-independent on U and $f \in N^{1,p}(X)$, then the p-weak differential $\mathrm{d}f$ with respect to (U, φ), if it exists, must be unique.*

Proof of Lemma 4.3. First, we show (1). For any ξ in the unit sphere of $(\mathbb{R}^N)^*$, we have $\Phi_\xi(x) = |D(\xi \circ \varphi)|_p$ almost everywhere by Lemma 4.1. Taking an infimum on the left then gives

$$\inf_{\|\zeta\|_*=1} \Phi_\zeta(x) \le |D(\xi \circ \varphi)|_p,$$

i.e., $\inf_{\|\zeta\|_*=1} \Phi_\zeta(x) \le \operatorname{ess\,inf}_{\|\xi\|_*=1} |D(\xi \circ \varphi)|_p$ almost everywhere by the definition of an essential infimum; see Definition A.4.

On the other hand, if ξ_n, for $n \in \mathbb{N}$, is a countably dense collection in the unit sphere of $(\mathbb{R}^N)^*$, then we have $\Phi_{\xi_n}(x) = |D(\xi_n \circ \varphi)|_p \ge \operatorname{ess\,inf}_{\|\xi\|_*=1} |D(\xi \circ \varphi)|_p$ almost everywhere. By intersecting the sets where this holds for different ξ_n and since the collection is countable, we have that these hold simultaneously on a full-measure set. Specifically, $\inf_{n \in \mathbb{N}} \Phi_{\xi_n}(x) \ge \operatorname{ess\,inf}_{\|\xi\|_*=1} |D(\xi \circ \varphi)|_p$. By Lemma 4.2, we have that $\xi \to \Phi_\xi(x)$ is Lipschitz. Thus, almost everywhere,

$$\inf_{\|\xi\|_*=1} \Phi(\xi, x) = \inf_{n \in \mathbb{N}} \Phi(\xi_n, x) \ge \operatorname*{ess\,inf}_{\|\xi\|_*=1} |D(\xi \circ \varphi)|_p,$$

which gives the claim.

Next fix $\boldsymbol{\xi} : U \to (\mathbb{R}^N)^*$ as in (2). Assume first that $\Phi^x(\boldsymbol{\xi}_x) = 0$ for μ-a.e. $x \in U$. Set $C = \{x : \Phi^x(\boldsymbol{\xi}_{\gamma_x}) \ne 0\}$ with $\mu(C) = 0$. Since $\mu(C) = 0$, we have $\mathrm{Mod}_p(\Gamma^+_C) = 0$. Let Γ_B be the family of curves from Lemma 4.2. We will show the claim for $\gamma \in \mathrm{AC}(I; X) \setminus (\Gamma_B \cup \Gamma^+_C)$. By Lemma 4.2(4), we obtain a null set E_γ so that for any $\xi \in (\mathbb{R}^N)^*$ we have $|(\xi \circ \varphi \circ \gamma)'_t| \le \Phi_\xi(\gamma_t)|\gamma'_t|$ and $t \notin E_\gamma$. Let F_γ be the set of $t \notin E_\gamma$ so that

$|\gamma_t'| > 0$ and $\Phi^{\gamma_t}(\boldsymbol{\xi}_{\gamma_t}) \neq 0$. Since $0 = \int_\gamma 1_C \, \mathrm{d}s \geq \int_{F_\gamma} |\gamma_t'| \, \mathrm{d}t$, we have that the measure of F_γ is null. Now, if $t \notin E_\gamma \cup F_\gamma$, then either $|\gamma_t'| = 0$ (and the condition is vacuously satisfied), or the claim follows from $\Phi^{\gamma_t}(\boldsymbol{\xi}_{\gamma_t}) = 0$.

On the other hand, suppose that $\boldsymbol{\xi}_{\gamma_t}((\varphi \circ \gamma)_t') = 0$ for a.e. $t \in \gamma^{-1}(U)$ and p-a.e. absolutely continuous curve γ. Let $\boldsymbol{\eta}$ be the q-plan from Lemma 4.1 and $\{\boldsymbol{\pi}_x\}$ the disintegration given there. The equality $\boldsymbol{\xi}_{\gamma_t}((\varphi \circ \gamma)_t') = 0$ holds then for $\boldsymbol{\eta}$-a.e. curve and a.e. $t \in \gamma^{-1}(U)$, since $\boldsymbol{\eta}$ is a q-plan (recall Remark 2.2). Then for μ-a.e. x we have $\Phi_{\boldsymbol{\xi}}(x) = 0$ or we have $\Phi_{\boldsymbol{\xi}_x}(x) = \|\boldsymbol{\xi}_x((\varphi \circ \gamma)_t')/|\gamma_t'|\|_{L^\infty(\boldsymbol{\pi}_x)}$. In the latter case, since $\boldsymbol{\eta}$ is a q-plan, we have for μ-a.e. such x and $\boldsymbol{\pi}_x$-a.e. $(\gamma, t) \in \mathrm{Diff}(f) \cap e^{-1}(x)$ that $\boldsymbol{\xi}_x((\varphi \circ \gamma)_t') = 0$. Thus, the claim follows together with the properties of disintegrations and Corollary A.3, since the essential supremum then vanishes.

The final claim about uniqueness follows since, if $\mathrm{d}_i f$ were two p-weak differentials for $i = 1, 2$, then we could define $\boldsymbol{\xi}_x = (\mathrm{d}_1 f - \mathrm{d}_2 f)/\|\mathrm{d}_1 f - \mathrm{d}_2\|_{x,*}$ when $\mathrm{d}_1 f \neq \mathrm{d}_2 f$ and otherwise $\boldsymbol{\xi}_x = 0$. We then get immediately from the definition and the second part that $\Phi^x(\boldsymbol{\xi}) = 0$ for μ-a.e. $x \in U$. This would contradict independence. □

4C. *Charts.* The presentation here should be compared to [Cheeger 1999, Section 4], and specifically to the proof of Theorem 4.38 there, where similar arguments are employed. We first consider 0-dimensional p-weak charts. These correspond to regions of the space where no curve spends positive time.

Proposition 4.4. *Suppose (U, φ) is a 0-dimensional p-weak chart. Then*

$$\mathrm{Mod}_p(\Gamma_U^+) = 0. \tag{4-2}$$

Conversely, if $U \subset X$ is Borel and satisfies (4-2), then $(U, 0)$ is a 0-dimensional p-weak chart of X.

Proof. Since (U, φ) is a 0-dimensional p-weak chart, we have

$$|Df|_p = 0 \quad \text{for } \mu\text{-a.e. in } U, \tag{4-3}$$

for every $f \in \mathrm{LIP}_b(X)$. Let $\{x_n\} \subset X$ be a countable dense subset, and $f_n := \max\{1 - d(x_n, \cdot), 0\}$. By [Ambrosio et al. 2008, Theorem 1.1.2] (see also its proof) and (4-3) we have

$$|\gamma_t'| = \sup_n |(f_n \circ \gamma)_t'| \leq \sup_n |Df_n|_p(\gamma_t)|\gamma_t'| = 0 \quad \text{for a.e. } t \in \gamma^{-1}(U),$$

for p-a.e. $\gamma \in \mathrm{AC}(I; X)$. It follows that $\int_\gamma \chi_U \, \mathrm{d}s = 0$ for p-a.e. $\gamma \in \mathrm{AC}(I; X)$, proving (4-2).

In the converse direction, (4-2) implies, for any $f \in \mathrm{LIP}_b(X)$, that

$$\int_0^1 \chi_U(\gamma_t)|(f \circ \gamma)_t'| \, \mathrm{d}t \leq \mathrm{LIP}(f) \int_0^1 \chi_U(\gamma_t)|\gamma_t'| \, \mathrm{d}t = 0$$

for p-a.e. $\gamma \in \mathrm{AC}(I; X)$. Thus $|(f \circ \gamma)_t'| = 0$ for p-a.e. $\gamma \in \mathrm{AC}(I; X)$ and a.e. $t \in \gamma^{-1}(U)$. Then, by Theorem 1.1, together with measurability considerations from Corollary A.3, this gives $|Df|_p = 0$ μ-a.e. on U for every $f \in \mathrm{LIP}_b(X)$, showing that $(U, 0)$ is a 0-dimensional p-weak chart. □

For the remainder of this subsection we assume that $N \geq 1$ and that (U, φ) is an N-dimensional chart of X. Denote by Φ the canonical minimal gradient of φ (see Lemma 4.1).

Lemma 4.5. *The function $\xi \mapsto \Phi^x(\xi)$ is a norm on $(\mathbb{R}^N)^*$ for μ-a.e. $x \in U$. Moreover, for every $f \in \mathrm{LIP}(X)$ there exists a p-weak differential $\mathrm{d}f$. That is, a Borel measurable map $\mathrm{d}f : U \to (\mathbb{R}^N)^*$ satisfying*

$$(f \circ \gamma)'_t = \mathrm{d}_{\gamma_t} f((\varphi \circ \gamma)'_t) \quad \text{for a.e. } t \in \gamma^{-1}(U),$$

for p-a.e. absolutely continuous curves γ in X. The map $\mathrm{d}f$ is uniquely determined a.e. in U and satisfies $|Df|_p(x) = \Phi^x(\mathrm{d}f)$ μ-a.e. in U.

Remark 4.6. The equation in the statement is an equivalent formulation of the definition of the p-weak differential in Definition 1.6. Indeed, the latter follows by integration of the first, and conversely, the first follows by Lebesgue differentiation. Further, it would be enough to consider only p-a.e. curve $\gamma \in \Gamma_U^+$. Indeed, if a curve γ does not spend positive length in the set U, then $|\gamma'_t| = 0$ for a.e. $t \in \gamma^{-1}(U)$ and both sides of the equation vanish.

Proof. First, consider $f \in \mathrm{LIP}_b(X)$. Since Φ^x is a norm if and only if $I(\varphi)(x) > 0$, Lemma 4.3(1) and (1-5) imply that Φ^x is a norm for μ-a.e. $x \in U$.

Next, let $f \in \mathrm{LIP}_b(X)$ and consider the map $\psi = (\varphi, f) : X \to \mathbb{R}^{N+1}$. Let Ψ be the canonical minimal gradient of ψ. Given $\xi \in (\mathbb{R}^N)^*$ and $a \in \mathbb{R}$, we use the notation

$$(\xi, a) \in (\mathbb{R}^{N+1})^*, \quad v = (v', v_{N+1}) \mapsto \xi(v') + a v_{N+1}.$$

For μ-a.e. $x \in U$, we have $\Psi^x(\xi, 0) = \Phi^x(\xi)$ and $\Psi^x(0, a) = |a|\,|Df|_p(x)$ for every $\xi \in (\mathbb{R}^N)^*$, $a \in \mathbb{R}$ (see Lemma 4.2(3) and (4)). Since φ is a chart, we have $I(\psi) = 0$ almost everywhere. Thus, given that $I(\varphi) > 0$, $\ker \Psi^x$ is a 1-dimensional subspace of $(\mathbb{R}^{N+1})^*$. Thus for μ-a.e. $x \in U$ there exists a unique $\xi := \mathrm{d}_x f \in (\mathbb{R}^N)^*$ such that $\Psi^x(\mathrm{d}_x f, -1) = 0$, and the map $x \mapsto \mathrm{d}_x f$ is Borel; see, e.g., [Bogachev 2007, Lemma 6.7.1]. By Lemma 4.3(2), $\mathrm{d}f : U \to (\mathbb{R}^N)^*$ satisfies

$$0 = (\mathrm{d}_{\gamma_t} f, -1)((\psi \circ \gamma)'_t) = \mathrm{d}_{\gamma_t} f((\varphi \circ \gamma)'_t) - (f \circ \gamma)'_t \quad \text{for a.e. } t \in \gamma^{-1}(U),$$

for p-a.e. γ. Moreover, we have

$$||Df|_p(x) - \Phi^x(\mathrm{d}_x f)| \le |\Psi^x(0, -1) - \Psi^x(\mathrm{d}_x f, 0)| \le \Psi^x(\mathrm{d}_x f, -1) = 0$$

for μ-a.e. $x \in U$, completing the proof in the case $f \in \mathrm{LIP}_b(X)$.

The case of $f \in \mathrm{LIP}(X)$ follows through localization. Indeed, let $x_0 \in X$ be arbitrary, and consider the functions $\eta_n(x) := \min\{\max\{n - d(x_0, d), 0\}, 1\}$ for $n \in \mathbb{N}$. Then, define $f_n = \eta_n f$ so that $f_n|_{B(x_0,n-1)} = f|_{B(x_0,n-1)}$. For each f_n we can define a differential $\mathrm{d}f_n$, and $\mathrm{d}f_n|_{B(x_0,\min(m,n)-1)} = \mathrm{d}f_m|_{B(x_0,\min(m,n)-1)}$ (a.e.) for each $n, m \in \mathbb{N}$. Thus, we can define $\mathrm{d}f(x) = \mathrm{d}f_n(x)$ for $x \in B(x_0, n-1)$ with only an ambiguity on a null set. It is easy to check that $\mathrm{d}f$ is a differential. □

4D. *Differential and pointwise norm.* Let $|\cdot|_x := \Phi^x$ and define

$$\Gamma_p(T^*U) = \{\boldsymbol{\xi} : U \to (\mathbb{R}^N)^* \text{ Borel} : \|\boldsymbol{\xi}\|_{\Gamma_p(T^*U)} < \infty\}, \quad \|\boldsymbol{\xi}\|_{\Gamma_p(T^*U)} := \left(\int_U |\boldsymbol{\xi}|_x^p \,\mathrm{d}\mu\right)^{1/p}$$

(with the usual identification of elements that agree μ-a.e.). Then $(\Gamma_p(T^*U), \|\cdot\|_{\Gamma_p(T^*U)})$ is a normed space. Observe that, if $V_j := U \cap \{I(\varphi) \ge 1/j\}$, the sets $U_j := V_j \setminus \bigcup_{i<j} V_i$ partition U up to a null-set

and we have an isometric identification

$$\Gamma_p(T^*U) \simeq \bigoplus_{l^p} \Gamma_p(T^*U_j), \quad \text{where } \Gamma_p(T^*U_j) \simeq L^p(U_j; (\mathbb{R}^N)^*). \tag{4-4}$$

Thus $(\Gamma_p(T^*U), \|\cdot\|_{\Gamma_p(T^*U)})$ is a Banach space. Recall, that an ℓ_p-direct sum of Banach spaces B_i with norms $\|\cdot\|_{B_i}$ with countable index set I is defined by

$$\bigoplus_{\ell^p} B_i := \{(v_i)_{i\in I} : \|(v_i)_{i\in I}\| = (\|v_i\|_{B_i}^p)^{1/p}, v_i \in B_i\}.$$

Lemma 4.7. *Suppose $(f_n) \subset \mathrm{LIP}_b(X)$ is a sequence such that $f_n \to f$ in $L^p(X)$ and $\mathrm{d}f_n \to \boldsymbol{\xi}$ in $\Gamma_p(T^*U)$ for some $f \in N^{1,p}(X)$ and $\boldsymbol{\xi} \in \Gamma_p(T^*U)$. Then $\boldsymbol{\xi}$ is the (uniquely defined) differential of f in U, and*

$$\lim_{n\to\infty} \int_U |D(f_n - f)|_p^p \, \mathrm{d}\mu = 0.$$

In particular, $\Phi(\boldsymbol{\xi}, \cdot) = |Df|_p$ μ-a.e. in U.

Proof. By Lemma 4.5 and Fuglede's theorem [1957, Theorem 3(f)] (applied to the sequence of functions $h_n = \chi_U(\gamma_t)|\mathrm{d}_{\gamma_t} f_n - \boldsymbol{\xi}_{\gamma_t}|_{\gamma_t}$ and f_n) we can pass to a subsequence so that

$$\begin{gathered}\lim_{n\to\infty} \int_0^1 \chi_U(\gamma_t)|(f_n\circ\gamma)'_t - \boldsymbol{\xi}_{\gamma_t}((\varphi\circ\gamma)'_t)| \, \mathrm{d}t \le \lim_{n\to\infty} \int_0^1 \chi_U(\gamma_t)|\,\mathrm{d}_{\gamma_t} f_n - \boldsymbol{\xi}_{\gamma_t}|_{\gamma_t}|\gamma'_t| \, \mathrm{d}t = 0, \\ \lim_{n\to\infty} \int_0^1 |f_n(\gamma_t) - f(\gamma_t)||\gamma'_t| \, \mathrm{d}t = 0\end{gathered} \tag{4-5}$$

for p-a.e. $\gamma \in \mathrm{AC}(I; X)$. Fix a curve γ where (4-5) holds and $f_n \circ \gamma$, $f \circ \gamma$ are absolutely continuous. We may assume that γ is constant-speed parametrized. By (4-5), $f_n \circ \gamma \to f \circ \gamma$ in $L^1([0, 1])$ and $(f_n \circ \gamma)' \to g$ in $L^1(\gamma^{-1}(U))$, where $g(t) := \chi_U(\gamma_t)\boldsymbol{\xi}_{\gamma_t}((\varphi\circ\gamma)'_t)$. It follows that

$$(f\circ\gamma)'_t = \boldsymbol{\xi}_{\gamma_t}((\varphi\circ\gamma)'_t) \quad \text{a.e. } t \in \gamma^{-1}(U).$$

This shows that $\boldsymbol{\xi}$ is the differential of f, and uniqueness follows from Lemma 4.3(3). The identity $((f - f_n)\circ\gamma)'_t = (\boldsymbol{\xi}_{\gamma_t} - \mathrm{d}f_n)((\varphi\circ\gamma)'_t)$ for a.e. $t \in \gamma^{-1}(U)$, for p-a.e. $\gamma \in \mathrm{AC}(I; X)$, together with Lemma 3.2(3), implies $\Phi^x(\boldsymbol{\xi} - \mathrm{d}f_n) \le |D(f - f_n)|_p$ for μ-a.e. $x \in U$. By the convergence $\mathrm{d}(f_m - f_n) \to \boldsymbol{\xi} - \mathrm{d}f_n$ (as $m \to \infty$) we have $|D(f_m - f_n)|_p \to_{m\to\infty} \Phi^x(\boldsymbol{\xi} - \mathrm{d}f_n)$ in $L^p(U)$, and thus $|D(f - f_n)|_p \le \Phi^x(\boldsymbol{\xi} - \mathrm{d}f_n)$ μ-a.e. in U. Thus $|D(f - f_n)|_p = \Phi^x(\boldsymbol{\xi} - \mathrm{d}f_n)$ converges to zero in $L^p(U)$. The equality $\Phi_{\boldsymbol{\xi}} = |Df|_p$ follows, completing the proof. □

We say that a sequence $(\boldsymbol{\xi}_n)_n \subset \Gamma_p(T^*U)$ is equi-integrable if the sequence $\{|\boldsymbol{\xi}_n|_x\}_n \subset L^p(U)$ is equi-integrable. Recall, that a collection of integrable functions $\mathcal{F}$ is called equi-integrable, if there is a M so that $\int_X |f|^p \, \mathrm{d}\mu \le M$ for every $f \in \mathcal{F}$ and if for every $\varepsilon > 0$, there is an $\delta > 0$ and a positive measure subset Ω_ε, so that for any measurable set E with $\mu(E) \le \delta$, we have $\int_{\Omega_\varepsilon^c \cup E} |f|^p \, \mathrm{d}\mu \le \varepsilon$ for each $f \in \mathcal{F}$. By the Dunford–Pettis theorem a set of L^1 functions is equi-integrable if and only if it is sequentially compact; see for example [Dunford and Schwartz 1958, Theorem IV.8.9].

Remark 4.8. It follows from (4-4) that, if $(\boldsymbol{\xi}_n)_n \subset \Gamma_p(T^*U)$ is equi-integrable, then there exists $\boldsymbol{\xi} \in \Gamma_p(T^*U)$ such that $\boldsymbol{\xi}_n \rightharpoonup \boldsymbol{\xi}$ weakly in $\Gamma_p(T^*U)$ up to a subsequence and, by Mazur's lemma, that a

convex combination of $\boldsymbol{\xi}_n$'s converges to $\boldsymbol{\xi}$ in $\Gamma_p(T^*U)$. Indeed, the $p > 1$ case is direct and the $p = 1$ case uses the Dunford–Pettis argument above.

Next, we show that any Sobolev function $f \in N^{1,p}$ has a uniquely defined differential with respect to a chart. Note, however, that here we still postulate the existence of charts.

Proof of Theorem 1.7. The measurable norm $|\cdot|_x$ is given by Lemma 4.5. Let $f \in N^{1,p}(X)$. Lemma 4.3(3) implies that $\mathrm{d}f$, if it exists, is a.e. uniquely determined on U. Let $(f_n) \subset \mathrm{LIP}_b(X)$ be such that $f_n \to f$ and $|Df_n|_p \to |Df|_p$ in $L^p(\mu)$ as $n \to \infty$, which exists by [Eriksson-Bique 2023, Theorem 1.1]. By Lemma 4.5, $(\mathrm{d}f_n)_n \subset \Gamma_p(T^*U)$ is equi-integrable. It follows that there exists $\boldsymbol{\xi} \in \Gamma_p(T^*U)$ such that $\mathrm{d}f_n \rightharpoonup \boldsymbol{\xi}$ weakly in $L^p(T^*U)$; see Remark 4.8. By Mazur's lemma, a sequence $(g_n) \subset \mathrm{LIP}_b(X)$ of convex combinations of the f_n's converges to f in $L^p(\mu)$ and $\mathrm{d}g_n \to \boldsymbol{\xi}$ in $\Gamma_p(T^*U)$. By Lemma 4.7, $\boldsymbol{\xi} =: \mathrm{d}f$ is the differential of f. The linearity of $f \mapsto \mathrm{d}f$ follows from the uniqueness of differentials; see Lemma 4.3(3). □

The proof above also yields the following corollary. Note that, while the claim initially holds only after passing to a subsequence, since the limit is unique, the convergence holds along the full sequence.

Corollary 4.9. *Let (U, φ) be a p-weak chart of X. Suppose that $f \in N^{1,p}(X)$ and $(f_n) \subset \mathrm{LIP}_b(X)$ converges to f in energy, that is, $f_n \to_{L^p} f$ and $|Df_n|_p \to_{L^p} |Df|_p$. Then we have that $\mathrm{d}f_n \rightharpoonup \mathrm{d}f$ weakly in $\Gamma_p(T^*U)$.*

Using Lemma 4.3 we prove that the differential satisfies natural rules of calculation. The following properties are stated for $f, g \in N^{1,p}(X)$, but they would equivalently hold if we assumed only we have the local assumption $f, g \in N^{1,p}_{\mathrm{loc}}(X)$.

For the following, recall that if $A \subset \mathbb{R}$ is a measurable set, then t is a *density point* of A if

$$\lim_{h \to 0} \frac{|A \cap [t-h, t+h]|}{2h} = 1.$$

Here $|\cdot|$ denotes the Lebesgue measure of the set.

Proposition 4.10. *Let (U, φ) be an N-dimensional p-weak chart of X, $f, g \in N^{1,p}(X)$, and $F : X \to Y$ be a Lipschitz map into a metric measure space (Y, d, ν) with $F_*\mu \le C\nu$ for some $C > 0$:*

(1) *If (V, ψ) is a p-weak chart with $\varphi|_{U\cap V} = \psi|_{U\cap V}$ then the p-weak differentials of f with respect to both charts agree μ-a.e. on $U \cap V$.*

(2) *If $f|_A = g|_A$ for some A, then $\mathrm{d}f = \mathrm{d}g$ μ-a.e. on $A \cap U$.*

(3) *If $f, g \in L^\infty(X) \cap N^{1,p}(X)$, then $\mathrm{d}(fg) = f\,\mathrm{d}g + g\,\mathrm{d}f$ μ-a.e. on U.*

(4) *If $h \in C^1(\mathbb{R})$ and if $h \circ f \in N^{1,p}(X)$, then $\mathrm{d}(h \circ f) = h'(f(x))\,\mathrm{d}f(x)$ holds μ-a.e. on U.*

(5) *Let (V, ψ) be an M-dimensional p-weak chart of Y with $\mu(U \cap F^{-1}(V)) > 0$. For μ-a.e. $U \cap F^{-1}(V)$ there exists a unique linear map $D_xF : \mathbb{R}^N \to \mathbb{R}^M$ satisfying the following: if $h \in N^{1,p}(Y)$ and E is the set of $y \in V$ where the differential $\mathrm{d}_y h$ does not exist, then $\mu(U \cap F^{-1}(E)) = 0$ and*

$$\mathrm{d}_x(h \circ F) = \mathrm{d}_{F(x)}h \circ D_xF \quad \text{for } \mu\text{-a.e. } x \in U \cap F^{-1}(V \setminus E).$$

Proof. Claim (1) follows from Lemma 4.3(2) and the fact that $(\varphi\circ\gamma)'_t = (\psi\circ\gamma)'_t$ for a.e. $t\in\gamma^{-1}(U\cap V)$, for p-a.e. $\gamma\in \mathrm{AC}(I;X)$. Indeed, for p-a.e. curve and a.e. $t\in\gamma^{-1}(U\cap V)$ both derivatives agree since a generic such t will satisfy either $|\gamma'_t|=0$ or that t is a density point of $\gamma^{-1}(U\cap V)$. In both cases the equality follows.

Claim (2) is similar. Define $\mathrm{d}'f=\mathrm{d}f$ if $x\in U\setminus A$ (when defined) and $\mathrm{d}'f=\mathrm{d}g$ for $x\in A$. Now, suppose for p-almost every absolutely continuous γ we have $(f\circ\gamma)'_t=\mathrm{d}f_{\gamma_t}(\varphi\circ\gamma)'_t$ for a.e. $t\in\gamma^{-1}(U)$ and $(g\circ\gamma)'_t=\mathrm{d}g_{\gamma_t}(\varphi\circ\gamma)'_t$. We will verify for almost every $t\in\gamma^{-1}(U)$ that $(f\circ\gamma)'_t=\mathrm{d}'f_{\gamma_t}(\varphi\circ\gamma)'_t$ so that $\mathrm{d}'f$ is a differential. Then, by uniqueness it agrees with $\mathrm{d}f$. Now, almost every $t\in\gamma^{-1}(U)$ will satisfy that $(f\circ\gamma)'_t$ and $(g\circ\gamma)'_t$ exist and one (or more) of the following: $|\gamma'_t|=0$, t is a density point of $\gamma^{-1}(A)$, or t is a density point of $\gamma^{-1}(U\setminus A)$. In the first and last cases the equality $(f\circ\gamma)'_t=\mathrm{d}'f_{\gamma_t}(\varphi\circ\gamma)'_t$ is obvious. In the second case $(f\circ\gamma)'_t=(g\circ\gamma)'_t$ because t is a density point.

To prove (3) note that, since we have $(fg\circ\gamma)'_t=g(\gamma_t)(f\circ\gamma)'_t+f(\gamma_t)(g\circ\gamma)'_t$ for a.e. t for p-a.e. curve $\gamma\in\mathrm{AC}(I;X)$, it follows from (1-6) that

$$\mathrm{d}_{\gamma_t}(fg)((\varphi\circ\gamma)'_t)=g(\gamma_t)\,\mathrm{d}f_{\gamma_t}((\varphi\circ\gamma)'_t)+f(\gamma_t)\,\mathrm{d}g_{\gamma_t}((\varphi\circ\gamma)'_t)\quad\text{for a.e. } t\in\gamma^{-1}(U),$$

for p-a.e. $\gamma\in\mathrm{AC}(I;X)$. By Lemma 4.3(2) and (3) the claimed equality holds.

The argument is similar to before. Indeed, for p-a.e. absolutely continuous γ we have that $f\circ\gamma$ is absolutely continuous and $(f\circ\gamma)'_t=\mathrm{d}f_{\gamma_t}(\varphi\circ\gamma)'_t$. Then $h\circ f\circ\gamma$ is differentiable whenever $f\circ\gamma$ is, with derivative $(h\circ f\circ\gamma)'_t=h'(f(\gamma_t))(f\circ\gamma)'_t$. Therefore, $h'(f(x))\mathrm{d}f_x$ is a p-weak differential, and by uniqueness it is the p-weak differential.

Finally, for (5), let $G=(G_1,\dots,G_M)=\psi\circ F\in\mathrm{LIP}(X;\mathbb{R}^M)$ and define the expression $D_xF:=(\mathrm{d}_xG_1,\dots,\mathrm{d}_xG_M):\mathbb{R}^N\to\mathbb{R}^M$ for μ-a.e. $x\in U\cap F^{-1}(V)$. We have that

$$(\psi\circ F\circ\gamma)'_t=D_{\gamma_t}F((\varphi\circ\gamma)'_t)\quad\text{for a.e. } t\in\gamma^{-1}(U),$$

for p-a.e. $\gamma\in\mathrm{AC}(I;X)$. Note that if h and E are as in the claim, then $\mu(U\cap F^{-1}(E))\le C\nu(E)=0$. To show the claimed identity, let $\Gamma_0\subset C(I;Y)$ be a path family with $\mathrm{Mod}_p\,\Gamma_0=0$ such that

$$(h\circ\alpha)'_t=\mathrm{d}_{\alpha_t}h((\psi\circ\alpha)'_t)\quad\text{for a.e. } t\in\alpha^{-1}(V),$$

for every absolutely continuous $\alpha\notin\Gamma_0$, and set $\Gamma_1=F^{-1}\Gamma_0:=\{\gamma\in C(I;X):F\circ\gamma\in\Gamma_0\}$. Since $\mathrm{Mod}_p\,\Gamma_1\le C\,\mathrm{LIP}(F)^p\,\mathrm{Mod}_p(\Gamma_0)=0$ it follows from the two identities above that

$$(h\circ F\circ\gamma)'_t=\mathrm{d}_{F(\gamma_t)}h((\psi\circ F\circ\gamma)'_t)=\mathrm{d}_{F(\gamma_t)}h(D_{\gamma_t}((\varphi\circ\gamma)'_t))\quad\text{for a.e. } t\in\gamma^{-1}(U\cap F^{-1}(V)),$$

for p-a.e. $\gamma\in\mathrm{AC}(I;X)$. Lemma 4.3(2) and (3) imply the claim. □

4E. *Dimension bound.* In this section we give a geometric condition which guarantees that finite dimensional weak p-charts exist. This involves a bound on the size of p-independent Lipschitz maps.

As a technical tool we need the notion of a decomposability bundle $V(\nu)$ of a Radon measure ν on $\mathbb{R}^m$; see [Alberti and Marchese 2016]. We will not fully define this here, as we only need some of its properties. Firstly, let $\mathrm{Gr}(m)$ be the set of linear subspaces of $\mathbb{R}^m$ equipped with a metric $d(V,V')$ defined as the Hausdorff distance of $V\cap\overline{B(0,1)}$ to $V'\cap\overline{B(0,1)}$. The linear dimension of a subspace V is denoted

by $\dim(V)$. The decomposability bundle is then a certain Borel measurable map $\mathbb{R}^m \to \mathrm{Gr}(m)$, which associates to every $x \in \mathbb{R}^m$ a subspace $V(\nu)_x \in \mathrm{Gr}(m)$. In a sense, this bundle measures the directions in which a Lipschitz function must be differentiable in (at almost every point). We collect the main properties we need for this bundle and briefly cite where the proofs of these claims can be found.

Theorem 4.11. *Suppose that ν is a Radon measure on $\mathbb{R}^m$. Then there exists a decomposability bundle $V(\nu)$ with the following properties*:

(1) *If* $\dim(V(\nu)_x) = m$ *for ν-a.e. $x \in \mathbb{R}^m$, then $\nu \ll \lambda$.*

(2) *There is a Lipschitz function $f : \mathbb{R}^m \to \mathbb{R}$ so that for ν-a.e. $x \in \mathbb{R}^m$ we have that the directional derivative of f does not exist in the direction v for any $v \notin V(\nu)_x$.*

(3) *If $\nu' \ll \nu$, then $V(\nu')_x = V(\nu)_x$ for ν'-a.e. $x \in \mathbb{R}^m$.*

Proof. The first follows from [De Philippis and Rindler 2016, Theorem 1.14] when combined with [Alberti and Marchese 2016, Theorem 1.1(i)]. The second claim follows from [loc. cit., Theorem 1.1(ii)]. Note that the second claim is vacuous for those points $x \in \mathbb{R}^m$ where the decomposability bundle has dimension m. The third claim is [loc. cit., Proposition 2.9(i)]. □

The following lemma gives a modulus perspective to the decomposability bundle.

Lemma 4.12. *Assume $N \geq 1$, $\varphi : X \to \mathbb{R}^N$ is Lipschitz, $U \subset X$ is a Borel set of bounded measure and $\nu = \varphi_*(\mu|_U)$. Then, for p-a.e. curve γ and almost every $t \in \gamma^{-1}(U)$ we have that $(\varphi \circ \gamma)'_t$ exists and $(\varphi \circ \gamma)'_t \in V(\nu)_{\varphi(\gamma_t)}$.*

Proof. By part (ii) of Theorem 4.11, there is a Lipschitz function $f : \mathbb{R}^N \to \mathbb{R}$, so that for ν-almost every $x \in \mathbb{R}^N$ and any $v \notin V(\nu)_x$ we have that the directional derivative $D_v(f) = \lim_{h\to 0}(f(x+hv) - f(x))/h$ does not exist. Let $A \subset \mathbb{R}^N$ be a full ν-measure Borel set so that this claim holds.

Let $B = \varphi^{-1}(\mathbb{R}^N \setminus A) \cap U$, which is μ-null. The family Γ_B^+ has null p-modulus. We will show that the claim holds for p-a.e. $\gamma \in \mathrm{AC}(I; X) \setminus \Gamma_B^+$. The derivatives $(\varphi \circ \gamma)'_t$ and $(f \circ \varphi \circ \gamma)'_t$ exist for almost every $t \in \gamma^{-1}(U)$. Also, for a.e. $t \in I$ we can either take $|\gamma'_t| = 0$ or $\gamma_t \notin B$ and so $(\varphi \circ \gamma)_t \notin A$, since $\gamma \notin \Gamma_B^+$. If $|\gamma'_t| = 0$, then $(\varphi \circ \gamma)'_t = 0 \in V(\nu)_{\varphi(\gamma_t)}$. In the other case, when $\gamma_t \notin B$, the function f does not have a directional derivative for $v \notin V(\nu)_{(\varphi\circ\gamma)_t}$. The only way for both $(\varphi \circ \gamma)'_t$ and $(f \circ \varphi \circ \gamma)'_t$ to exist then is if $(\varphi \circ \gamma)'_t \in V(\nu)_{\varphi(\gamma_t)}$, which gives the claim. □

The following should be compared to [Cheeger 1999, Lemma 4.37].

Proposition 4.13. *Suppose $\varphi \in \mathrm{LIP}(X; \mathbb{R}^N)$ is p-independent on U. Then $N \leq \dim_H U$.*

Proof. By restriction to a subset of the form $U \cap B(x_0, R)$ for $x_0 \in X$, $R > 0$, of positive measure, it suffices to assume that U has finite measure. The claim is automatic, if $\dim_H U = \infty$. Thus, assume that the Hausdorff dimension is finite. Set $\nu = \varphi_*(\mu|_U)$ and let $V(\nu)$ be the decomposability bundle of ν. If $V(\nu)_x$ has dimension N for almost every x with respect to ν, then $\nu \ll \lambda$ by Theorem 4.11(1) and thus $\mathcal{H}^N(\varphi(U)) > 0$, since ν is concentrated on $\varphi(U)$. Then $N \leq \dim_H(\varphi(U)) \leq \dim_H(U)$.

Suppose then to the contrary, that there exists a subset $A \subset U$ with positive ν-measure where $V(\nu)_x$ has dimension less than $\dim_H(U)$ for each $x \in A$. We can take A to be Borel. Consider $\mu' = \mu|_{\varphi^{-1}(A)}$, which has

push-forward $\nu' = \nu|_A = \varphi_*(\mu')$. By the third part in Theorem 4.11 we have that $V(\nu')_x = V(\nu)_x$ for ν'-a.e. $x \in A$. Further $\varphi^{-1}(A) \subset U$, so φ is still p-independent on $\varphi^{-1}(A) = U'$. Now, by considering U' instead of U and ν' instead of ν, we have that $V(\nu')_{\varphi(x)}$ has dimension less than N for ν'-almost every $x \in U$. In the following, we simplify notation by dropping the primes, and restricting to the positive measure subset U' so constructed. For ν-almost every $x \in U$, we have that $V(\nu)_{\varphi(x)}$ is a strict subspace of $\mathbb{R}^N$, and thus there are vectors perpendicular to these. Since $x \to V(\nu)_{\varphi(x)}$ is Borel, we can choose a Borel map $x \to \boldsymbol{\xi}_x \in (\mathbb{R}^N)^*$ so that $\boldsymbol{\xi}_x$ is a unit vector that vanishes on $V(\nu)_{\varphi(x)}$ for μ-a.e. $x \in U$ (see, e.g., [Bogachev 2007, Theorem 6.9.1], which is an instance of a Borel selection theorem). Let $\widetilde{U} \subset U$ be the full measure subset where these properties hold for every $x \in \widetilde{U}$. Now, by Lemma 4.12 we have for p-a.e. curve γ that $(\varphi\circ\gamma)'_t \in V(\nu)_{\varphi(\gamma_t)}$ for almost every $t \in \gamma^{-1}(U)$. The set $U \setminus \widetilde{U}$ has null measure, and thus $\Gamma^+_{U\setminus\widetilde{U}}$ has null modulus.

Thus, for p-a.e. curve $\gamma \in \mathrm{AC}(I; X)$ and a.e. $t \in \gamma^{-1}(U)$ we can further assume $\gamma_t \in U$ or $|\gamma'_t| = 0$. Therefore, $\boldsymbol{\xi}_{\gamma_t}((\varphi\circ\gamma)'_t) = 0$ for almost every $t \in \gamma^{-1}(U)$ and such curves γ. By part (2) of Lemma 4.3, we have that $I(\varphi) \le \Phi^x(\boldsymbol{\xi}_x) = 0$ for μ-a.e. $x \in U$. This contradicts p-independence and proves the claim. □

4F. *Sobolev charts.* By definition, a p-weak chart is a Lipschitz map which has target of maximal dimension with respect to Lipschitz maps. The notions of p-independence and maximality, however, are well-defined for any Sobolev map, and in fact p-weak charts *could* be required to have Sobolev (instead of Lipschitz) regularity. Despite the apparent difference of the alternative definition, the existence of maximal p-independent Sobolev maps also guarantees the existence of p-weak chart of the same dimension. This follows from the energy density of Lipschitz functions, see [Eriksson-Bique 2023], together with results of the previous subsection.

Proposition 4.14. *Suppose $p \ge 1$, and $\varphi \in N^{1,p}(X; \mathbb{R}^N)$ is p-independent and p-maximal in a bounded Borel set $U \subset X$. For any $\varepsilon > 0$ there exists $V \subset U$ with $\mu(U \setminus V) < \varepsilon$, and a Lipschitz function $\psi : X \to \mathbb{R}^N$ such that (V, ψ) is an N-dimensional p-weak chart.*

Proof. For any $V \subset U$ with $\mu(V) > 0$, let n_V be the supremum of numbers n so that there exists $\psi \in \mathrm{LIP}_b(X; \mathbb{R}^n)$ which is p-independent on a positive measure subset of V. By the maximality of N we have that $n_V \le N$. Thus n_V is attained for every such V and, by [Keith 2004a, Proposition 3.1], there is a partition of U up to a null-set by p-weak charts V_i, $i \in \mathbb{N}$, of dimension $\le N$. By [Eriksson-Bique 2023, Theorem 1.1], Corollary 4.9 (with a diagonal argument) and Mazur's lemma we have that, for each component $\varphi_k \in N^{1,p}(X)$ of φ, there exists a sequence $(\psi^n_k) \subset \mathrm{LIP}_b(X)$ with $|D(\varphi_k - \psi^n_k)|_p \to 0$ in $L^p(V_i)$. Thus, $|D(\varphi_k - \psi^n_k)|_p \to 0$ in $L^p(U)$. Here, we use that $|D(\varphi_k - \psi^n_k)|_p \le |D\varphi_k|_p + |D\psi^n_k|_p$ and the L^p-convergence of the right-hand side from [Eriksson-Bique 2023].

If Φ and Ψ_n denote the canonical minimal gradients associated to φ and $\psi^n := (\psi^n_1, \dots, \psi^n_N)$, we have

$$\sup_{\|\xi\|_*=1} |\Phi(\xi, \cdot) - \Psi_n(\xi, \cdot)| \le \operatorname*{ess\,sup}_{\|\xi\|_*=1} |D(\xi \circ (\varphi - \psi^n))|_p \le \sum_{k=1}^N |D(\varphi_k - \psi^n_k)|_p \quad \mu\text{-a.e. in } U.$$

It follows that

$$\lim_{n\to\infty} \mu(U \setminus \{I(\psi^n) > 0\}) = 0,$$

completing the proof, since ψ^n is p-independent and maximal on the set $\{I(\psi^n) > 0\}$. □

Another condition in this context is strong maximality: a map $\varphi \in N^{1,p}(X;\mathbb{R}^N)$ is *strongly maximal* in $U \subset X$ if no positive measure subset $V \subset U$ admits a p-independent Sobolev map into a higher-dimensional Euclidean space. This condition excludes not only Lipschitz, but also Sobolev functions into higher-dimensional targets, and is thus a priori stronger than maximality. However, it follows from Proposition 4.14 that a maximal p-independent Sobolev map is also strongly maximal. Conversely, if one has a Lipschitz chart, then the Lipschitz chart is also strongly maximal.

4G. *p-weak charts in Poincaré spaces.* Recall that a metric measure space $X = (X, d, \mu)$ is said to be a p-PI space if μ is doubling, and X supports a weak $(1, p)$-Poincaré inequality: there exist constants $C, \sigma > 0$ so that, for any $f \in L^1(X)$ with upper gradient g, we have

$$\fint_B |f - f_B|\,\mathrm{d}\mu \le Cr\left(\fint_{\sigma B} g^p\,\mathrm{d}\mu\right)^{1/p}$$

for all balls $B \subset X$ of radius r. Here $h_B = \fint_B h\,\mathrm{d}\mu = (1/\mu(B))\int_B h\,\mathrm{d}\mu$ for a ball $B \subset X$ and $h \in L^1(B)$. The celebrated result from [Cheeger 1999] states that a PI-space admits a Lipschitz differentiable structure. We will return to this structure in Section 6B, but here recall the constructions from [Cheeger 1999, Section 4]. Cheeger's paper does not employ the following terminology, but it simplifies and clarifies our presentation.

Given a Lipschitz map $\varphi : X \to \mathbb{R}^N$ and a positive measure subset $U \subset X$ the pair (U, φ) is called *a Cheeger chart* if for every Lipschitz map $f : X \to \mathbb{R}$ and a.e. $x \in U$ there is a unique element $\mathrm{d}_{C,x} f \in (\mathbb{R}^N)^*$ satisfying

$$\mathrm{Lip}(\mathrm{d}_{C,x} f \circ \varphi - f)(x) = 0. \tag{4-6}$$

This equality is equivalent to (1-4).

Proof of Theorem 1.8. Let (U, φ) be a p-weak chart of dimension N and let $f \in \mathrm{LIP}(X)$. Denote by Φ the canonical minimal gradient of $(\varphi, f) : X \to \mathbb{R}^{N+1}$; see Lemma 4.1. Since X is a p-PI space, it follows that $\mathrm{Lip}\, h = |Dh|_p$ μ-a.e. for any $h \in \mathrm{LIP}(X)$; see [Cheeger 1999, Theorem 6.1]. (In fact, the slightly easier comparability from Lemma 4.35 of that work suffices for the following.) Then, for any $\xi \in (\mathbb{R}^N)^*$ and for μ-a.e. $x \in U$, we have

$$\mathrm{Lip}(\xi \circ \varphi - f)(x) = \Phi^x(\xi, -1), \quad \xi \in (\mathbb{R}^N)^*.$$

Arguing using in the proof of Lemmas 4.1 and 4.5 we obtain this equality, simultaneously, for a.e. $x \in U$ and for any $\xi \in A$ for a dense subset of $A \subset (\mathbb{R}^N)^*$. From this, and the continuity of both sides in ξ, we obtain that for μ-a.e. $x \in U$, the equality holds simultaneously for all $\xi \in (\mathbb{R}^N)^*$.

Since the p-weak differential $\mathrm{d}f$ is characterized by the property $\Phi^x(\mathrm{d}f, -1) = 0$ for μ-a.e. $x \in U$, it follows that, for μ-a.e. $x \in U$, $\mathrm{d}_x f \in (\mathbb{R}^N)^*$ satisfies (4-6). Thus (U, φ) is a Cheeger chart. The uniqueness follows from the equality in a similar way. □

Remark 4.15. The proof of Theorem 1.8 also yields the claim under the weaker assumption $\mathrm{Lip}\, f \le \omega(|Df|_p)$ for some collection of moduli of continuity ω (compare Theorem 1.10) since the equality $\mathrm{Lip}\, f = |Df|_p$ follows from this by [Ikonen et al. 2022, Theorem 1.1].

5. The p-weak differentiable structure

5A. *The p-weak cotangent bundle.* A *measurable L^∞-bundle* $\mathcal{T}$ over X consists of a collection $(\{U_i, V_{i,x}\})_{i\in I}$ together with a collection $(\{\phi_{i,j,x}\})$ of transformations with a countable index set I, where:

(1) $U_i \subset X$ are Borel sets for each $i \in I$, and cover X up to a μ-null set.

(2) For any $i \in I$ and μ-a.e. $x \in U_i$, $V_{i,x} = (V_i, |\cdot|_{i,x})$ is a finite-dimensional normed space so that $x \mapsto |v|_{i,x}$ is Borel for any $v \in V_i$.

(3) For any $i, j \in I$ and μ-a.e. $x \in U_i \cap U_j$, $\phi_{i,j,x} : V_{i,x} \to V_{j,x}$ is an isometric bijective linear map satisfying the *cocycle condition*: for any $i, j, k \in I$ and μ-a.e. $x \in U_i \cap U_j \cap U_k$, we have $\phi_{j,k,x} \circ \phi_{i,j,x} = \phi_{i,k,x}$.

For each $i \in I$ and μ-a.e. $x \in U_i$, we denote by $\mathcal{T}_x$ the equivalence class of the normed vector space $V_{i,x}$ under identification by isometric isomorphisms. By (3), $\mathcal{T}_x$ is well-defined for μ-a.e. $x \in X$.

We now show that a p-weak differentiable structure $\mathscr{A}$ on X gives rise to a measurable bundle.

Proposition 5.1. *Let $p \geq 1$, and let $\{(U_i, \varphi_i)\}$ be an atlas of p-weak charts on X. The collection $\{(U_i, (\mathbb{R}^{N_i})^*, |\cdot|_{i,x})\}$ forms a measurable bundle over X, the transformations given by the collection $\{D\Phi_{i,j,x}\}$ constructed in Lemma 5.2.*

First, we construct the transformation maps.

Lemma 5.2. *Let (U_i, φ^i) be N_i-dimensional p-weak charts on X, with corresponding differentials d^i and norms $|\cdot|_{i,x}$ for $i = 1, 2$. If $\mu(U_1 \cap U_2) > 0$, then $N_1 = N_2 := N$ and, for μ-a.e. $x \in U_1 \cap U_2$, there exists a unique bijective isometric isomorphism $D\Phi_{1,2,x} : ((\mathbb{R}^N)^*, |\cdot|_{1,x}) \to ((\mathbb{R}^N)^*, |\cdot|_{2,x})$ such that $\mathrm{d}^1 f = \mathrm{d}^2 f \circ D\Phi_{1,2,x}$. Further $D\Phi_{1,2,x}$ satisfies the measurability constraint* (2).

In the proof, we denote by $\varphi^i_1, \dots, \varphi^i_{N_i}$ the components of φ^i.

Proof. For μ-a.e. $x \in U_1 \cap U_2$, define

$$D_x = D = (\mathrm{d}^1\varphi^1_1, \dots, \mathrm{d}^2\varphi^1_{N_1}) : \mathbb{R}^{N_2} \to \mathbb{R}^{N_1}.$$

D is a linear map satisfying, for all $\xi \in (\mathbb{R}^{N_1})^*$,

$$\xi \circ D((\varphi^2 \circ \gamma)'_t) = \xi((\varphi^1 \circ \gamma)'_t) \quad \text{for a.e. } t \in \gamma^{-1}(U_1 \cap U_2), \tag{5-1}$$

for p-a.e. $\gamma \in \Gamma^+_{U_1 \cap U_2}$. Note that, by the uniqueness of differentials, D is the unique linear map satisfying (5-1) for p-a.e. curve. By Lemma 4.3(2) it follows that

$$|\xi \circ D|_{2,x} = |\xi|_{1,x}, \quad \xi \in (\mathbb{R}^{N_1})^*,$$

for μ-a.e. $x \in U_1 \cap U_2$. Thus D^* is an isometric embedding and in particular $N_1 \leq N_2$. Reversing the roles of φ^1 and φ^2 we obtain that $N_1 = N_2$ and consequently $D\Phi_{1,2,x} := D^*_x : ((\mathbb{R}^{N_1})^*, |\cdot|_{1,x}) \to ((\mathbb{R}^{N_2})^*, |\cdot|_{2,x})$ is an isometric isomorphism for μ-a.e. $x \in U_i \cap U_j$.

For any $f \in N^{1,p}(X)$, the identity $\mathrm{d}^1_x f = \mathrm{d}^2_x f \circ D\Phi_{1,2,x}$ for μ-a.e. $x \in U_1 \cap U_2$ follows from (5-1) and (1-6). $\square$

Proof of Proposition 5.1. Conditions (1) and (2) are satisfied by Lemma 4.2. The cocycle condition follows from Lemma 5.2. □

Definition 5.3. We call the measurable bundle given by Proposition 5.1 the p-weak cotangent bundle and denote it by T_p^*X. We define $T_{p,x}^*X = ((\mathbb{R}^N)^*, |\cdot|_x)$ and $T_{p,x}X = (\mathbb{R}^N, |\cdot|_{*,x})$ for almost every $x \in U$, where (U, φ) is an N-dimensional p-weak chart and $|\cdot|_x$ the norm given by the canonical minimal gradient Φ; see Lemmas 4.1 and 4.5. The spaces $T_{p,x}$ are here defined pointwise almost everywhere. By considering the adjoints of transition maps in the definition above, one can patch these together to form a measurable L^∞ tangent bundle, which is dual to T_p^*X, whose fibers are $T_{p,x}X$.

The next proposition establishes the existence of a p-weak differentiable structure under a mild finite dimensionality condition.

Proposition 5.4. *Suppose X is a metric measure space and $\{X_i\}_{i\in\mathbb{N}}$ a covering of X with $\dim_H X_i < \infty$. Then, for any $p \geq 1$, X admits a p-weak differentiable structure. Moreover, $N \leq \dim_H X_i$ whenever (U, φ) is an N-dimensional p-weak chart with $\mu(U \cap X_i) > 0$.*

Proof. For any Borel set $U \subset X$ with $\mu(U) > 0$ there exists $i \in \mathbb{N}$ such that $\mu(U \cap X_i) > 0$. By Proposition 4.13 we have that $N \leq \dim_H(U \cap X_i)$ whenever $\varphi \in \mathrm{LIP}_b(X; \mathbb{R}^N)$ is p-independent in a positive measure subset of $U \cap X_i$. Using [Keith 2004b, Proposition 3.1] we can cover X up to a null-set by Borel sets U_k for which there exist $\varphi_k \in \mathrm{LIP}_b(X; \mathbb{R}^{N_k})$ that are p-independent and p-maximal on U_k. The collection $\{(U_k, \varphi_k)\}_{k\in\mathbb{N}}$ is a p-weak differentiable structure on X. The last claim follows by the argument above. □

5B. *Sections of measurable bundles.* A measurable bundle $\mathcal{T}$ over X comes with a projection map $\pi : \mathcal{T} \to X$, $(x, v) \mapsto x$, and a *section* of $\mathcal{T}$ is a collection $\omega = \{\omega_i : U_i \to V_i\}$ of Borel measurable maps satisfying $\pi \circ \omega_i = \mathrm{id}_{U_i}$ μ-a.e. and $\phi_{i,j,x}(\omega_i) = \omega_j$ for each $i, j \in I$ and almost every $x \in U_i \cap U_j$. Observe that the map $x \mapsto |\omega(x)|_x$ given by

$$|\omega(x)|_x := |\omega_i(x)|_{i,x} \quad \text{for } \mu\text{-a.e. } x \in U_i \tag{5-2}$$

is well-defined up to negligible sets by the cocycle condition and the fact that $\phi_{i,j,x}$ is isometric.

Definition 5.5. For $p \in [1, \infty]$, let $\Gamma_p(\mathcal{T})$ be the space of sections ω of $\mathcal{T}$ with

$$\|\omega\|_p := \|x \mapsto |\omega(x)|_x\|_{L^p(\mu)} < \infty.$$

We call $\Gamma_p(\mathcal{T})$ the space of p-integrable sections of $\mathcal{T}$. The space $\Gamma_p(T_p^*X)$ is called the p-weak cotangent module.

Note that $\Gamma_p(\mathcal{T})$, equipped with the pointwise norm (5-2) and the natural addition and multiplication operations, is a normed module in the sense of [Gigli 2015]. Recall that an L^p-normed L^∞-module over X is a Banach module $(\mathscr{M}, \|\cdot\|)$ over $L^\infty(X)$, equipped with a pointwise norm $|\cdot| : X \to \mathbb{R}$ that satisfies

$$|gm| = |g|\,|m| \quad \text{and} \quad \|m\| = \left(\int_X |m|_x^p \,\mathrm{d}\mu(x)\right)^{1/p}$$

for all $m \in \mathscr{M}$ and $g \in L^\infty(X)$. We refer to [Gigli 2015; 2018] for a detailed account of the theory of normed modules.

Next we consider the p-weak cotangent module $\Gamma_p(T_p^* X)$. For a p-weak chart (U, φ) of X and $f \in N^{1,p}(X)$, denote by $\mathrm{d}_{(U,\varphi)} f$ the differential of f with respect to (U, φ). Lemma 5.2 implies that the collection of differentials with respect to different charts satisfies the compatibility condition above.

Definition 5.6. Let $p \geq 1$, and suppose $\mathscr{A}$ is a p-weak differentiable atlas of X. For any $f \in N^{1,p}(X)$, the differential $\mathrm{d}f \in \Gamma_p(T_p^* X)$ is the element in the p-weak cotangent module defined by the collection $\{\mathrm{d}_{(U,\varphi)} f : U \to (\mathbb{R}^N)^*\}_{(U,\varphi)\in\mathscr{A}}$.

We record the following properties of the differential.

Proposition 5.7. *Let $A \subset X$ be a Borel set and $F : X \to Y$ a Lipschitz map to a metric measure space (Y, d, ν) admitting a p-weak differentiable structure, with $F_*\mu \leq C\nu$.*

(1) *If $f, g \in N^{1,p}(X)$ agree on $A \subset X$, then $\mathrm{d}f = \mathrm{d}g$ μ-a.e. on A.*

(2) *If $f, g \in N^{1,p}(X) \cap L^\infty(X)$, then $\mathrm{d}(fg) = g\,\mathrm{d}f + f\,\mathrm{d}g$ μ-a.e.*

(3) *If E is the set of $y \in Y$ for which $T^*_{p,y}Y$ does not exist, then $\mu(F^{-1}(E)) = 0$ and, for μ-a.e. $x \in X \setminus F^{-1}(E)$ there exists a unique linear map $D_x F : T_{p,x}X \to T_{p,F(x)}Y$ such that*

$$\mathrm{d}_x(h \circ F) = \mathrm{d}_{F(x)} h \circ D_x F \quad \text{for } \mu\text{-a.e. } x,$$

for every $h \in N^{1,p}(Y)$.

(4) *If $h \in C^1(\mathbb{R})$ and if $h \circ f \in N^{1,p}(X)$, then $\mathrm{d}(h \circ f) = h'(f(x))\,\mathrm{d}f$.*

(5) *If $f_i \in N^{1,p}(X)$ and there is a function $f \in L^p$ and a $w \in \Gamma_p(T_p^* X)$ so that $\lim_{i\to\infty} f_i = f(x)$ converges in $L^p(X)$ and $\mathrm{d}f_i \to w$ converges in $\Gamma_p(T_p^* X)$, then, there is a function $\tilde{f} \in N^{1,p}(X)$ so that $\tilde{f} = f$ almost everywhere with $\mathrm{d}\tilde{f} = w$.*

Proof. The proofs of the first four claims follow directly from Proposition 4.10 together with the compatibility condition of sections. Indeed, one can verify the identities for each chart (U, φ), from which the identities follows for everywhere.

Consider now $f_i \in N^{1,p}(X)$ which converge in $L^p(X)$ to $f \in L^p(X)$ and so that $\mathrm{d}f_i$ converge in $\Gamma_p(T_p^* X)$ to $w \in \Gamma_p(T_p^* X)$. We have therefore that $g_i = |Df_i|_p = |\mathrm{d}f_i|$ converges in L^p to $g = |w|$. By Fuglede's theorem [1957, Theorem 3(f)] we can pass to a subsequence so that $f_i \to \tilde{f}$ converges pointwise and so that

$$\int_\gamma |f_i - \tilde{f}|\,\mathrm{d}s \to 0 \quad \text{and} \quad \int_\gamma |g_i - g|\,\mathrm{d}s \to 0$$

for p-a.e. absolutely continuous curves $\gamma : [0, 1] \to X$. Then, for all such curves, we have that $|f_i(\gamma(0)) - f_i(\gamma(1))| \leq \int_\gamma g_i \mathrm{d}s$, which converges to

$$|\tilde{f}(\gamma(0)) - \tilde{f}(\gamma(1))| \leq \int_\gamma g\,\mathrm{d}s.$$

Thus $\tilde{f} \in N^{1,p}$. Finally, one only needs to show that $w = \mathrm{d}\tilde{f}$. This follows by another diagonal argument and computing $\mathrm{d}\tilde{f}$ in charts using the argument from Lemma 4.7. □

We finish the subsection with a proof of the density of Lipschitz functions in Newtonian spaces.

Proof of Theorem 1.9. Let $f \in N^{1,p}(X)$. By [Eriksson-Bique 2023], there exists a sequence $(f_n) \subset \mathrm{LIP}_b(X)$ with $f_n \to f$ and $|Df_n|_p \to |Df|_p$ in $L^p(\mu)$. It follows that $(\mathrm{d}f_n) \subset \Gamma_p(T^*X)$ is equi-integrable, and Remark 4.8 and Lemma 4.7, together with a diagonalization argument over a union of charts covering X, show that $\mathrm{d}\tilde{f}_n \to \mathrm{d}f$ in $\Gamma_p(T^*X)$ for convex combinations $\tilde{f}_n \in \mathrm{LIP}_b(X)$ of f_n's. Consequently $|D(\tilde{f}_n - f)|_p \to 0$ in $L^p(\mu)$. □

5C. *Dependence of the p-weak differentiable structures on p.* Suppose $1 \le p < q$. We have that $|Df|_p \le |Df|_q$ μ-a.e. for every $f \in \mathrm{LIP}_b(X)$, and the inequality may be strict; see [Di Marino and Speight 2015]. As a consequence, if $\varphi \in \mathrm{LIP}_b(X; \mathbb{R}^N)$ is q-maximal in $U \subset X$, then it is p-maximal. It follows (using this dimension upper bound and [Keith 2004b, Proposition 3.1]) that if X admits a q-weak differentiable structure then X also admits a p-weak differentiable structure. We remark that the structures may be different.

For the following statement we say that a *bundle map* $\pi : \mathcal{T} \to \mathcal{T}'$ between two measurable bundles $\mathcal{T} = (\{U_i, V_{i,x}\}, \{\phi_{i,l,x}\})_{i \in I}$ and $\mathcal{T}' = (\{U'_j, V'_{j,x}\}, \{\psi_{j,k,x}\})_{j \in J}$ over X is a collection of linear maps $\{\pi_{i,j,x} : V_i \to V'_j\}$ for μ-a.e. $x \in U_i \cap U'_j$ such that

(a) for each $i \in I$, $j \in J$ the map $x \mapsto \pi_{i,j,x}(v) : U_i \cap U'_j \to V'_j$ is Borel for any $v \in V_i$,

(b) for each $i, l \in I$, $j, k \in J$ and μ-a.e. $x \in U_i \cap U_l \cap U'_j \cap U'_k$, we have the compatibility condition $\psi_{j,k,x} \circ \pi_{i,j,x} = \phi_{l,j,x} \circ \pi_{i,l,x}$.

When the underlying index sets agree and $U_i = V_i$ for all $i \in I$, it is sufficient to consider the family $\{\pi_{i,x} := \pi_{i,i,x}\}$, since these determine a unique bundle map.

Proposition 5.8. *Suppose $q > p \ge 1$ and X admits a q-weak differentiable structure. Then X admits p-weak differentiable structure and there is a bundle map $\pi_{p,q} : T^*_q X \to T^*_p X$ which is a linear 1-Lipschitz surjection μ-a.e. Moreover, this map satisfies $\pi_{p,q} = \pi_{p,s} \circ \pi_{s,q}$ for $q > s > p$, and $\pi_{p,q}(\mathrm{d}_q f) = \mathrm{d}_p f$ for any $f \in \mathrm{LIP}_b(X)$, where $\mathrm{d}_q f, \mathrm{d}_p f$ are the p- and q-weak differentials respectively.*

Proof. Since X admits a q-differential structure, we can find q-charts $(U_i, \varphi_{q,i})$ so that $X = \bigcup_{i \in \mathbb{N}} U_i \cup N$, with $\mu(N) = 0$, and $\varphi_{q,i} \in N^{1,p}(X; \mathbb{R}^{m_i})$ is Lipschitz. Assume that U_i are chosen to be pairwise disjoint. As $|Df|_p \le |Df|_q$ (a.e.) for any $f \in \mathrm{LIP}_b(X)$, any p-independent map is also q-independent. Any map $\varphi \in N^{1,p}(X; \mathbb{R}^n)$ which is p-independent on some positive-measure subset of U_i must have $n \le m_i$; see Proposition 4.14. By [Keith 2004b, Proposition 3.1] and this dimension bound we can cover X by maximal p-independent maps, i.e., charts, $(V_j, \varphi_{p,j})$. By considering the countable collection of sets $V_i \cap U_j$, and reindexing, we may assume that $(U_i, \varphi_{q,i})$ and $(U_i, \varphi_{p,i})$ are q- and p- charts, respectively.

We define the matrix A_x for $x \in U_i$ by taking as rows the vectors $\mathrm{d}_{i,p}\varphi^k_{q,i}$ for each component $k = 1, \dots, m_i$. We define the bundle map $\pi_{p,q}$ by setting $\pi^x_{p,q}(\xi) = \xi \circ A_x$ for μ-a.e. $x \in U_i$. For each ξ we get $\mathrm{d}_p(\xi \circ \varphi_{q,i}) = \xi \circ A_x$. Thus, for p-a.e. curve $\gamma \in \mathrm{AC}(I; X)$ and a.e. $t \in \gamma^{-1}(U)$ we have

$$\xi(\varphi_{q,i} \circ \gamma)'_t = (\xi \circ A_x)(\varphi_{p,i} \circ \gamma)'_t.$$

By the definition of the differential, we get immediately that $\pi_{p,q}(\mathrm{d}_q f) = \mathrm{d}_p f$ for every $f \in \mathrm{LIP}_b(X)$. Thus, the 1-Lipschitz property follows immediately from the definition of norms combined with $|Df|_p \leq |Df|_q$. The map is clearly a surjective bundle map as well, and by uniqueness of the p-differential, we automatically get $\pi_{p,s} \circ \pi_{s,q} = \pi_{p,q}$. □

6. Relationship with Cheeger's and Gigli's differentiable structures

6A. ***Gigli's cotangent module.*** Fix $p \geq 1$. Gigli's cotangent module is the L^p-normed L^∞-module given by the following theorem.

Theorem 6.1. *There exists an L^p-normed L^∞-module $L^p(T^*X)$, with pointwise norm denoted by $|\cdot|_G$, and a bounded linear map $\mathrm{d}_G : N^{1,p}(X) \to L^p(T^*X)$ satisfying*

$$|\mathrm{d}_G f|_G = |Df|_p, \quad f \in N^{1,p}(X), \tag{6-1}$$

such that the subspace $\mathcal{V}$ defined by

$$\mathcal{V} := \left\{ \sum_j^M \chi_{A_j} \mathrm{d}_G f_j : (A_j)_j \text{ Borel partition of } X,\ f_j \in N^{1,p}(X) \right\}$$

is dense in $L^p(T^*X)$. The module $L^p(T^*X)$ is uniquely determined up to isometric isomorphism of normed modules by these properties.

Following [Gigli 2018, Definition 1.4.1] we say that a collection $\{v_1, \dots, v_N\} \subset L^p(T^*X)$ is *linearly independent* in a Borel set $U \subset X$ if, whenever $g_1, \dots, g_N \in L^\infty(X)$ satisfy $\left|\sum_j^N g_j v_j\right|_G = 0$ μ-a.e. on U, we have $g_1 = \cdots = g_N = 0$ μ-a.e. in U. A linearly independent collection $\{v_1, \dots, v_N\}$ in U is a *basis* of $L^p(T^*X)$ in U if, for any $v \in L^p(T^*X)$, there exists a Borel partition $\{U_i\}_{i\in\mathbb{N}}$ of U and $g_1^i, \dots, g_N^i \in L^\infty(X)$ such that $\left|v - \sum_j^N g_j^i v_j\right|_G = 0$ μ-a.e. on U_i, for every $i \in \mathbb{N}$.

Definition 6.2. Let $p \geq 1$. The cotangent module $L^p(T^*X)$ is locally finitely generated if there exists a Borel partition such that $L^p(T^*X)$ has a finite basis in each set of the partition.

By [Gigli 2018, Proposition 1.4.5], there exists a Borel partition $\{A_N\}_{N\in\mathbb{N}\cup\{\infty\}}$ of X such that $L^p(T^*X)$ has a basis of N elements on A_N for each $N \in \mathbb{N} \cup \{\infty\}$. We call the partition $\{A_N\}$ the *dimensional decomposition* of X. Notice that $L^p(T^*X)$ is locally finitely generated if and only if $\mu(A_\infty) = 0$.

In the forthcoming discussion we identify vectors (and vector fields) $\xi \in \mathbb{R}^N$ with their dual element $v \mapsto v \cdot \xi$ where necessary.

Lemma 6.3. *Let $p \geq 1$, $N \geq 0$, $\varphi = (\varphi_1, \dots, \varphi_N) \in N^{1,p}(X)^N$, and Φ be the canonical minimal gradient associated to φ. If $\boldsymbol{g} = (g_1, \dots, g_N) \in L^\infty(X; (\mathbb{R}^N)^*)$, then*

$$\left|\sum_{k=1}^N g_k \mathrm{d}_G \varphi_k\right|_{G,x} = \Phi^x(\boldsymbol{g}) \quad \text{for } \mu\text{-a.e. } x \in X.$$

In particular, φ is p-independent on $U \subset X$ if and only if $\mathrm{d}_G\varphi_1, \dots, \mathrm{d}_G\varphi_N \in L^p(T^*X)$ are linearly independent on U.

Proof. If $g_1, \dots, g_N$ are simple functions, then $\boldsymbol{g} = \sum_j^M \chi_{A_j} \xi_j$ for disjoint Borel A_j and some $\xi_j \in (\mathbb{R}^N)^*$. It follows that $\sum_{k=1}^N g_k \, \mathrm{d}_G \varphi_k = \sum_j^M \chi_{A_j} \, \mathrm{d}_G(\xi_j \circ \varphi)$ as elements of $L^p(T^*X)$. Thus

$$\left|\sum_{k=1}^N g_k \, \mathrm{d}_G \varphi_k\right|_x = \left|\sum_j^M \chi_{A_j} \, \mathrm{d}_G(\xi_j \circ \varphi)\right|_x = \sum_j^M \chi_{A_j} |D(\xi_j \circ \varphi)|_p = \Phi^x(\boldsymbol{g})$$

for μ-a.e. $x \in X$.

The estimate

$$\Phi^x(\boldsymbol{g}) \le \left(\sum_k^N |g_k|^q\right)^{1/q} \left(\sum_k^N |D\varphi_k|_p^p\right)^{1/p} \le C|\boldsymbol{g}| \sum_k^N |D\varphi_k|_p,$$

valid for all simple vector-valued $\boldsymbol{g}$, implies that the equality in the claim is stable under local L^∞-convergence of $\boldsymbol{g}$. Since simple functions are dense in L^∞, the claim follows. The remaining claim follows in a straightforward way from the equality. □

Remark 6.4. If $\varphi \in \mathrm{LIP}(X; \mathbb{R}^N)$ is a chart in U, and $f \in N^{1,p}(X)$, then for the canonical minimal upper gradient $\Phi^x(a, \boldsymbol{\xi}))$ of $(f, \varphi) \in N^{1,p}_{\mathrm{loc}}(X; \mathbb{R}^{N+1})$ we have by Lemma 4.3(2) that $\Phi^x(1, -\mathrm{d}f) = 0$. Thus, by the previous lemma, we get $\mathrm{d}_G f - \sum_{k=1}^N g^k \, \mathrm{d}_G \varphi_k = 0$, where g^k are the components of $\mathrm{d}f$ and φ_k are the components of φ. Indeed, this follows by considering this first on the sets $U_M = \{x \in U : |g^k(x)| \le M,\ k = 1, \dots, N\}$ and sending $M \to \infty$ combined with locality.

Lemma 6.5. *If (U, φ) is an N-dimensional p-weak chart in X, then the differentials of the component functions $\mathrm{d}_G\varphi_1, \dots, \mathrm{d}_G\varphi_N$ form a basis of $L^p(T^*X)$ in U.*

Proof. By Lemma 6.3, $\mathrm{d}_G\varphi_1, \dots, \mathrm{d}_G\varphi_N \in L^p(T^*X)$ are linearly independent on U. To see that they span $L^p(T^*X)$ in U, let $f \in N^{1,p}(X)$, and set $g_k := \mathrm{d}f(e_k)$ for each $k = 1, \dots, N$, where e_k is the standard basis of $\mathbb{R}^N$. Then, since $\mathrm{d}\varphi_k = e^k$, where e^k is the dual basis of $(\mathbb{R}^N)^*$, we get $\mathrm{d}f = \sum_{k=1}^N g_k \, \mathrm{d}\varphi_k$. Thus, by Remark 6.4 we have $\mathrm{d}_G f = \sum_{k=1}^N g_k \, \mathrm{d}_G \varphi_k$. Since the abstract differentials $\mathrm{d}_G f$ span $L^p(T^*X)$, this completes the proof. □

Lemma 6.6. *Suppose $p \ge 1$ and X admits a p-weak differentiable structure. There exists an isometric isomorphism $\iota : \Gamma_p(T_p^*X) \to L^p(T^*X)$ of normed modules satisfying*

$$\iota(\mathrm{d}f) = \mathrm{d}_G f, \quad f \in N^{1,p}(X). \tag{6-2}$$

The map ι is uniquely determined by (6-2).

Uniqueness here means that if $A : \Gamma_p(T_p^*X) \to L^p(T^*X)$ is L^∞-linear and satisfies (6-2) then $A = \iota$.

Proof. The set

$$\mathcal{W} = \left\{\sum_j^M \chi_{A_j} \mathrm{d}f_j : (A_j)_j \text{ Borel partition of } X,\ f_j \in N^{1,p}(X)\right\}$$

is dense in $\Gamma_p(T_p^*X)$, since it contains all the simple Borel sections of T_p^*X. We set

$$\iota(v) := \sum_j^M \chi_{A_j} \mathrm{d}_G f_j, \quad v = \sum_j^M \chi_{A_j} \mathrm{d}f_j \in \mathcal{W}.$$

We have that

$$|\iota(v)|_G = \sum_j^M \chi_{A_j} |Df_j|_p = \sum_j^M \chi_{A_j} |\mathrm{d}f_j| = |v| \quad \mu\text{-a.e.}$$

for $v \in \mathcal{W}$. This implies that ι is well-defined and preserves the pointwise norm on the dense set $\mathcal{W}$. By Remark 6.4 we have that ι is linear. Since $\iota(\mathcal{W}) = \mathcal{V}$, it follows that ι extends to an isometric isomorphism $\iota : \Gamma_p(T_p^*X) \to L^p(T^*X)$. Note that $\iota(\mathrm{d}f) = \mathrm{d}_G f$ for every $f \in N^{1,p}(X)$, establishing (6-2).

To prove uniqueness, note that if $A : \Gamma_p(T_p^*X) \to L_p(T^*X)$ is linear and satisfies (6-2), then $A(v) = \iota(v)$ for all $v \in \mathcal{W}$ which implies that $A = \iota$ by the density of $\mathcal{W}$. □

Proof of Theorem 1.11. If X admits a p-weak differentiable structure, Lemma 6.5 implies that $L^p(T^*X)$ is locally finitely generated. To prove the converse implication, suppose $\{A_N\}_{N\in\mathbb{N}\cup\{\infty\}}$ is the dimensional decomposition of X and $\mu(A_\infty) = 0$.

Let $N \in \mathbb{N}$ be such that $\mu(A_N) \geq \mu(V) > 0$ for some Borel set V, and $v_1, \dots, v_N \in L^p(T^*X)$ is a basis of $L^p(T^*X)$ on V. By possibly passing to a smaller subset of V, we may assume that there exists $C > 0$ for which

$$\int_V \left|\sum_k^N g_k v_k\right|_G^p \mathrm{d}\mu \geq \frac{1}{C}\int_V |\boldsymbol{g}|^p \,\mathrm{d}\mu \quad \text{for all } \boldsymbol{g} = (g_1, \dots, g_N) \in L^\infty. \tag{6-3}$$

For each $k = 1, \dots, N$ there are sequences

$$v_k^n = \sum_j^{M_k^n} \chi_{A_{j,k}^n} \mathrm{d}_G f_{j,k}^n,$$

with $\{A_{j,k}^n\}_j$ a Borel partition of X and $(f_j^n) \subset N^{1,p}(X)$ such that $v_k^n \to v_k$ in $L^p(T^*X)$ as $n \to \infty$, by the definition of $L^p(T^*X)$. We set $J^n = \{1, \dots, M_1^n\} \times \cdots \times \{1, \dots, M_N^n\}$ and define new partitions $A_{\bar{j}}^n := A_{j_1,1}^n \cap \cdots \cap A_{j_N,N}^n$ indexed by $\bar{j} = (j_1, \dots, j_N) \in J^n$. Then

$$\begin{gathered} v_k^n = \sum_{\bar{j}\in J^n} \chi_{A_{\bar{j}}^n} \mathrm{d}_G(f_{j_k,k}^n), \quad \mu(V) = \sum_{\bar{j}\in J^n} \mu(A_{\bar{j}}^n \cap V), \\ \lim_{n\to\infty} \int_X |v_k^n - v_k|_G^p \,\mathrm{d}\mu = \lim_{n\to\infty} \sum_{\bar{j}\in J^n} \int_{A_{\bar{j}}^n} |\mathrm{d}_G \varphi_k^{n,\bar{j}} - v_k|_G^p \,\mathrm{d}\mu = 0 \end{gathered} \tag{6-4}$$

for all n and $k = 1, \dots, N$. We claim that there exists n so that $\varphi^{n,\bar{j}} := (f_{j_1,1}^n, \dots, f_{j_N,N}^n) \in N^{1,p}(X; \mathbb{R}^N)$ is p-independent on a positive measure subset of $A_{\bar{j}}^n \cap V$ for some $\bar{j} \in J^n$.

By (6-3) we have the inequality

$$\begin{aligned} \frac{1}{C}\int_{A_{\bar{j}}^n\cap V} |\boldsymbol{g}|^p \,\mathrm{d}\mu &\leq \int_{A_{\bar{j}}^n\cap V} \left|\sum_k^N g_k v_k\right|^p \mathrm{d}\mu \\ &\leq C' \int_{A_{\bar{j}}^n\cap V} \left|\sum_k^N g_k \mathrm{d}_G \varphi_k^{n,\bar{j}}\right|_G^p \mathrm{d}\mu + C' \int_{A_{\bar{j}}^n\cap V} \left|\sum_k^N g_k (\mathrm{d}_G \varphi_k^{n,\bar{j}} - v_k)\right|_G^p \mathrm{d}\mu \\ &\leq C' \int_{A_{\bar{j}}^n\cap V} \Phi_{n,\bar{j}}(\boldsymbol{g}(x), x) \,\mathrm{d}\mu + C'' \int_{A_{\bar{j}}^n\cap V} |\boldsymbol{g}|^p \left(\sum_k^N |\mathrm{d}_G \varphi_k^{n,\bar{j}} - v_k|_G^p\right) \mathrm{d}\mu \end{aligned}$$

for all $\boldsymbol{g} = (g_1, \dots, g_N) \in L^\infty$, where $\Phi_{n,\bar{j}}$ is the canonical minimal gradient of $\varphi^{n,\bar{j}}$ (see Lemma 6.3). By (6-4) there exists $n \in \mathbb{N}$, $\bar{j} \in J^n$ and a Borel set $U \subset A_{\bar{j}}^n \cap V$ with $0 < \mu(U) \leq \mu(A_{\bar{j}}^n \cap V)$ such that

$\sum_k^N |\mathrm{d}_G \varphi_k^{n,\bar{J}} - v_k|_G^p < \varepsilon$ on U, where $C''\varepsilon < 1/(2C)$. Thus

$$\frac{1}{C}\int_U |\boldsymbol{g}|^p \,\mathrm{d}\mu \le C' \int_U \Phi_{n,\bar{J}}(\boldsymbol{g}(x), x)\,\mathrm{d}\mu + \frac{1}{2C}\int_U |\boldsymbol{g}|^p \,\mathrm{d}\mu$$

for all $\boldsymbol{g} = (g_1, \dots, g_N) \in L^\infty(U; \mathbb{R}^N)$ by extending g by zero to $V \setminus U$. This readily implies that $I(\varphi^{n,\bar{J}}) > 0$ a.e. in U, proving the p-independence of $\varphi^{n,\bar{J}}$ in U. Note that $\varphi^{n,\bar{J}}$ is also maximal, since the existence of a Lipschitz map on a positive measure subset of U with a higher-dimensional target would imply that the local dimension of $L^p(T^*X)$ in V would be $> N$; see Lemma 6.3. By Proposition 4.14, U contains an N-dimensional p-weak chart, and [Keith 2004b, Proposition 3.1] implies that X admits a differentiable structure.

The argument above shows that each A_N with $\mu(A_N) > 0$ can be covered up to a null-set by N-dimensional p-weak charts, proving (b), while (a) follows directly from Lemma 6.6. Finally, (c) is implied by Proposition 4.13. □

Theorem 1.11 and [Gigli 2018, Chapter 2] immediately yield the following corollary.

Corollary 6.7. *Let $p \ge 1$ and suppose X admits a p-weak differentiable structure.*

(i) *If $p > 1$, then $N^{1,p}(X)$ is reflexive.*

(ii) *If $p = 2$, then $N^{1,2}(X)$ is infinitesimally Hilbertian if and only if, for μ-a.e. $x \in X$, the pointwise norm $|\cdot|_x$ (see Theorem 1.7) is induced by an inner product.* □

6B. *Lipschitz differentiability spaces.* A space X is said to be a Lipschitz differentiability space if it admits a Cheeger structure. Recall that a Cheeger structure is a countable collection of Cheeger charts (U_i, φ_i), see Section 4G, so that $\mu\big(X \setminus \bigcup_i U_i\big) = 0$. Following [Cheeger 1999, Section 4, p. 458], we note that the differentials $\mathrm{d}_{C,i} f$ of a Lipschitz function f with respect to overlapping charts satisfy a cocycle condition almost everywhere and the transition maps preserve the pointwise norm. Thus, they define a measurable L^∞-bundle T_C^*X called the measurable cotangent bundle.

Suppose now that X admits a Cheeger structure. Denote by T_C^*X the associated measurable cotangent bundle, and by

$$|\xi|_{C,x} := \mathrm{Lip}(\xi \circ \varphi)(x), \quad \xi \in (\mathbb{R}^N)^*,$$

the pointwise norm for μ-a.e. $x \in U$, where (U, φ) is an N-dimensional Cheeger chart of X.

Fix $p \ge 1$. Any Lipschitz differentiability space X admits a p-weak differentiable structure. Indeed, the asymptotic doubling property of the measure (see [Bate and Speight 2013]) implies, by [Bate 2015, Lemma 8.3], that X decomposes into finite-dimensional pieces. The existence of the p-weak differentiable structure now follows from Proposition 5.4, and the associated measurable cotangent bundle is denoted by T_p^*X. We have the following result from [Ikonen et al. 2022, Theorem 3.4]:

Theorem 6.8. *Let $p \ge 1$. There exists a morphism $P : \Gamma_p(T_C^*X) \to L^p(T^*X)$ of normed modules such that*

(a) *$P(\mathrm{d}_C f) = \mathrm{d}_G f$ for every $f \in \mathrm{LIP}(X)$,*

(b) *$|P(\omega)|_G \le |\omega|_C$ for every $\omega \in \Gamma_p(T_C^*X)$, and*

(c) *for every $w \in L^p(T^*X)$ there exists $\omega \in P^{-1}(w)$ with $|w|_G = |\omega|_c$.*

Remark 6.9. The proof of [Ikonen et al. 2022, Theorem 3.4] can be modified to cover the case $p = 1$: the energy density of Lipschitz functions holds for $p = 1$ by [Eriksson-Bique 2023], and equicontinuity can be used instead of L^p-boundedness to obtain the weakly convergent subsequence in the proof.

Proof of Theorem 1.10. Arguing as in the proof of Proposition 5.8 we may assume that X has a Borel partition $\{U_i\}$ and Lipschitz maps $\varphi^i_p = (\varphi^i_{p,1}, \dots, \varphi^i_{p,N_i})$, $\varphi^i_C = (\varphi^i_{C,1}, \dots, \varphi^i_{C,M_i})$ such that (U_i, φ^i_p) is a p-weak chart and (U_i, φ^i_C) is a Cheeger chart on X (of possibly different dimensions N_i and M_i) for each $i \in \mathbb{N}$. For each i and μ-a.e. $x \in U_i$ define

$$\sigma_{i,x} = (\mathrm{d}_{p,x}\varphi^i_{C,1}, \dots, \mathrm{d}_{p,x}\varphi^i_{C,M_i}) : \mathbb{R}^{N_i} \to \mathbb{R}^{M_i}.$$

It is easy to see that the collection $\{\pi_{i,x} = \sigma^*_{i,x}\}$ defines a bundle map $T^*_C X \to T^*_p X$ satisfying

$$\mathrm{d}_{p,x} f = \mathrm{d}_{C,x} f \circ \sigma_x \quad \text{for } \mu\text{-a.e. } x \in X,$$

for every $f \in N^{1,p}(X)$. This proves (1-7). In particular, for each $i \in \mathbb{N}$ and $\xi \in (\mathbb{R}^{M_i})^*$ we have $\pi_{i,x}(\xi) = \xi \circ \sigma_{i,x} = \mathrm{d}_{p,x}(\xi \circ \varphi^i_C)$, and consequently

$$|\pi_{i,x}(\xi)|_x = |D(\xi \circ \varphi^i_C)|_p(x) \le \mathrm{Lip}(\xi \circ \varphi^i_C)(x) = |\xi|_{C,x}$$

for μ-a.e. $x \in U_i$. Moreover, for any $\zeta \in (\mathbb{R}^{N_i})^*$, setting $\xi := \mathrm{d}_{C,x}(\zeta \circ \varphi^i_p)$, we have

$$\pi_{i,x}(\xi) = \mathrm{d}_{C,x}(\zeta \circ \varphi^i_p) \circ \sigma_x = \mathrm{d}_{p,x}(\zeta \circ \varphi^i_p) = \zeta,$$

proving that $\pi_{i,x}$ is surjective for μ-a.e. $x \in U_i$.

To prove that $\pi_{i,x}$ is a submetry for μ-a.e. $x \in U_i$, suppose to the contrary that there exists a Borel set $B \subset U_i$, with $0 < \mu(B) < \infty$, such that $\pi_{i,x}$ is not a submetry for $x \in B$. Then there exists a Borel map $\boldsymbol{\zeta} : B \to (\mathbb{R}^{N_i})^*$ with $|\boldsymbol{\zeta}_x|_x = 1$ and

$$|\boldsymbol{\zeta}_x|_x = 1 \quad \text{and} \quad \inf_{\xi \in \pi^{-1}_{i,x}(\boldsymbol{\zeta}_x)} |\xi|_{C,x} > 1 \quad \text{for } \mu\text{-a.e. } x \in B. \tag{6-5}$$

We derive a contradiction using Theorem 6.8 and the isometric isomorphism $\iota : \Gamma_p(T^*_p X) \to L^p(T^*X)$ from Theorem 1.11(a). We may view $\boldsymbol{\zeta}$ as an element of $\Gamma_p(T^*_p X)$ by extending it by zero outside B. Set $w := \iota(\boldsymbol{\zeta}) \in L^p(T^*X)$. Then $|w|_G = \chi_B$. By Theorem 6.8(c) there exists $\omega \in \Gamma_p(T^*_C X)$ with $P(\omega) = w$ and $|\omega|_C = |w|_G = \chi_B$ μ-a.e. However, since $\omega_x \in \pi^{-1}_{i,x}(\boldsymbol{\zeta}_x)$ for μ-a.e. $x \in B$, we have $|\omega|_{C,x} \ge \inf_{\xi \in \pi^{-1}_{i,x}(\boldsymbol{\zeta}_x)} |\xi|_{C,x} > 1$ for μ-a.e. $x \in B$ by (6-5), which is a contradiction. This completes the proof that $\pi_{i,x}$ is a submetry for μ-a.e. $x \in U_i$.

If $\mathrm{Lip}\, f \le \omega(|Df|_p)$ holds for every $f \in \mathrm{LIP}_b(X)$, then by [Ikonen et al. 2022, Theorem 1.1] we have $|Df|_p = \mathrm{Lip}\, f$ μ-a.e. for every $f \in \mathrm{LIP}_b(X)$. It follows that p-weak charts are Cheeger charts (see Theorem 1.8 and Remark 4.15) and that the pointwise norms agree μ-almost everywhere. This implies that the maps $\pi_{i,x}$ are isometric bijections for μ-a.e. x. □

Appendix: General measure theory

A1. ***Measurability questions.*** Here we record a host of measurability statements that are needed throughout the paper. See [Gigli and Pasqualetto 2020; Ambrosio et al. 2008; Bogachev 2007] for more details.

Given $f \in N^{1,p}(X)$ and a Borel representative g of p-weak upper gradient of f, we define

$$\Gamma(f) := \{\gamma \in \mathrm{AC}(I; X) : f \circ \gamma \in \mathrm{AC}(I; \mathbb{R})\},$$
$$\Gamma(f, g) := \{\gamma \in \mathrm{AC}(I; X) : g \text{ upper gradient of } f \text{ along } \gamma\} \subset \Gamma(f)$$

and

$$MD = \{(\gamma, t) \in \mathrm{AC}(I; X) \times I : |\gamma_t'| \text{ exists}\},$$
$$\mathrm{Diff}(f) = \{(\gamma, t) \in \mathrm{AC}(I; X) \times I : \gamma \in \Gamma(f),\ (f \circ \gamma)_t' \text{ and } |\gamma_t'| > 0 \text{ exist}\},$$
$$\mathrm{Diff}(f, g) = \{(\gamma, t) \in \mathrm{Diff}(f) : \gamma \in \Gamma(f, g),\ |(f \circ \gamma)_t'| \le g_f(\gamma_t)|\gamma_t'|\}.$$

Also, let $\mathrm{Len}(\gamma)$ be the length of a curve γ, if the curve is rectifiable, and otherwise infinity. The function der is defined by $\mathrm{der}(\gamma, t) := |\gamma_t'| = \lim_{h\to 0} d(\gamma_{t+h}, \gamma_t)/|h|$, when the limit exists, and otherwise is infinity.

Lemma A.1. (1) *The functions* $\mathrm{Len} : C(I; X) \to [0, \infty]$ *and* $\mathrm{der} : \mathrm{AC}(I; X) \times I \to [0, \infty]$ *are Borel measurable.*

(2) *If* $g : X \to [0, \infty]$ *is a Borel function, then* $I : \mathrm{AC}(I; X) \to \mathbb{R}$, *given by* $\gamma \mapsto \int_\gamma g \,\mathrm{d}s$ *or* ∞ *if the curve is not rectifiable, is Borel.*

(3) *If* $H : \mathrm{AC}(I; X) \times I \to [0, \infty]$ *is Borel, then* $I_H(\gamma) := \int_0^1 H(\gamma, s)\,\mathrm{d}s : \mathrm{AC}(I; X) \to [0, \infty]$ *is Borel.*

(4) *The set* MD *is Borel, and the map* $MD \to \mathbb{R}$ *defined by* $(\gamma, t) \to |\gamma_t'|$ *is Borel.*

Proof. (1) The length function is a lower semicontinuous function with respect to uniform convergence, and thus is Borel. Fix $r, p \in \mathbb{Q}$ positive. Then define

$$A_{p,r} = \bigcup_{n\in\mathbb{N}} \bigcap_{q\in\mathbb{Q}\cap(-1/n,1/n)} \{(\gamma, t) : |d(\gamma_{t+q}, \gamma_t) - qp| < r|q|\},$$

which is Borel. The set M where the metric derivative exists is of the form $\bigcap_{r\in\mathbb{Q}\cap(0,\infty)} \bigcup_{p\in\mathbb{Q}\cap(0,\infty)} A_{p,r}$. On this set we have $M \cap A_{p,r} = \mathrm{der}^{-1}(B(p, r))$ and thus $\mathrm{der}(\gamma, t)$ is Borel.

(2) The claims for the integral function being Borel follow from a monotone family argument, and considering g first a characteristic function of an open set and using lower semicontinuity of the integral in that case.

(3) If H is a characteristic function of a product set $A \times B$, where A and B are open sets such that $A \subset C(I; X)$, $B \subset I$, then the claim follows just as in statement (2). Again, by a monotone family argument, we obtain the claim for all Borel measurable functions.

(4) Define for every $q \in \mathbb{Q}$ and $\varepsilon, h > 0$ the sets $A(\varepsilon, q, h)$ and $B(\varepsilon, q)$ by

$$A(\varepsilon, q, h) := \left\{(\gamma, t) \in C(I; X) \times I : \left|\frac{d(\gamma_{t+h}, \gamma_t)}{|h|} - q\right| < \varepsilon\right\},$$
$$B(\varepsilon, q) := \bigcup_{\delta\in\mathbb{Q}_+} \bigcap_{h\in(0,\delta)\cap\mathbb{Q}} A(\varepsilon, q, h).$$

We note that $|\gamma_t'|$ exists if and only if $(\gamma, t) \in \bigcap_{j\in\mathbb{N}} \bigcup_{q\in\mathbb{Q}} B(2^{-j}, q) = MD$. On the set MD, where the limit exists, we can write $|\gamma_t'| = \lim_{n\to\infty} n(d(\gamma_{t+n^{-1}}, \gamma_t))$, which shows measurability. □

Lemma A.2. *Let g be a Borel p-weak upper gradient of $f \in N^{1,p}(X)$. There exists a Borel set $\Gamma_0 \subset \mathrm{AC}(I; X)$ with $\mathrm{Mod}_p(\Gamma_0) = 0$ such that $AC \setminus \Gamma_0 \subset \Gamma(f, g)$.*

Suppose moreover that f is Borel. Then the set $A := \Gamma_0^c \times I \cap \mathrm{Diff}(f, g)$ is Borel, and $\boldsymbol{\pi}(A^c) = 0$ whenever $\boldsymbol{\pi} = \mathcal{L}^1 \times \boldsymbol{\eta}$ and $\boldsymbol{\eta}$ is a q-test plan.

If f is Lipschitz, and $g = \mathrm{Lip}[f]$, then we can choose $\Gamma_0 = \varnothing$, and $\mathrm{Diff}(f, g) = \mathrm{Diff}(f)$ is Borel.

Note that we make no claims about the Borel measurability of the set $\Gamma(f, g)$.

Proof. We model the argument after [Pasqualetto 2022, Lemma 1.9]. Since $\mathrm{Mod}_p(\Gamma(f, g)^c) = 0$, there exists an L^p-integrable Borel function $\rho : X \to [0, \infty]$ with $\int_\gamma \rho \, \mathrm{d}s = \infty$ for every $\gamma \notin \Gamma_{f,g}$. Then $\Gamma_0 := \left\{\gamma \in \mathrm{AC}(I; X) : \int_\gamma \rho \, \mathrm{d}s = \infty\right\} \supset \Gamma_{f,g}^c$ is a Borel set, by Lemma A.1 and $\boldsymbol{\eta}(\Gamma_0) = 0$ for every q-plan $\boldsymbol{\eta}$ (see Remark 2.2). If f is Lipschitz, then $\Gamma(f, g) = \mathrm{AC}(I; X)$. Thus, we can choose $\Gamma_0 = \varnothing$.

For the second part assume $f \in N^{1,p}(X)$ is Borel, and set

$$A(\varepsilon, q, h) = \left\{(\gamma, t) \in \Gamma_0^c \times I : \left|\frac{f(\gamma_{t+h}) - f(\gamma_t)}{h} - q\right| < \varepsilon\right\},$$
$$B(\varepsilon, q) = \bigcup_{\delta \in \mathbb{Q}_+} \bigcap_{h \in (0,\delta) \cap \mathbb{Q}} A(\varepsilon, q, h)$$

for each $q \in \mathbb{Q}$ and $\varepsilon, h > 0$. It is easy to see that for each $\gamma \notin \Gamma_0$, $(f \circ \gamma)'_t$ exists if and only if

$$(\gamma, t) \in \bigcap_{j \in \mathbb{N}} \bigcup_{q \in \mathbb{Q}} B(2^{-j}, q) =: A.$$

Note that A is a Borel set with $A \cap MD \subset \mathrm{Diff}(f)$. Moreover, $(\gamma, t) \mapsto (f \circ \gamma)'_t$ is Borel when restricted to $A \cap MD$.

Define the Borel function $H(\gamma, t) = (f \circ \gamma)'_t$ if $(\gamma, t) \in A \cap MD$ and $H = +\infty$ otherwise, and $G(\gamma, t) = |H| - g(\gamma_t)|\gamma'_t|$ (here we use the convention $\infty - \infty = \infty$). Then the set

$$\{G \le 0\} = \Gamma_0^c \times I \cap \mathrm{Diff}(f, g)$$

is Borel.

Set $N := \{G > 0\}$, suppose $\boldsymbol{\eta}$ is a q-test plan and $\boldsymbol{\pi} := \mathcal{L}^1 \times \boldsymbol{\eta}$. Note that

$$N \subset \Gamma_0 \times I \cup \{(\gamma, t) \in \Gamma_0^c \times I : G(\gamma, t) > 0\}.$$

But, for all $\gamma \notin \Gamma_0$, we have $G(\gamma, t) \le 0$ for $\mathcal{L}^1$-a.e. $t \in I$. Thus

$$\boldsymbol{\pi}(N) \le \boldsymbol{\eta}(\Gamma_0) + \int_{\Gamma_0^c} \int_0^1 \chi_{\{G(\gamma,\cdot)>0\}}(t) \, \mathrm{d}t \, \mathrm{d}\boldsymbol{\eta}(\gamma) = 0,$$

finishing the proof of the second part. □

Corollary A.3. *Every pointwise defined function $f \in N^{1,p}(X)$ has a Borel representative $\bar{f} \in N^{1,p}(X)$. Moreover, if $f \in N^{1,p}(X)$ and g is a Borel p-weak upper gradient of f, there exists a Borel set $N \subset C(I; X) \times I$, with $N^c \subset \mathrm{Diff}(f, g)$ and $\boldsymbol{\pi}(N) = 0$ whenever $\boldsymbol{\pi} = \mathcal{L}^1 \times \boldsymbol{\eta}$, $\boldsymbol{\eta}$ a q-test plan. The map $(\gamma, t) \mapsto (f \circ \gamma)'_t$ if $(\gamma, t) \notin N$ and $+\infty$ otherwise is Borel. If f is Lipschitz the representative can be chosen as the same function.*

Proof. The first claim follows directly from [Eriksson-Bique 2023, Theorem 1.1]. To see the second, let $\bar{f} \in N^{1,p}(X)$ be a Borel representative of f. The set $E := \{f \neq \bar{f}\}$ is p-exceptional, i.e., $\Gamma_E := \{\gamma : \gamma^{-1}(E) \neq \varnothing\}$ has zero p-modulus. Note that, if f is Lipschitz, then f is automatically Borel and we do not need to change representatives, and we can set $\Gamma_E = \varnothing$.

If $\bar{A}$ is the set in Lemma A.2 for $\bar{f}, g$, then $A := \bar{A} \setminus (\Gamma_E \times I) \subset \mathrm{Diff}(f, g)$ and $N := A^c$ satisfies the claim since it is Borel and $N \subset \Gamma_E \times I \cup \bar{A}^c$.

The last claim follows since N^c is Borel and, if $(\gamma, t) \notin N$, we have

$$(f \circ \gamma)'_t = \lim_{n\to\infty} n(f(\gamma_{t+1/n}) - f(\gamma_t)). \qquad \square$$

A2. *Essential supremum.*

Definition A.4. Let X be a σ-finite measure space and $\mathcal{F}$ a collection of measurable functions on X, then there exists a function $g : X \to \mathbb{R} \cup \{\infty, -\infty\}$ which is measurable, and:

(A) For each $f \in \mathcal{F}$,

$$f \leq g$$

almost everywhere.

(B) For each g' that satisfies (A), will satisfy $g \leq g'$ almost everywhere.

We call $g = \operatorname{ess\,sup}_{f\in\mathcal{F}} f$. Similarly, we define $g = \operatorname{ess\,inf}_{f\in\mathcal{F}} f$, by switching the directions of the inequalities and assuming $g : X \to \mathbb{R} \cup \{\infty, -\infty\}$.

We will need the following standard lemma. While its proof is standard, we provide it for the sake of completeness.

Lemma A.5. *If X is any σ-finite measure space and $\mathcal{F}$ is any collection of measurable functions, then* $\operatorname{ess\,sup}_{f\in\mathcal{F}} f$ *and* $\operatorname{ess\,inf}_{f\in\mathcal{F}} f$ *exists and is unique, and further, there are sequences $f_n, g_n \in \mathcal{F}$ so that* $\operatorname{ess\,sup}_{f\in\mathcal{F}} f = \sup_n f_n$ *and* $\operatorname{ess\,inf}_{f\in\mathcal{F}} f = \inf_n g_n$ *almost everywhere.*

Proof. The uniqueness follows from (B) in Definition A.4. Indeed, if g and g' are essential suprema, they both satisfy A, and thus $g \leq g'$ and $g' \leq g$.

By considering $\{\arctan(f) : f \in \mathcal{F}\}$, we can assume that the collection is bounded. Further, by σ-finiteness, and after exhausting the space by finite measure sets, it suffices to consider a bounded measure. Define $\mathcal{G}$ to be the collection of all functions of the form $\max(f_1, \dots, f_k)$ for some $f_i \in \mathcal{F}$. By construction, if $g, g' \in G$, then $\max(g, g') \in G'$.

Consider $U = \sup_{g\in\mathcal{G}} \int g \, \mathrm{d}\mu$. There is a sequence g_n so that $\lim_{n\to\infty} \int g_n \, \mathrm{d}\mu = U$. By modifying the sequence if necessary, we may take it increasing in n, and define $g = \lim_{n\to\infty} g_n$.

We claim that g is an essential supremum for $\mathcal{F}$. First, if $f \in \mathcal{F}$, and $f > g$ on a positive measure set, then $\lim_{n\to\infty} \int \max(f, g_n) \, \mathrm{d}\mu > U$, contradicting the definition of U. Thus the condition A in the definition is satisfied.

Now, if h is any other function satisfying A, then $h \geq g_n$, and thus $h \geq g$ almost everywhere, by construction. Thus B is also satisfied. Finally, the construction gives a countable collection g_n formed each from finitely many $f_i \in \mathcal{F}$, and thus gives the final claim in the statement. $\square$

Acknowledgements

Eriksson-Bique was partially supported by National Science Foundation under grant no. DMS-1704215 and by the Finnish Academy under research postdoctoral grant no. 330048. Soultanis was supported by the Swiss National Science Foundation Grant 182423. Throughout the project the authors have had insightful discussions with Nageswari Shanmugalingam, which have been tremendously useful. A further thanks goes to Jeff Cheeger and Nicola Gigli for helpful comments and inspiration for the project. The authors thank IMPAN for hosting the semester "Geometry and analysis in function and mapping theory on Euclidean and metric measure space" where this research was started. Through this workshop the authors were partially supported by grant no. 346300 for IMPAN from the Simons Foundation and the matching 2015-2019 Polish MNiSW fund. Eriksson-Bique additionally thanks University of Jyväskylä, where much of this work was done.

References

[Alberti and Marchese 2016] G. Alberti and A. Marchese, "On the differentiability of Lipschitz functions with respect to measures in the Euclidean space", *Geom. Funct. Anal.* **26**:1 (2016), 1–66. MR Zbl

[Ambrosio et al. 2008] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, 2nd ed., Birkhäuser, Basel, 2008. MR Zbl

[Ambrosio et al. 2013] L. Ambrosio, N. Gigli, and G. Savaré, "Density of Lipschitz functions and equivalence of weak gradients in metric measure spaces", *Rev. Mat. Iberoam.* **29**:3 (2013), 969–996. MR Zbl

[Ambrosio et al. 2014] L. Ambrosio, N. Gigli, and G. Savaré, "Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below", *Invent. Math.* **195**:2 (2014), 289–391. MR Zbl

[Ambrosio et al. 2015a] L. Ambrosio, M. Colombo, and S. Di Marino, "Sobolev spaces in metric measure spaces: reflexivity and lower semicontinuity of slope", pp. 1–58 in *Variational methods for evolving objects* (Sapporo, Japan, 2012), edited by L. Ambrosio et al., Adv. Stud. Pure Math. **67**, Math. Soc. Japan, Tokyo, 2015. MR Zbl

[Ambrosio et al. 2015b] L. Ambrosio, S. Di Marino, and G. Savaré, "On the duality between p-modulus and probability measures", *J. Eur. Math. Soc.* **17**:8 (2015), 1817–1853. MR Zbl

[Ambrosio et al. 2015c] L. Ambrosio, A. Pinamonti, and G. Speight, "Tensorization of Cheeger energies, the space $H^{1,1}$ and the area formula for graphs", *Adv. Math.* **281** (2015), 1145–1177. MR Zbl

[Bate 2015] D. Bate, "Structure of measures in Lipschitz differentiability spaces", *J. Amer. Math. Soc.* **28**:2 (2015), 421–482. MR Zbl

[Bate and Speight 2013] D. Bate and G. Speight, "Differentiability, porosity and doubling in metric measure spaces", *Proc. Amer. Math. Soc.* **141**:3 (2013), 971–985. MR Zbl

[Björn and Björn 2011] A. Björn and J. Björn, *Nonlinear potential theory on metric spaces*, EMS Tracts in Math. **17**, Eur. Math. Soc., Zürich, 2011. MR Zbl

[Bogachev 2007] V. I. Bogachev, *Measure theory, II*, Springer, 2007. MR Zbl

[Cheeger 1999] J. Cheeger, "Differentiability of Lipschitz functions on metric measure spaces", *Geom. Funct. Anal.* **9**:3 (1999), 428–517. MR Zbl

[Cheeger and Kleiner 2009] J. Cheeger and B. Kleiner, "Differentiability of Lipschitz maps from metric measure spaces to Banach spaces with the Radon–Nikodým property", *Geom. Funct. Anal.* **19**:4 (2009), 1017–1028. MR Zbl

[Cheeger et al. 2016] J. Cheeger, B. Kleiner, and A. Schioppa, "Infinitesimal structure of differentiability spaces, and metric differentiation", *Anal. Geom. Metr. Spaces* **4**:1 (2016), 104–159. MR Zbl

[David and Eriksson-Bique 2020] G. C. David and S. Eriksson-Bique, "Infinitesimal splitting for spaces with thick curve families and Euclidean embeddings", preprint, 2020. arXiv 2006.10668

[De Philippis and Rindler 2016] G. De Philippis and F. Rindler, "On the structure of $\mathscr{A}$-free measures and applications", *Ann. of Math.* (2) **184**:3 (2016), 1017–1039. MR Zbl

[Di Marino and Speight 2015] S. Di Marino and G. Speight, "The p-weak gradient depends on p", *Proc. Amer. Math. Soc.* **143**:12 (2015), 5239–5252. MR Zbl

[Di Marino and Squassina 2019] S. Di Marino and M. Squassina, "New characterizations of Sobolev metric spaces", *J. Funct. Anal.* **276**:6 (2019), 1853–1874. MR Zbl

[Dunford and Schwartz 1958] N. Dunford and J. T. Schwartz, *Linear operators*, *I*: *General theory*, Pure Appl. Math. **7**, Interscience, New York, 1958. MR Zbl

[Durand-Cartagena et al. 2021] E. Durand-Cartagena, S. Eriksson-Bique, R. Korte, and N. Shanmugalingam, "Equivalence of two BV classes of functions in metric spaces, and existence of a Semmes family of curves under a 1-Poincaré inequality", *Adv. Calc. Var.* **14**:2 (2021), 231–245. MR Zbl

[Eriksson-Bique 2023] S. Eriksson-Bique, "Density of Lipschitz functions in energy", *Calc. Var. Partial Differential Equations* **62**:2 (2023), art. id. 60. MR Zbl

[Fuglede 1957] B. Fuglede, "Extremal length and functional completion", *Acta Math.* **98** (1957), 171–219. MR Zbl

[Gigli 2015] N. Gigli, *On the differential structure of metric measure spaces and applications*, Mem. Amer. Math. Soc. **1113**, Amer. Math. Soc., Providence, RI, 2015. MR Zbl

[Gigli 2018] N. Gigli, *Nonsmooth differential geometry: an approach tailored for spaces with Ricci curvature bounded from below*, Mem. Amer. Math. Soc. **1196**, Amer. Math. Soc., Providence, RI, 2018. MR Zbl

[Gigli and Pasqualetto 2020] N. Gigli and E. Pasqualetto, *Lectures on nonsmooth differential geometry*, SISSA Springer Series **2**, Springer, 2020. MR Zbl

[Hajłasz 1996] P. Hajłasz, "Sobolev spaces on an arbitrary metric space", *Potential Anal.* **5**:4 (1996), 403–415. MR Zbl

[Hajłasz 2003] P. Hajłasz, "Sobolev spaces on metric-measure spaces", pp. 173–218 in *Heat kernels and analysis on manifolds*, *graphs*, *and metric spaces* (Paris, 2002), edited by P. Auscher et al., Contemp. Math. **338**, Amer. Math. Soc., Providence, RI, 2003. MR Zbl

[Heinonen and Koskela 1998] J. Heinonen and P. Koskela, "Quasiconformal maps in metric spaces with controlled geometry", *Acta Math.* **181**:1 (1998), 1–61. MR Zbl

[Heinonen et al. 2015] J. Heinonen, P. Koskela, N. Shanmugalingam, and J. T. Tyson, *Sobolev spaces on metric measure spaces: an approach based on upper gradients*, New Math. Monogr. **27**, Cambridge Univ. Press, 2015. MR Zbl

[Honzlová Exnerová et al. 2021] V. Honzlová Exnerová, O. F. K. Kalenda, J. Malý, and O. Martio, "Plans on measures and AM-modulus", *J. Funct. Anal.* **281**:10 (2021), art. id. 109205. MR Zbl

[Ikonen et al. 2022] T. Ikonen, E. Pasqualetto, and E. Soultanis, "Abstract and concrete tangent modules on Lipschitz differentiability spaces", *Proc. Amer. Math. Soc.* **150**:1 (2022), 327–343. MR Zbl

[Keith 2003] S. Keith, "Modulus and the Poincaré inequality on metric measure spaces", *Math. Z.* **245**:2 (2003), 255–292. MR Zbl

[Keith 2004a] S. Keith, "A differentiable structure for metric measure spaces", *Adv. Math.* **183**:2 (2004), 271–315. MR Zbl

[Keith 2004b] S. Keith, "Measurable differentiable structures and the Poincaré inequality", *Indiana Univ. Math. J.* **53**:4 (2004), 1127–1150. MR Zbl

[Lučić et al. 2021] D. Lučić, E. Pasqualetto, and T. Rajala, "Characterisation of upper gradients on the weighted Euclidean space and applications", *Ann. Mat. Pura Appl.* (4) **200**:6 (2021), 2473–2513. MR Zbl

[Mackay et al. 2013] J. M. Mackay, J. T. Tyson, and K. Wildrick, "Modulus and Poincaré inequalities on non-self-similar Sierpiński carpets", *Geom. Funct. Anal.* **23**:3 (2013), 985–1034. MR Zbl

[Pasqualetto 2022] E. Pasqualetto, "Testing the Sobolev property with a single test plan", *Studia Math.* **264**:2 (2022), 149–179. MR Zbl

[Rudin 1980] W. Rudin, *Function theory in the unit ball of $\mathbb{C}^n$*, Grundlehren der Math. Wissenschaften **241**, Springer, 1980. MR Zbl

[Schioppa 2016a] A. Schioppa, "Derivations and Alberti representations", *Adv. Math.* **293** (2016), 436–528. MR Zbl

[Schioppa 2016b] A. Schioppa, "Metric currents and Alberti representations", *J. Funct. Anal.* **271**:11 (2016), 3007–3081. MR Zbl

[Semmes 1996] S. Semmes, "Finding curves on general spaces through quantitative topology, with applications to Sobolev and Poincaré inequalities", *Selecta Math.* (*N.S.*) **2**:2 (1996), 155–295. MR Zbl

[Shanmugalingam 2000] N. Shanmugalingam, "Newtonian spaces: an extension of Sobolev spaces to metric measure spaces", *Rev. Mat. Iberoam.* **16**:2 (2000), 243–279. MR Zbl

Received 2 Jun 2021. Revised 11 May 2022. Accepted 11 Jul 2022.

SYLVESTER ERIKSSON-BIQUE: sylvester.d.eriksson-bique@jyu.fi
Research Unit of Mathematical Sciences, University of Oulu, Oulu, Finland

ELEFTERIOS SOULTANIS: elefterios.e.soultanis@jyu.fi
Department of Mathematics and Statistics, University of Jyväskylä, Jyväskylä, Finland

ANALYSIS AND PDE
Vol. 17 (2024), No. 2, pp. 499–533
DOI: 10.2140/apde.2024.17.499

SMOOTH EXTENSIONS FOR INERTIAL MANIFOLDS OF SEMILINEAR PARABOLIC EQUATIONS

ANNA KOSTIANKO AND SERGEY ZELIK

The paper is devoted to a comprehensive study of smoothness of inertial manifolds (IMs) for abstract semilinear parabolic problems. It is well known that in general we cannot expect more than $C^{1,\varepsilon}$-regularity for such manifolds (for some positive, but small ε). Nevertheless, as shown in the paper, under natural assumptions, the obstacles to the existence of a C^n-smooth inertial manifold (where $n \in \mathbb{N}$ is any given number) can be removed by increasing the dimension and by modifying properly the nonlinearity outside of the global attractor (or even outside the $C^{1,\varepsilon}$-smooth IM of a minimal dimension). The proof is strongly based on the Whitney extension theorem.

1. Introduction

It is believed that in many cases the long-time behaviour of infinite-dimensional dissipative dynamical systems generated by evolutionary PDEs (at least in bounded domains) can be effectively described by finitely many parameters (the so-called order parameters in the terminology of I. Prigogine) which obey a system of ODEs. This system of ODEs (if it exists) is usually referred as an inertial form (IF) of the considered PDE; see [Hale 1988; Robinson 2001; 2011; Temam 1988; Zelik 2014] and references therein for more details. However, despite the fundamental significance of this reduction from both theoretical and applied points of view and big interest during the last 50 years, the nature of such a reduction and its rigorous justification remains a mystery.

Indeed, it is well understood now that the key question of the theory is how smooth the desired IF can/should be. For instance, in the case of *Hölder* continuous IFs, there is a highly developed machinery for constructing them based on the theory of global attractors and the Mañé projection theorem. We recall that, by definition, a global attractor is a compact invariant set in the phase space of the dissipative system

This work is partially supported by the RSF grant 19-71-30004 as well as the EPSRC grant EP/P024920/1. The authors also would like to thank Dmitry Turaev for many fruitful discussions.
MSC2020: 35B40, 35B42, 37D10, 37L25.
Keywords: inertial manifolds, finite-dimensional reduction, smoothness, Whitney extension theorem.

considered which attracts as time goes to infinity the images of bounded sets under the evolutionary semigroup related to the considered problem. Thus, on the one hand, a global attractor (if it exists) contains all of the nontrivial dynamics and, on the other hand, it is usually essentially "smaller" than the initial phase space and this second property allows us to speak about the reduction of degrees of freedom in the limit dynamics. In particular, one of the main results of the attractors theory tells us that, under relatively weak assumptions on a dissipative PDE (in a bounded domain), the global attractor exists and has finite Hausdorff and fractal dimensions. In turn, due to the Mañé projection theorem, this finite-dimensionality guarantees that this attractor can be projected one-to-one to a generic finite-dimensional plane of the phase space and that the inverse map is Hölder continuous. Finally, this scheme gives us an IF with Hölder continuous vector field defined on some compact set of $\mathbb{R}^N$ which is treated as a rigorous justification of the above-mentioned finite-dimensional reduction. This approach works, for instance, for 2-dimensional Navier–Stokes equations, reaction-diffusion systems, pattern formation equations, damped wave equations, etc.; see [Babin and Vishik 1992; Ben-Artzi et al. 1993; Chepyzhov and Vishik 2002; Hale 1988; Henry 1981; Hunt and Kaloshin 1999; Miranville and Zelik 2008; Robinson 2011; Sell and You 2002; Temam 1988].

However, the above-described scheme has a very essential intrinsic drawback which prevents us from treating it as a satisfactory solution of the finite-dimensional reduction problem. Namely, the vector field in the IF thus constructed is Hölder continuous *only* and there is no way in general to get even its Lipschitz continuity. As a result, we may lose the uniqueness of solutions for the obtained IF and have to use the initial infinite-dimensional system at least in order to select the correct solution of the reduced IF. Another drawback is that the Mañé projection theorem is not constructive, so it is not clear how to choose this "generic" plane for projection in applications; in addition, the IF constructed in such a way is defined only on a complicated compact set (the image of the attractor under the projection) and it is not clear how to extend it on the whole $\mathbb{R}^N$ preserving the dynamics (surprisingly, this is also a deep open problem; a partial solution of it is given in [Robinson 1999]).

It is also worth noting that the restriction for IF to be only Hölder continuous is far from being just a technical problem here. As relatively simple counterexamples show (see [Eden et al. 2013; Kostianko and Zelik 2018; Mallet-Paret et al. 1993; Romanov 2000; Zelik 2014]) the fractal dimension of the global attractor may be finite and not big, but the attractor cannot be embedded into any finite-dimensional Lipschitz (or even log-Lipschitz) finite-dimensional submanifold of the phase space. Even more importantly, the dynamics on this attractor does not look finite-dimensional at all (despite the existence of a *Hölder continuous* (with the Hölder exponent arbitrarily close to 1) IF provided by the Mañé projection theorem). For instance, it may contain limit cycles with superexponential rate of attraction, decaying travelling waves in Fourier space and other phenomena which are impossible in the classical dynamics generated by smooth ODEs. These examples suggest that, in contradiction to the widespread paradigm, Hölder continuous IF is probably not an appropriate tool for distinguishing between finite and infinite-dimensional limit behaviour and, as a result, fractal-dimension is not so good for estimating the number of degrees of freedom for the reduced dynamics; see [Eden et al. 2013; Kostianko and Zelik 2018; Zelik 2014] for more details.

An alternative, probably more transparent approach to the finite-dimensional reduction problem which has been suggested in [Foias et al. 1988] is related to the concept of an inertial manifold (IM). By definition, an IM is a finite-dimensional smooth (at least Lipschitz) invariant submanifold of the phase space which is globally exponentially stable and possesses the so-called exponential tracking property (that is, existence of asymptotic phase). Usually this manifold is $C^{1,\varepsilon}$-smooth for some positive ε and is normally hyperbolic, so the exponential tracking is an immediate corollary of normal hyperbolicity. Then the corresponding IF is just a restriction of the initial PDE to IM and is also $C^{1,\varepsilon}$-smooth. However, being a sort of centre manifold, an IM requires a separation of the dependent variable to the "slow" and "fast" components and this, in turn, leads to extra rather restrictive assumptions which are usually formulated in terms of *spectral gap* conditions. Namely, let us consider the following abstract semilinear parabolic equation in a real Hilbert space H:

$$\partial_t u + Au = F(u), \quad u|_{t=0} = u_0, \tag{1-1}$$

where $A : D(A) \to H$ is a self-adjoint positive operator such that A^{-1} is compact and $F : H \to H$ is a given nonlinearity which is globally Lipschitz in H with Lipschitz constant L. Let also $0 < \lambda_1 \le \lambda_2 \le \cdots$ be the eigenvalues of A enumerated in the nondecreasing order and $\{e_n\}_{n=1}^{\infty}$ be the corresponding eigenvectors. Then, the sufficient condition for the existence of an N-dimensional IM reads

$$\lambda_{N+1} - \lambda_N > 2L. \tag{1-2}$$

If this condition is satisfied, the desired IM $\mathcal{M}_N$ is actually a graph of a Lipschitz function $M_N : H_N \to (H_N)^{\perp}$, where $H_N = \operatorname{span}\{e_1, \dots, e_N\}$ is a spectral subspace spanned by the first N eigenvectors, and the corresponding IF has the form

$$\frac{d}{dt} u_N + Au_N = \boldsymbol{P}_N F(u_N + M_N(u_N)), \quad u_N \in H_N \sim \mathbb{R}^N, \tag{1-3}$$

where $\boldsymbol{P}_N$ is the orthoprojector to H_N; see [Chow et al. 1992; Constantin et al. 1989; Foias et al. 1988; Koksch 1998; Miklavčič 1991; Romanov 1993; Rosa and Temam 1996; Zelik 2014] and also Section 2 below.

We see that, in contrast to the IF constructed via the Mañé projection theorem, the IF which corresponds to the IM is explicit (uses the spectral projections) and is as smooth as the functions F and M_N are. We mention that although the spectral gap condition (1-2) is rather restrictive (e.g., in the case where A is a Laplacian in a bounded domain, it is satisfied in 1-dimensional case only) and is known to be sharp in the class of abstract semilinear parabolic equations (see [Eden et al. 2013; Miklavčič 1991; Romanov 1993; Zelik 2014] for more details), it can be relaxed for some concrete classes of PDEs. For instance, for scalar 3-dimensional reaction-diffusion equations (using the so-called spatial averaging principle, see [Mallet-Paret and Sell 1988]), for 1-dimensional reaction-diffusion-advection systems (using the proper integral transforms, see [Kostianko and Zelik 2017; 2018]), for 3-dimensional Cahn–Hilliard equations and various modifications of 3-dimensional Navier–Stokes equations (using various modifications of spatial-averaging, see [Gal and Guo 2018; Kostianko 2018; Kostianko and Zelik 2015; Li and Sun 2020]), for the 3-dimensional complex Ginzburg–Landau equation (using the so-called spatiotemporal

averaging, see [Kostianko 2020]), etc. Note also that the global Lipschitz continuity assumption for the nonlinearity F is not an essential extra restriction since usually one proves the well-posedness and dissipativity of the PDE under consideration *before* constructing the IM. Cutting off the nonlinearity outside the absorbing ball does not affect the limit dynamics, but reduces the case of locally Lipschitz continuous nonlinearity (satisfying the proper dissipativity restrictions) to the model case where the nonlinearity is globally Lipschitz continuous. Of course, this cut-off procedure is not unique and, as we will see below, choosing it correctly is extremely important in the theory of IMs.

The main aim of the present paper is to study the smoothness of the IFs for semilinear parabolic equations (1-1) in the ideal situation where the nonlinearity F is smooth and the spectral gap condition (1-2) is satisfied. As we have already mentioned, in this case we have a $C^{1,\varepsilon}$-smooth IM $\mathcal{M}_N$ for some $\varepsilon > 0$ and the associated IF (1-3) which is also $C^{1,\varepsilon}$-smooth; see [Zelik 2014]. But, unfortunately, the exponent $\varepsilon > 0$ here is usually very small (depending on the spectral gap) and in a more or less general situation, we cannot expect even the C^2-regularity of the IM. The spectral gap condition for C^2-regular IM is

$$\lambda_{N+1} - 2\lambda_N > 3L \tag{1-4}$$

and such exponentially big spectral gaps are not available if A is a finite-order elliptic operator in a bounded domain. The corresponding counterexamples were given in [Chow et al. 1992]; see also Example 3.11 below. Thus, the existing IM theory does not allow us, even in the ideal situation, to construct more regular than $C^{1,\varepsilon}$ IFs (where $\varepsilon > 0$ is small). This looks to be an essential drawback for at least two reasons:

(1) The lack of regularity prevents us from using higher-order methods for numerical simulations of the reduced IF (as a result, direct simulations for the initial *smooth* PDE using the standard methods may be more effective than simulations based on the reduced nonsmooth ODEs).

(2) $C^{1,\varepsilon}$-regularity is not enough to build up normal forms and/or study the bifurcations properly (for instance, the simplest saddle-node bifurcation requires C^2-smoothness, the Hopf bifurcation needs C^3, etc.; see [Katok and Hasselblatt 1995; Kielhöfer 2004] for more details) and, therefore, we need to return back to the initial PDE to study these bifurcations.

Thus, the natural question,

> "*Is it possible to construct a smooth* (C^k*-smooth for any finite* k) *or to extend the existing* $C^{1,\varepsilon}$*-smooth IF to a more regular one*?"

becomes crucial for the theory of inertial manifolds.

Here we give an affirmative answer to this question under the slightly stronger spectral gap assumption

$$\limsup_{N\to\infty}(\lambda_{N+1} - \lambda_N) = \infty. \tag{1-5}$$

In contrast to (1-4), this assumption does not require exponentially big spectral gaps (and is satisfied for most of the examples where the IMs exist), but guarantees the existence of infinitely many spectral gaps of size larger than $2L$ and, consequently, the existence of an infinite tower of the embedded IMs

$$\mathcal{M}_{N_1} \subset \mathcal{M}_{N_2} \subset \cdots \subset \mathcal{M}_{N_n} \subset \cdots \tag{1-6}$$

and the corresponding IFs

$$\frac{d}{dt}u_{N_n} + Au_{N_n} = \boldsymbol{P}_{N_n}F(u_{N_n} + M_{N_n}(u_{N_n})), \quad u_{N_n} \in H_{N_n}. \tag{1-7}$$

Let $n \in \mathbb{N}$ be given. We say that a $C^{n,\varepsilon}$-smooth submanifold $\widetilde{\mathcal{M}}_{N_n}$ of the phase space H (which is a graph of $C^{n,\varepsilon}$-smooth $\widetilde{M}_{N_n} : H_{N_n} \to (H_{N_n})^{\perp}$) is a $C^{n,\varepsilon}$-smooth extension of the initial IM $\mathcal{M}_{N_1}$ for some $\varepsilon > 0$ if

(1) $\mathcal{M}_{N_1} \subset \widetilde{\mathcal{M}}_{N_n}$,

(2) the manifold $\widetilde{\mathcal{M}}_{N_n}$ is C^1_b-close to the IM $\mathcal{M}_{N_n}$.

Then, the first condition guarantees that the $C^{n,\varepsilon}$-smooth system of ODEs

$$\frac{d}{dt}u_{N_n} + Au_{N_n} = \boldsymbol{P}_{N_n}F(u_{N_n} + \widetilde{M}_{N_n}(u_{N_n})), \quad u_{N_n} \in H_{N_n}, \tag{1-8}$$

will possess the initial IM $\boldsymbol{P}_{N_n}\mathcal{M}_{N_1}$ as an invariant submanifold. The second condition together with the robustness theorem for normally hyperbolic manifolds ensures that this manifold will be globally exponentially stable and normally hyperbolic (in particular, it will possess an exponential tracking property in H_{N_n}). In this case we refer to the system (1-8) as a $C^{n,\varepsilon}$-smooth extension of the corresponding IF (1-3); see Section 3 for more details. Thus, the extended IF is C^n-smooth on the one hand and, on the other hand, its limit dynamics *coincides* with the dynamics of the IF which corresponds to the IM $\mathcal{M}_{N_1}$ and, in turn, coincides with the limit dynamics of the initial abstract parabolic problem (1-1). Note that the manifold $\widetilde{\mathcal{M}}_{N_n}$ is *not necessarily* invariant under the solution semigroup $S(t)$ generated by the initial equation (1-1) and this allows us to overcome the standard obstacles to the smoothness of an invariant manifold (e.g., such as resonances, see Examples 3.11 and 5.6 below).

The main result of the paper is the following theorem which suggests a solution of the smoothness problem for IMs.

Theorem 1.1. *Let the nonlinearity $F \in C^\infty_b(H, H)$ and let the operator A satisfy the spectral gap condition (1-5). Let also $N_1 \in \mathbb{N}$ be the smallest number for which the spectral gap condition (1-2) is satisfied and $\mathcal{M}_{N_1}$ be the corresponding IM. Then, for every $n \in \mathbb{N}$, one can find $\varepsilon = \varepsilon_n > 0$ for which there exists a $C^{n,\varepsilon}$-smooth extension of the IM $\mathcal{M}_{N_1}$ as well as the $C^{n,\varepsilon}$-smooth extension of the corresponding IF in the sense described above.*

The proof of this theorem is given in Section 4 and the Appendix. To construct the desired extension $\widetilde{M}_{N_n}$, we first define it on the manifold $\boldsymbol{P}_{N_n}\mathcal{M}_{N_1}$ only in a natural way $\widetilde{M}_{N_n}(p) = (1 - \boldsymbol{P}_{N_n})M_{N_1}(\boldsymbol{P}_{N_1}p)$. Then, we present an explicit construction of Taylor jets of order n for this function via an inductive procedure; see Section 4. Finally, we check (in the Appendix) the compatibility conditions for the constructed Taylor jets and get the desired extension by the Whitney extension theorem.

Our main result can be reformulated in the following way.

Corollary 1.2. *Let the assumptions of Theorem 1.1 hold. Then, for every $n \in \mathbb{N}$, there exists $\varepsilon = \varepsilon_n > 0$ and a $C^{n-1,\varepsilon}$-smooth "correction" $\widetilde{F}_n(u)$ of the initial nonlinearity F such that:*

(1) *$\widetilde{F}_n(u) = F(u)$ for all $u \in \mathcal{M}_{N_1}$ and $\mathcal{M}_{N_1}$ is an IM for the modified equation*

$$\partial_t u + Au = \widetilde{F}_n(u), \quad u|_{t=0} = u_0, \tag{1-9}$$

as well. In particular, the dynamics of (1-9) *on* $\mathcal{M}_{N_1}$ *coincides with the initial dynamics* (*generated by* (1-1)) *and* $\mathcal{M}_{N_1}$ *possesses an exponential tracking property for solutions of* (1-9).

(2) *The extended manifold* $\widetilde{\mathcal{M}}_{N_n}$ *constructed in Theorem 1.1 is an IM* (*of smoothness* $C^{n,\varepsilon}$) *for the modified equation* (1-9); *see Corollary 5.4 below.*

In this interpretation, the modified nonlinearity $\widetilde{F}_n$ can be considered as a "cut-off" version of the initial function F and the main result claims that all obstacles for the existence of C^n-smooth IM can be removed by increasing the dimension of the IM and using a properly chosen cut-off procedure.

To conclude, we note that the main aim of this paper is to verify the principal possibility to get smooth extensions of an IM rather than to obtain the optimal bounds for the dimensions N_n of the constructed extensions. For this reason, the obtained bounds look far from being optimal, but we believe that they can be essentially improved; see Remark 5.7 for the discussion of this problem.

The paper is organized as follows. In Section 2 we recall the standard facts about smooth functions in Banach spaces, their Taylor jets, direct and converse Taylor theorems and the Whitney extension theorem, which is the main technical tool for what follows. In Section 3 we collect basic facts about the construction of IMs for semilinear parabolic equations via the Perron method and discuss known facts about the smoothness of these IMs. The main result (Theorem 1.1) is presented in Section 4. The proof of it is also given there by modulo of compatibility conditions for Whitney extension theorem which are verified in the Appendix. Finally, the applications of the proved theorem as well as a discussion of open problems and related topics are given in Section 5.

2. Preliminaries, I: Taylor expansions and the Whitney extension theorem

In this section we briefly recall the standard results on Taylor expansions of smooth functions in Banach spaces and the related Whitney extension theorem, as well as prepare some technical tools which will be used later. We start with some basic facts from multilinear algebra; see, e.g., [Hájek and Johanis 2014] for a more detailed exposition. Let X and Y be two normed spaces. For any $n \in \mathbb{N}$, we denote by $\mathcal{L}_s(X^n, Y)$ the space of multilinear continuous symmetric maps from X^n to Y endowed by the standard norm

$$\|M\|_{\mathcal{L}_s(X^n,Y)} := \sup_{\xi_i \in X,\, \xi_i \neq 0} \left\{ \frac{\|M(\xi_1, \dots, \xi_n)\|}{\|\xi_1\| \cdots \|\xi_n\|} \right\}.$$

Every element $M \in \mathcal{L}_s(X^n, Y)$ defines a homogeneous continuous polynomial P_M of order n on X with values in Y via

$$P_M(\xi) := M(\{\xi\}^n), \quad \text{where } \{\xi\}^n := \underbrace{\xi, \dots, \xi}_{n\text{-times}}.$$

Vice versa, the multilinear symmetric map $M = M_P$ can be restored in a unique way if the corresponding homogeneous polynomial is known via the polarization equality:

$$M_P(\xi_1, \dots, \xi_n) = \frac{1}{2^n n!} \sum_{\varepsilon_i = \pm 1, i=1,\dots,n} \varepsilon_1 \cdots \varepsilon_n P\left(a + \sum_{j=1}^n \varepsilon_j \xi_j\right)$$

for all $a, \xi_1, \dots, \xi_n \in X$; see, e.g., [Hájek and Johanis 2014]. Thus, there is a one-to-one correspondence between homogeneous polynomials and multilinear symmetric maps. Moreover, if we introduce the norm

$$\|P\|_{\mathcal{P}_n(X,Y)} := \sup_{\xi \neq 0} \left\{ \frac{\|P(\xi)\|}{\|\xi\|^n} \right\}$$

on the space $\mathcal{P}_n(X, Y)$ of n-homogeneous polynomials, this correspondence becomes an isometry. For this reason, we will identify below multilinear forms and the corresponding homogeneous polynomials where this does not lead to misunderstandings. We also mention here the generalization of the Newton binomial formula; namely, for any $P \in \mathcal{P}_n(X, Y)$ and $\xi, \eta \in X$, we have

$$P(\xi + \eta) = \sum_{j=0}^{n} C_n^j P(\{\xi\}^j, \{\eta\}^{n-j}), \quad C_n^j := \frac{n!}{j!(n-j)!}; \tag{2-1}$$

see, e.g., [Hájek and Johanis 2014]. Finally, we denote by $\mathcal{P}^n(X, Y)$ the space of all continuous polynomials of order less than or equal to n on X with values in Y, i.e., $P(\xi) \in \mathcal{P}^n(X, Y)$ if

$$P(\xi) = \sum_{j=0}^{n} \frac{1}{j!} P_j(\xi), \quad P_j(\xi) \in \mathcal{P}_j(X, Y).$$

The following standard result is crucial for our purposes.

Lemma 2.1. *For every $n \in \mathbb{N}$ there exist real numbers $a_{kj} \in \mathbb{R}$, $k, j \in \{0, \dots, n\}$, such that for every $P = \sum_{k=0}^{n} \frac{1}{k!} P_k$, $P_k \in \mathcal{P}_k(X, Y)$ and every $k \in \{0, \dots, n\}$, we have*

$$P_k(\xi) = \sum_{j=0}^{n} a_{kj} P\left(\frac{j}{n}\xi\right) \tag{2-2}$$

and, therefore,

$$\|P_k(\xi)\| \le K_{n,k} \max_{j=0,\dots,n} \left\| P\left(\frac{j}{n}\xi\right) \right\| \tag{2-3}$$

for some constants $K_{n,k}$ which are independent of P.

For the proof of this lemma, see [Hájek and Johanis 2014].

Corollary 2.2. *Let $P(\xi, \delta) \in P^n(X, Y)$ be a family of polynomials of ξ depending on a parameter $\delta \in B$, where B is a set in X containing zero. Assume that*

$$\|P(\xi, \delta)\| \le C(\|\xi\| + \|\delta\|)^{n+\alpha}, \quad \xi \in X,\ \delta \in B, \tag{2-4}$$

for some $\alpha \ge 0$. Then, for any $k \in \{0, \dots, n\}$,

$$\|P_k(\cdot, \delta)\|_{\mathcal{P}_k(X,Y)} \le C_k \|\delta\|^{n-k+\alpha} \tag{2-5}$$

for some constants C_k depending on C, n and k.

Proof. Indeed, according to (2-3) and (2-4), we have

$$\|P_k(\xi, \delta)\| \le C'(\|\xi\| + \|\delta\|)^{n+\alpha}.$$

Assuming that $\delta \neq 0$ (there is nothing to prove otherwise), replacing ξ by $\|\delta\|\xi$ and using that P_k is homogeneous of order k, we get

$$\|P_k(\xi, \delta)\| \leq C'(1 + \|\xi\|)^{n+\alpha}\|\delta\|^{n-k+\alpha}.$$

Using once more that P_k is homogeneous of order k in ξ, we finally arrive at

$$\|P_k(\xi, \delta)\| \leq C'\|\xi\|^k(1 + \|\xi/\|\xi\|\|^{n+\alpha})\|\delta\|^{n-k+\alpha} \leq C''\|\xi\|^k\|\delta\|^{n-k+\alpha},$$

which gives (2-5) and finishes the proof. □

Let now $U \subset X$ be an open set and let $F : U \to Y$ be a map. As usual, for any $u \in U$, we denote by $F'(u) \in \mathcal{L}(X, Y)$ the Fréchet derivative of F at u (if it exists). Analogously, for any $n \in \mathbb{N}$, we denote by $F^{(n)}(u) \in \mathcal{L}_s(X^n, Y)$ its n-th Fréchet derivative. The space of all functions $F : U \to Y$ such that $F^{(n)}(u)$ exists and is continuous as a function from U to $\mathcal{L}_s(X^n, Y)$ is denoted by $C^n(U, Y)$. For any $\alpha \in (0, 1]$, we denote by $C^{n,\alpha}(U, Y)$ the space of functions $F \in C^n(U, Y)$ such that $F^{(n)}$ is Hölder continuous with exponent α on U. The action of $F^{(n)}(u)$ to vectors $\xi_1, \dots, \xi_n \in X$ is denoted by $F^{(n)}(u)[\xi_1, \dots, \xi_n]$. The Taylor jet of length $n + 1$ of the function F at the point u and vector $\xi \in X$ will be denoted by $J^n_\xi F(u)$:

$$J^n_\xi F(u) := F(u) + \frac{1}{1!}F'(u)\xi + \frac{1}{2!}F''(u)[\xi, \xi] + \cdots + \frac{1}{n!}F^{(n)}(u)[\{\xi\}^n]. \tag{2-6}$$

Obviously, the function $\xi \to J^n_\xi F(u)$ is in $\mathcal{P}^n(X, Y)$ for every $u \in U$. We will also systematically use the truncated Taylor jets

$$j^n_\xi F(u) := \frac{1}{1!}F'(u)\xi + \frac{1}{2!}F''(u)[\xi, \xi] + \cdots + \frac{1}{n!}F^{(n)}(u)[\{\xi\}^n], \tag{2-7}$$

which do not contain zero-order terms.

Theorem 2.3 (direct Taylor theorem). *Let $F \in C^n(U, Y)$ and take $u_1, u_2 \in U$ such that $u_t := tu_1 + (1 - t)u_2 \in U$ for all $t \in [0, 1]$. Let also $\xi := u_2 - u_1$. Then*

$$F(u_2) = J^n_\xi F(u_1) + \frac{1}{n!}\int_0^1 (1 - s)^{n-1}(F^{(n)}(u_1 + s\xi) - F^{(n)}(u_1))\, ds[\{\xi\}^n]. \tag{2-8}$$

In particular, if $F \in C^{n,\alpha}(U, Y)$, then

$$\|F(u_2) - J^n_\xi F(u_1)\| \leq C\|\xi\|^{n+\alpha} \tag{2-9}$$

for some positive C.

For the proof of this classical result; see, e.g., [Hájek and Johanis 2014]. We also mention that in terms of truncated jets formula (2-9) reads

$$F(u_2) - F(u_1) = j^n_\xi F(u_1) + O(\|\xi\|^{n+\alpha}), \quad \xi := u_2 - u_1. \tag{2-10}$$

The above theorem can be inverted as follows.

Theorem 2.4 (converse Taylor theorem). *Let F be a function such that, for any $u \in U$, there exists a polynomial $\xi \to P(\xi, u) \in \mathcal{P}^n(X, Y)$ such that, for all $u_1, u_2 \in U$,*

$$\|F(u_2) - P(\xi, u_1)\| \le C\|\xi\|^{n+\alpha}, \quad \xi := u_2 - u_1, \tag{2-11}$$

for some $C > 0$ and $\alpha \in (0, 1]$. Then, $F \in C^n(U, Y)$,

$$P(\xi, u) = J^n_\xi F(u)$$

for all $u \in U$ and $F^{(n)}(u)$ is locally Hölder continuous in U with exponent α. If, in addition, U is convex, then $F \in C^n(U, Y)$ and

$$\|F^{(n)}(u_2) - F^{(n)}(u_1)\| \le C'\|u_2 - u_1\|^\alpha,$$

where C' depends only on n, α and the constant C from (2-11).

For the proof of this theorem, see [Hájek and Johanis 2014].

Keeping in mind the Whitney extension problem, we recall that an arbitrarily chosen set of polynomials $P(\xi, u)$, $u \in U$, does not define in general a $C^{n,\alpha}$- smooth function, but some compatibility conditions must be satisfied for that. Indeed, let $u_1 \in U$ and let $\delta, \xi \in X$ be such that $u_2 := u_1 + \delta \in U$ and $u_3 := u_1 + \delta + \xi = u_2 + \xi \in X$. Then, from (2-9), we have

$$\|F(u_3) - P(\xi + \delta, u_1)\| \le C\|\xi + \delta\|^{n+\alpha},$$
$$\|F(u_3) - P(\xi, u_1 + \delta)\| \le C\|\xi\|^{n+\alpha}.$$

Therefore,

$$\|P(\xi + \delta, u_1) - P(\xi, u_1 + \delta)\| \le C_1(\|\xi\| + \|\delta\|)^{n+\alpha}. \tag{2-12}$$

These are the desired compatibility conditions. In other words, if we are given a set $V \subset X$ and a family of polynomials

$$\{P(\xi, u) : u \in V\} \subset P^n(X, Y)$$

and want to find a function $F \in C^{n,\alpha}(X, Y)$ such that $J^n_\xi F(u) = P(\xi, u)$ for all $u \in V$, then the compatibility condition (2-12) must be satisfied for all $u_1, u_1 + \delta \in V$ and all $\xi \in X$.

Inequality (2-12) can be rewritten in a more standard form, which usually appears in the statement of the Whitney extension theorem. Namely, using (2-1), we see that

$$P(\xi + \delta, u_1) = \sum_{l=0}^{n} \frac{1}{l!} \sum_{k=l}^{n} \frac{1}{(k-l)!} P_k([\{\xi\}^l, \{\delta\}^{k-l}], u_1),$$

where $P(\xi, u_1) = \sum_{l=0}^{n} (1/l!) P_l([\{\xi\}^l], u_1)$, $P_l(\cdot, u_1) \in \mathcal{P}_l(X, Y)$. Applying now Corollary 2.2 to (2-12), we get the desired alternative form of the compatibility conditions:

$$\left\| P_l(\{\xi\}^l, u_1 + \delta) - \sum_{k=0}^{n-l} \frac{1}{k!} P_{l+k}([\{\xi\}^l, \{\delta\}^k], u_1) \right\| \le C\|\xi\|^l \|\delta\|^{n-l+\alpha} \tag{2-13}$$

for $l = \{0, \dots, n\}$. The compatibility condition (2-13) has a natural interpretation: if $P_k(\{\xi\}, u_1) = F^{(k)}(u_1)[\{\xi\}^k]$ as we expect, then (2-13) is nothing more than Taylor expansions of $F^{(l)}(u_1 + \delta)[\{\xi\}^l]$ at u_1.

The next theorem shows that the introduced compatibility conditions are sufficient for the existence of F in the case when X is finite-dimensional.

Theorem 2.5 (Whitney extension theorem). *Let* $\dim X < \infty$ *and let* V *be an arbitrary subset of* X. *Assume also that we are given a family of polynomials* $\{P(\xi, u) : u \in V\} \subset P^n(X, Y)$ *which satisfies the compatibility condition* (2-12) *with some* $\alpha \in (0, 1]$. *Then, there exists a function* $F \in C^{n,\alpha}(X, Y)$ *such that* $J^n_\xi F(u) = P(\xi, u)$ *for all* $u \in V$.

For the proof of this theorem, see [Stein 1970; Fefferman 2005]. Note that the theorem fails if the dimension of X is infinite, but there are no restrictions on the dimension of the space Y; see [Wells 1973].

3. Preliminaries, II: Spectral gaps and the construction of an inertial manifold

In this section we briefly discuss the classical theory of inertial manifolds for semilinear parabolic equations; see, e.g., [Zelik 2014] for a more detailed exposition.

Let H be an infinite-dimensional real Hilbert space. Let us consider an abstract parabolic equation in H:

$$\partial_t u + Au = F(u), \quad u|_{t=0} = u_0, \tag{3-1}$$

where $A : D(A) \to H$ is a linear self-adjoint positive operator in H with compact inverse and $F \in C^\infty_b(H, H)$ is a smooth bounded function on H such that all its derivatives are also bounded on H.

It is well known that under the above assumptions (3-1) is globally well-posed for any $u_0 \in H$ in the class of solutions $u \in C([0, T], H)$ for all $T > 0$ and, therefore, generates a semigroup in H:

$$S(t) : H \to H, \quad t \geq 0, \quad S(t)u_0 := u(t). \tag{3-2}$$

Moreover, the solution operators $S(t)$ are in $C^\infty(H, H)$ for every fixed $t \geq 0$; see [Henry 1981; Zelik 2014] for the details.

Let $0 < \lambda_1 \leq \lambda_2 \leq \cdots$ be the eigenvalues of the operator A enumerated in the nondecreasing order and let $\{e_n\}_{n=1}^\infty$ be the corresponding orthonormal system of eigenvectors. Then, by the Parseval equality, for every $u \in H$, we have

$$\|u\|_H^2 = \sum_{n=1}^\infty (u, e_n)^2, \quad u = \sum_{n=1}^\infty (u, e_n) e_n,$$

where $(\cdot, \cdot)$ is an inner product in H. For a given $N \in \mathbb{N}$, we denote by $\boldsymbol{P}_N$ and $\boldsymbol{Q}_N$ the orthoprojectors on the first N and the rest of eigenvectors of A respectively:

$$\boldsymbol{P}_N u := \sum_{n=1}^N (u, e_n) e_n, \quad \boldsymbol{Q}_N u := \sum_{n=N+1}^\infty (u, e_n) e_n.$$

We are now ready to introduce the main object of study in this paper — an inertial manifold (IM).

Definition 3.1. A set $\mathcal{M} = \mathcal{M}_N$ is an inertial manifold of dimension N for problem (3-1) (with the base $H_N := \boldsymbol{P}_N H$) if

(1) $\mathcal{M}$ is invariant with respect to the semigroup $S(t)$: $S(t)\mathcal{M} = \mathcal{M}$.

(2) $\mathcal{M}$ is a graph of a Lipschitz continuous function $M : H_N \to \boldsymbol{Q}_N H$:

$$\mathcal{M} = \{p + M(p) : p \in H_N\}.$$

(3) $\mathcal{M}$ possesses an exponential tracking property, namely, for every trajectory $u(t)$ of (3-1) there exists a trace solution $\bar{u}(t) \in \mathcal{M}$ such that

$$\|u(t) - \bar{u}(t)\| \le Ce^{-\theta t}, \quad t \ge 0, \tag{3-3}$$

for some $\theta > \lambda_N$ and constant $C = C_u$ which depends on u.

Note that, although only Lipschitz continuity is traditionally required in the definition, usually IMs are $C^{1,\varepsilon}$-smooth for some $\varepsilon > 0$ (see the discussion below) and are normally hyperbolic. Then the exponential tracking property (that is, existence of an asymptotic phase), as well as robustness with respect to perturbations, are the standard corollaries of this normal hyperbolicity; see [Bates et al. 1999; Fenichel 1972; Katok and Hasselblatt 1995; Rosa and Temam 1996] for the details. We also mention that these results can be obtained without formally referring to normal hyperbolicity; see, e.g., [Foias et al. 1988], as well as Theorem 3.2 and Corollary 4.4 below.

Note also the dynamics of (3-1) restricted to the IM $\mathcal{M}$ is governed by the system of ODEs

$$\frac{d}{dt}u_N + Au_N = \boldsymbol{P}_N F(u_N + M(u_N)), \quad u_N := \boldsymbol{P}_N u \in \mathbb{R}^N, \tag{3-4}$$

which is called an inertial form (IF) associated with (3-1). In the case where the spectral subspace H_N is used as a base for IM (like in Definition 3.1), the regularity of the corresponding vector field in the IF is determined by the regularity of the IM only.

The following theorem is the key result in the theory of IMs.

Theorem 3.2. *Let the function F in (3-1) be globally Lipschitz continuous with Lipschitz constant L and let, for some $N \in \mathbb{N}$, the spectral gap condition*

$$\lambda_{N+1} - \lambda_N > 2L \tag{3-5}$$

be satisfied. Then (3-1) possesses an IM $\mathcal{M}_N$ of dimension N.

Proof. Although this statement is classical, see, e.g., [Foias et al. 1988; Miklavčič 1991; Romanov 1993; Zelik 2014], the elements of its proof will be crucially used in what follows, so we sketch them below.

To construct the IM, we will use the so-called Perron method; namely, we will prove that, for every $p \in H_N$, the problem

$$\partial_t u + Au = F(u), \quad t \le 0, \quad \boldsymbol{P}_N u|_{t=0} = p \tag{3-6}$$

possesses a unique backward solution $u(t) = V(p, t)$, $t \le 0$, belonging to an appropriately weighted space, and then define the desired map $M : H_N \to \boldsymbol{Q}_N H$ via

$$M(p) := \boldsymbol{Q}_N V(p, 0). \tag{3-7}$$

To solve (3-6) we use the Banach contraction theorem treating the nonlinearity F as a perturbation. To this end we need the following two lemmas.

Lemma 3.3. *Let $\theta \in (\lambda_N, \lambda_{N+1})$ and let us consider the equation*

$$\partial_t v + Av = h(t), \quad t \in \mathbb{R},\ h \in L^2_{e^{\theta t}}(\mathbb{R}, H), \tag{3-8}$$

where the space $L^2_{e^{\theta t}}(\mathbb{R}, H)$ is defined via the weighted norm

$$\|h\|^2_{L^2_{e^{\theta t}}(\mathbb{R},H)} := \int_{t\in\mathbb{R}} e^{2\theta t}\|h(t)\|^2\,dt < \infty. \tag{3-9}$$

Then, problem (3-8) possesses a unique solution $u \in L^2_{e^{\theta t}}(\mathbb{R}, H)$ and the solution operator $\mathcal{T} : L^2_{e^{\theta t}} \to L^2_{e^{\theta t}}$, $\mathcal{T} : h \mapsto u$ satisfies

$$\|\mathcal{T}\|_{\mathcal{L}(L^2_{e^{\theta t}}, L^2_{e^{\theta t}})} = \frac{1}{\min\{\theta - \lambda_N, \lambda_{N+1} - \theta\}}. \tag{3-10}$$

The proof of this identity is just a straightforward calculation based on decomposition of the solution $u(t)$ with respect to the base $\{e_n\}_{n=1}^{\infty}$ and solving the corresponding ODEs; see [Zelik 2014].

The second lemma gives the analogue of this formula for the linear equation on a negative semiaxis.

Lemma 3.4. *Let $\theta \in (\lambda_N, \lambda_{N+1})$. Then, for any $p \in H_N$ and any $h \in L^2_{e^{\theta t}}(\mathbb{R}_-, H)$, the problem*

$$\partial_t v + Av = h(t), \quad t \le 0, \quad \boldsymbol{P}_N v|_{t=0} = p \tag{3-11}$$

possesses a unique solution $v \in L^2_{e^{\theta t}}(\mathbb{R}_-, H)$. This solution can be written in the form

$$v = \mathcal{T}h + \mathcal{H}p,$$

where $\mathcal{T}$ is exactly the solution operator constructed in Lemma 3.3 applied to the extension of the function $h(t)$ by zero for $t \ge 0$ and $\mathcal{H} : H_N \to L^2_{e^{\theta t}}(\mathbb{R}_-, H)$ is a solution operator for the problem with zero right-hand side:

$$\mathcal{H}(p, t) := \sum_{n=1}^{N} (p, e_n) e^{-\lambda_n t}.$$

Indeed, this lemma is an easy corollary of Lemma 3.3; see [Zelik 2014].

We are now ready to prove the theorem. To this end, we fix the optimal value $\theta = (\lambda_{N+1} + \lambda_N)/2$ and write (3-6) as a fixed-point problem

$$u = \mathcal{T} \circ F(u) + \mathcal{H}(p) \tag{3-12}$$

in the space $L^2_{e^{\theta t}}(\mathbb{R}_-, H)$. Since the norm of the operator $\mathcal{T}$ is equal to $2/(\lambda_{N+1} - \lambda_N)$ and the Lipschitz constant of F is L, the spectral gap condition (3-5) guarantees that the right-hand side of (3-12) is a contraction for every $p \in H_N$. Thus, by the Banach contraction theorem, for every $p \in H_N$, there exists a unique solution $u(t) = V(p, t)$ of problem (3-6) belonging to $L^2_{e^{\theta t}}(\mathbb{R}_-, H)$ and the map $p \mapsto V(p, \cdot)$ is Lipschitz continuous. Due to the parabolic smoothing property, we know that

$$\|u(0)\| \le C(1 + \|u\|_{L^2([-1,0],H)}) \quad \text{and} \quad \|u(0) - w(0)\| \le C\|u\|_{L^2([-1,0],H)}$$

for any two backward solutions u, w of (3-1); see, e.g., [Zelik 2014]. In particular, these formulas show that the solution $V(p, t)$ is continuous in time ($V(p, \cdot) \in C_{e^{\theta t}}(\mathbb{R}_-, H)$, where the weighted space of continuous functions is defined analogously to (3-9)) and the map $p \mapsto V(p, \cdot)$ is Lipschitz continuous as

a map from H_N to $C_{e^{\theta t}}(\mathbb{R}_-, H)$. Thus, formula (3-7), defines indeed a Lipschitz manifold of dimension N over the base H_N as graph of Lipschitz continuous function $M : H_N \to \boldsymbol{Q}_N H$.

The invariance of this manifold follows by the construction, so we only need to verify the exponential tracking property.

Let $u(t) = S(t)u_0$ be an arbitrary solution of problem (3-1) and let $\phi(t) \in C^\infty(R)$ be a cut-off function such that $\phi(t) \equiv 0$ for $t \le 0$ and $\phi(t) \equiv 1$ for $t \ge 1$. Then the function $\phi(t)u(t)$ is defined for all $t \in \mathbb{R}$. We seek for the desired solution $\bar{u}(t) \in \mathcal{M}$ (by the construction of $\mathcal{M}$ such solutions are defined for all $t \in \mathbb{R}$) in the form

$$\bar{u}(t) = \phi(t)u(t) + v(t). \tag{3-13}$$

Inserting this anzatz to (3-1), we end up with the following equation for $v(t)$:

$$\partial_t v + Av = F(\phi u + v) - \phi F(u) - \phi' u. \tag{3-14}$$

Let $v \in L^2_{e^{\theta t}}(\mathbb{R}, H)$ be a solution of this equation. Then, since $\bar{u} = v$ for $t \le 0$, we necessarily have $\bar{u} \in \mathcal{M}$ by the construction of the IM. On the other hand, for $t \ge 1$, we have $v = \bar{u} - u \in L^2_{e^{\theta t}}([1, \infty), H)$ and using the parabolic smoothing again, we get the desired estimate (3-3). Thus, we only need to find such a solution $v(t)$. To this end, we invert the linear part of (3-14) to get the fixed-point equation

$$v = \mathcal{T}(F(\phi u + v) - \phi F(u) - \phi' u). \tag{3-15}$$

It is straightforward to verify using Lemma 3.3 that the right-hand side of (3-15) is a contraction on the space $L^2_{e^{\theta t}}(\mathbb{R}, H)$ if the spectral gap condition holds; see [Zelik 2014]. Thus, the Banach contraction theorem finishes the proof of exponential tracking. □

Remark 3.5. It is well known that the spectral gap condition (3-5) is sharp in the sense that if it is violated for some N and L, one can find a nonlinearity F such that (3-1) does not possess an IM of dimension N with base H_N; see [Romanov 1993].

More recent examples show that if the condition

$$\sup_{N \in \mathbb{N}} \{\lambda_{N+1} - \lambda_N\} < 2L$$

is violated for all N, one can construct a smooth nonlinearity F such that (3-1) does not possess any Lipschitz or even log-Lipschitz finite-dimensional manifold (not necessarily invariant) which contains the global attractor; see [Eden et al. 2013; Zelik 2014].

Remark 3.6. Theorem 3.2 guarantees the existence of an IM $\mathcal{M}_N$ for *every* N such that the spectral gap condition (3-5) is satisfied. Typically, this N is not unique, instead, we have a whole sequence $\{N_k\}_{k=1}^\infty$ of N's satisfying the spectral gap condition. Therefore, according to the theorem, we will have a sequence of IMs $\{\mathcal{M}_{N_k}\}_{k=1}^\infty$ of increasing dimensions: $N_1 < N_2 < N_3 < \cdots$. Moreover, from the explicit description of an IM using backward solutions of (3-6), we see that

$$\mathcal{M}_{N_1} \subset \mathcal{M}_{N_2} \subset \mathcal{M}_{N_3} \subset \cdots; \tag{3-16}$$

see [Foias et al. 1988] for more details. In this case it can be also proved that $\mathcal{M}_{N_{k-1}}$ is a normally hyperbolic submanifold of $\mathcal{M}_{N_k}$.

Let us now discuss the further regularity of the IM $\mathcal{M}$. To this end, we need one more auxiliary statement.

Proposition 3.7. *Let the spectral gap condition (3-5) hold and let $u(t) \in C(R_-, H)$ be an arbitrary function. Let also the exponent $\theta \in (\lambda_N, \lambda_{N+1})$ satisfy*

$$\theta_- := L + \lambda_N < \theta < \lambda_{N+1} - L := \theta_+. \tag{3-17}$$

Then, for any $h \in L^2_{e^{\theta t}}(\mathbb{R}_-, H)$ and every $p \in H_N$, the corresponding equation of variations

$$\partial_t v + Av - F'(u(t))v = h(t), \quad t \le 0, \quad \boldsymbol{P}_N v|_{t=0} = p \tag{3-18}$$

possesses a unique solution $v \in L^2_{e^{\theta t}}(\mathbb{R}_-, H) \cap C_{e^{\theta t}}(\mathbb{R}_-, H)$ and the following estimate holds:

$$\|v\|_{C_{e^{\theta t}}(\mathbb{R}_-,H)} \le C\|v\|_{L^2_{e^{\theta t}}(\mathbb{R}_-,H)} \le C_{L,\theta}(\|h\|_{L^2_{e^{\theta t}}(\mathbb{R}_-,H)} + \|p\|), \tag{3-19}$$

where the constant $C_{L,\theta}$ is independent of u, h and p.

Indeed, (3-18) can be solved via the Banach contraction theorem treating the term $F'(u)v$ as a perturbation analogously to the nonlinear case. Inequalities (3-17) guarantee that the map $\mathcal{T}F'(u)v$ is a contraction on $L^2_{e^{\theta t}}(\mathbb{R}_-, H)$, due to (3-10).

Corollary 3.8. *Let the assumptions of Theorem 3.2 hold and let, in addition, the exponent $\varepsilon \in (0, 1]$ be such that*

$$\lambda_{N+1} - (1+\varepsilon)\lambda_N > (2+\varepsilon)L. \tag{3-20}$$

Assume also that $F \in C^{1,\varepsilon}(H, H)$. Then the associated IM $\mathcal{M}_N$ is $C^{1,\varepsilon}$-smooth, for any $p, \xi \in H_N$, the derivative $M'(p)\xi$ can be found as the value of the $\boldsymbol{Q}_N$ projection of $V'(t) = V'(p,t)\xi$ at $t = 0$, where the function V' solves the equation of variations

$$\partial_t V' + AV' - F'(u(t))V' = 0, \quad t \le 0, \quad \boldsymbol{P}_N V'|_{t=0} = \xi, \quad u(t) := V(p,t), \tag{3-21}$$

and

$$\|M'(p_1) - M'(p_2)\|_{\mathcal{L}(H_N,H)} \le C\|p_1 - p_2\|^{\varepsilon}$$

for some constant C independent of $p_1, p_2 \in H_N$.

Proof. Let $p_1, p_2 \in H_N$ and $u_i(t) := V(p_i, t)$ be the corresponding trajectories belonging to the IM. Let also $v(t) := u_1(t) - u_2(t)$ and $\xi := p_1 - p_2$. Then v solves

$$\partial_t v + Av - L_{u_1,u_2}(t)v = 0, \quad t \le 0, \quad \boldsymbol{P}_N v|_{t=0} = \xi, \tag{3-22}$$

where $L_{u_1,u_2}(t) := \int_0^1 F'(su_1(t) + (1-s)u_2(t))\,ds$. Since the norm of $L_{u_1,u_2}(t)$ does not exceed L, Proposition 3.7 is applicable to (3-22) and, therefore, for every θ satisfying (3-17), we have the estimate

$$\|v\|_{C_{e^{\theta t}}(\mathbb{R}_-,H)} \le C\|v\|_{L^2_{e^{\theta t}}(\mathbb{R}_-,H)} \le C_\theta\|p_1 - p_2\|. \tag{3-23}$$

Note also that the function $V'(p,t)\xi$ is well-defined for all $p, \xi \in H_N$ due to Proposition 3.7 and satisfies the analogue of (3-23). Let $w(t) := v(t) - V'(p_1, t)\xi$, with $\xi := p_1 - p_2$. Then, this function solves

$$\partial_t w + Aw - F'(u_1)w = F(u_1) - F(u_2) - F'(u_1)v := h_{u_1,u_2}(t), \quad \boldsymbol{P}_N w|_{t=0} = 0. \tag{3-24}$$

Since $F \in C^{1,\varepsilon}(H, H)$, by the Taylor theorem, we have

$$\|h_{u_1,u_2}(t)\| \le C\|v(t)\|^{1+\varepsilon},$$

which, due to (3-23), gives

$$\|h_{u_1,u_2}\|_{L^2_{e^{(1+\varepsilon)\theta t}}(\mathbb{R}_-,H)} \le C\|v\|_{L^2_{e^{\theta t}}(\mathbb{R}_-,H)}\|v\|^{\varepsilon}_{C_{e^{\theta t}}(\mathbb{R}_-,H)} \le C'\|\xi\|^{1+\varepsilon}.$$

Fixing now θ in such a way that $\theta > \theta_-$ and $(1+\varepsilon)\theta < \theta_+$ (this is possible to do due to assumption (3-20)) and applying Proposition 3.7 to (3-24), we finally arrive at

$$\|M(p_2) - M(p_1) - M'(p_1)\xi\| = \|w(0)\| \le C_1\|w\|_{L^2_{e^{(1+\varepsilon)\theta t}}(\mathbb{R}_-,H)} \le C_2\|\xi\|^{1+\varepsilon}$$

and the converse Taylor theorem finishes the proof of the corollary. □

The next corollary claims that the constructed manifold $\mathcal{M}$ actually lives in a more regular space $H^2 := D(A)$.

Corollary 3.9. *Let the assumptions of Corollary 3.8 hold. Then the manifold $\mathcal{M}$ is simultaneously a $C^{1,\varepsilon}$-smooth IM for (3-1) in the phase space $H^2 = D(A)$.*

Proof. This is an almost immediate corollary of the parabolic smoothing property. Indeed, let us first check that $\mathcal{M} \in H^2$. To this end, it is enough to check that the backward solution (3-6) actually belongs to $C_{e^{\theta t}}(\mathbb{R}_-, H^2)$. First, using the $L^2(H^2)$-maximal regularity for the solutions of a linear parabolic equation

$$\partial_t v + Av = h(t), \quad t \le 0, \tag{3-25}$$

namely, that

$$\|v\|_{C^\alpha(-1,0;H)} + \|\partial_t v\|_{L^2(-1,0;H)} + \|Av\|_{L^2(-1,0;H)} \le C_\alpha(\|h\|_{L^2(-2,0,H)} + \|v\|_{L^2(-2,0,H)}), \tag{3-26}$$

where $\alpha \in \left(0, \frac{1}{2}\right)$, we end up with the estimate

$$\|u\|_{C^\alpha(-1,0;H)} \le C_\alpha(\|F(u)\|_{L^2(-2,0;H)} + \|u\|_{L^2(-2,0;H)}) \le C_{\alpha,\theta}(1 + \|u\|_{L^2_{e^{\theta t}}(\mathbb{R}_-,H)}) \le C(1 + \|p\|), \tag{3-27}$$

where $\alpha \in \left(0, \frac{1}{2}\right)$. Second, using the $C^\alpha(H)$-maximal regularity for solutions of (3-25) and the obvious estimate

$$\|F(u)\|_{C^\alpha(-2,0;H)} \le \|F\|_{C^\alpha(H,H)}(1 + \|u\|^{\alpha}_{C^\alpha(-2,0;H)}),$$

we arrive at

$$\begin{aligned}\|\partial_t u\|_{C^\alpha(-1,0;H)} + \|Au\|_{C^\alpha(-1,0;H)} &\le C(\|F(u)\|_{C^\alpha(-2,0;H)} + \|u\|_{C^\alpha(-2,0;H)}) \\ &\le C_1(1 + \|u\|_{C^\alpha(-2,0;H)}) \le C_2(1 + \|p\|)\end{aligned} \tag{3-28}$$

and the fact that $M(p)$ belongs to H^2 is proved. The fact that M is $C^{1,\varepsilon}$-smooth as a map from H_N to H^2 can be verified analogously and the corollary is proved. □

Remark 3.10. The analogue of Corollary 3.8 holds for higher derivatives as well. For instance, if we want to have a $C^{n,\varepsilon}$-smooth IM, we need to require that

$$\lambda_{N+1} - (n+\varepsilon)\lambda_N > (n+1+\varepsilon)L. \tag{3-29}$$

To verify this, we just need to define the higher-order Taylor jets for the IM $\mathcal{M}$ using second, third, etc., equations of variations for (3-6) and use again Proposition 3.7. For instance, the second derivative $V'' = V''(p,t)[\xi,\xi]$ solves

$$\partial_t V'' + AV'' - F'(u(t))V'' = F''(u(t))[V'(p,t)\xi, V'(p,t)\xi], \quad \boldsymbol{P}_N V''|_{t=0} = 0, \quad u(t) := V(p,t). \tag{3-30}$$

According to Proposition 3.7, in order to be able to solve this equation, we need $\theta_+ > 2\theta_-$ (since $V' \in L^2_{e^{\theta t}}$ with $\theta > \theta_-$ and the right-hand side $F''(u)[V', V'] \in L^2_{e^{2\theta t}}$), which gives (3-29) for $n = 2$.

We believe that sufficient condition (3-29) for the existence of a $C^{n,\varepsilon}$-smooth IM is sharp for any n and ε, but we restrict ourselves by recalling below the classical counterexample of G. Sell to the existence of C^2-smooth IM which demonstrates the sharpness of (3-29) for $n = 2$; see [Chow et al. 1992].

Example 3.11. Let $H := l^2$ (the space of square summable sequences with the standard inner product) and let us consider the following particular case of (3-1):

$$\frac{d}{dt}u_1 + u_1 = 0, \quad \frac{d}{dt}u_n + 2^{n-1}u_n = u_{n-1}^2, \quad n = 2, 3, \dots. \tag{3-31}$$

Here $\lambda_n = 2^{n-1}$ and we have a set of resonances $2\lambda_n = \lambda_{n+1}$ which prevent the existence of any finite-dimensional invariant local manifold of dimension greater than zero which is C^2-smooth and contains zero. Note that the nonlinearity here is locally smooth near zero and since we are interested in *local* invariant manifolds near zero, the behaviour of it outside the small neighbourhood of zero is not important (we may always cut-off it outside of the neighbourhood to get global Lipschitz continuity). Moreover, since $F'(0) = 0$, decreasing the size of the neighbourhood we may make the Lipschitz constant L as small as we want. Thus, according to Corollary 3.8, for any $N \in \mathbb{N}$, there exists a local invariant manifold $\mathcal{M}_N$ of dimension N with the base H_N which is $C^{1,\varepsilon}$-smooth for any $\varepsilon < 1$.

Let us check that a C^2-smooth invariant local manifold does not exist. Indeed, let $\mathcal{M}_N$ be such a manifold of dimension N. Then, since the tangent plane $T\mathcal{M}_N(0)$ to this manifold at zero is invariant with respect to A (due to the fact that $F'(0) = 0$), we must have

$$H'_N := T\mathcal{M}_N(0) = \operatorname{span}\{e_{n_1}, \dots, e_{n_N}\}$$

for some $n_1 < n_2 < \dots < n_N$. Thus, the manifold $\mathcal{M}_N$ can be presented locally near zero as a graph of a C^2-function $M : H'_N \to (H'_N)^\perp$ such that $M(0) = M'(0) = 0$. In particular, expanding M in Taylor series near zero, we have

$$u_{n_N+1} = (M(u_{n_1}, \dots, u_{n_N}), e_{n_N+1}) = cu_{n_N}^2 + \cdots.$$

Let us try to compute the constant c. Inserting this in the (n_N+1)-th equation and using the invariance, we get

$$\begin{aligned} \partial_t u_{n_N+1} + 2^{n_N}u_{n_N+1} &= 2c\partial_t u_{n_N}u_{n_N} + 2^{n_N}cu_{n_N}^2 + \cdots \\ &= -2c2^{n_N-1}u_{n_N}^2 + 2^{n_N}cu_{n_N}^2 + \cdots = 0 + \cdots = u_{n_N}^2, \end{aligned} \tag{3-32}$$

which gives $0 = 1$. Thus, the manifold $\mathcal{M}_N$ cannot be C^2-smooth.

Remark 3.12. Note that in the case where A is an elliptic operator of order $2k$ in a bounded domain Ω of $\mathbb{R}^d$, we have $\lambda_n \sim Cn^{2k/d}$ due to the Weyl asymptotic. Thus, one may expect in general only gaps of

the size

$$\lambda_{N+1} - \lambda_N \sim CN^{2k/d-1} \sim C'\lambda_N^{1-d/(2k)}, \tag{3-33}$$

which is much weaker than (3-29) with $n > 1$. Sometimes the exponent in the right-hand side of (3-33) may be improved due to multiplicity of eigenvalues (e.g., for the Laplace–Beltrami operator on a sphere S^d, we have $\lambda_N^{1/2}$ instead of $\lambda_N^{1-d/(2k)}$ in the right-hand side of (3-33) for infinitely many values of N, no matter how big the dimension d is), but this exponent is always *less than one* in all more or less realistic examples. Thus, the existence of C^n-smooth IMs with $n > 1$ looks unrealistic and could be obtained in general, but only for bifurcation problems where, e.g., $\lambda_1, \dots, \lambda_N$ are close to zero, λ_{N+1} is of order 1 and L is small.

In contrast to this, if the spectral gap condition (3-5) is satisfied for some N, i.e., $\lambda_{N+1} - \lambda_N > 2L$, we always can find a small positive $\varepsilon = \varepsilon_N$ such that $\lambda_{N+1} - (1+\varepsilon)\lambda_N > (2+\varepsilon)N$ and, therefore, (3-20) will be also satisfied. Thus, if the nonlinearity F is smooth enough, we automatically get a $C^{1,\varepsilon}$-smooth IM for some small ε depending on N and L.

Remark 3.13. Let $\bar{u}(t)$ be a trajectory of (3-1) belonging to the IM, i.e.,

$$\boldsymbol{Q}_N \bar{u}(t) \equiv M_N(\boldsymbol{P}_N \bar{u}(t))$$

and let $\bar{u}_N := \boldsymbol{P}_N \bar{u}(t)$. Then, we may write a linearization near the trajectory $\bar{u}(t)$ in two natural ways. First, we may just linearize (3-1) without using the fact that $\bar{u} \in \mathcal{M}_N$. This gives the equation

$$\partial_t v + Av - F'(\bar{u})v = h(t), \tag{3-34}$$

which we have used above to get the existence of the IM, its smoothness and exponential tracking.

Alternatively, we may linearize the reduced ODEs (3-4):

$$\partial_t v_N + Av_N - F'(\bar{u})(v_N + M_N'(\bar{u})v_N) = h_N(t). \tag{3-35}$$

Of course, these two equations are closely related. Namely, if $v_N(t)$ solves (3-35), then the function

$$v(t) := v_N(t) + M_N'(\bar{u}(t))v_N(t) \tag{3-36}$$

solves (3-34) with

$$h(t) := h_N(t) + M_N'(\bar{u}(t))h_N(t). \tag{3-37}$$

Vice versa, if $h(t)$ satisfies (3-37) and the solution $v(t)$ of (3-34) satisfies (3-36) for some t, then it satisfies (3-36) for all t and $v_N(t) := \boldsymbol{P}_N v(t)$ solves (3-35).

This equivalence is a straightforward corollary of the invariance of the manifold $\mathcal{M}_N$ and we leave its rigorous proof to the reader.

4. Main result

In this section we develop an alternative approach for constructing C^n-smooth IFs which does not require huge spectral gaps. The key idea is to require instead the existence of *many* spectral gaps and to use the second spectral gap in order to solve (3-30) for the second derivative, the third gap to solve the appropriate equation for the third derivative, etc. Of course, this will not allow us to construct a C^n-smooth IM (we

know that it may not exist for $n > 1$, see Example 3.11). Instead, for every $p \in \mathcal{M}_{N_2}$ and the corresponding trajectory $u = V(p,t)$, we construct the corresponding Taylor jet $J^n_\xi V(p,t)$ of length $n+1$ belonging to the space $\mathcal{P}^n(H_{N_n}, H)$ for all $t \le 0$, where N_k is the dimension of the IM $\mathcal{M}_{N_k}$ built up on the k-th spectral gap. These jets must be constructed in such a way that the compatibility conditions are satisfied. Then, the Whitney embedding theorem will give us the desired smooth extension of the initial IM. To be more precise, we give the following definition of such a smooth extension.

Definition 4.1. Let (3-1) possess at least two spectral gaps which correspond to the dimensions K_1 and K_2 and let $\varepsilon > 0$ be a small number. Denote the corresponding IMs by $\mathcal{M}_{K_1}$ and $\mathcal{M}_{K_2}$ respectively; the corresponding $C^{1,\varepsilon}$-functions generating these manifolds are denoted by M_{K_1} and M_{K_2} respectively. A $C^{n,\varepsilon}$-smooth submanifold $\widetilde{\mathcal{M}}_{K_2}$ (not necessarily invariant) of dimension K_2 is called a C^n-extension of the IM $\mathcal{M}_{K_1}$ if the following conditions hold:

(1) $\widetilde{\mathcal{M}}_{K_2}$ is a graph of a $C^{n,\varepsilon}$-smooth function $\widetilde{M}_{K_2} : \boldsymbol{P}_{K_2} H \to \boldsymbol{Q}_{K_2} H$.

(2) $\widetilde{M}_{K_2}|_{\boldsymbol{P}_{K_2}\mathcal{M}_{K_1}} = \boldsymbol{Q}_{K_2} M_{K_1}$ and therefore $\mathcal{M}_{K_1} \subset \widetilde{\mathcal{M}}_{K_2}$.

(3) $\widetilde{M}_{K_2}$ is μ-close in the C^1_b-norm to M_{K_2} for a sufficiently small μ.

Remark 4.2. The $C^{n,\varepsilon}$ dynamics on the extended IM $\widetilde{\mathcal{M}}_{K_2}$ is naturally defined via

$$\partial_t u_{K_2} + A u_{K_2} = \boldsymbol{P}_{K_2} F(u_{K_2} + \widetilde{M}_{K_2}(u_{K_2})), \quad u_{K_2} \in H_{K_2}, \tag{4-1}$$

and $u(t) := u_{K_2}(t) + \widetilde{M}_{K_2}(u_{K_2}(t))$. Obviously, the manifold $\widetilde{\mathcal{M}}_{K_2}$ is invariant with respect to the dynamical system thus defined. Moreover, due to the second condition of Definition 4.1, the $C^{1,\varepsilon}$-submanifold $\boldsymbol{P}_{K_2}\mathcal{M}_{K_1} \subset H_{K_2}$ is invariant with respect to (4-1) and the restriction of (4-1) coincides with the initial IF (3-4) generated by the IM $\mathcal{M}_{K_1}$. Thus, system of ODEs (4-1) is indeed a smooth extension of the IF (3-4).

Finally, the third condition of Definition 4.1 guarantees that $\boldsymbol{P}_{K_2}\mathcal{M}_{K_1}$ is a normally hyperbolic stable invariant manifold for (4-1) (since it is so for the IF generated by the function M_{K_2}). This means that $\boldsymbol{P}_{K_2}\mathcal{M}_{K_1}$ also possesses an exponential tracking property. Thus, the limit dynamics generated by the extended IF coincides with the one generated by the initial abstract parabolic equation (3-1).

We are now ready to state the main result of the paper.

Theorem 4.3. *Let the nonlinearity $F : H \to H$ in (3-1) be smooth and all its derivatives be globally bounded. Let also the following form of spectral gap conditions be satisfied*:

$$\limsup_{N\to\infty} (\lambda_{N+1} - \lambda_N) = \infty. \tag{4-2}$$

Then, for any $n \in \mathbb{N}$ and any $\mu > 0$, equation (3-1) possesses a $C^{n,\varepsilon}$-smooth extension $\widetilde{\mathcal{M}}_{N_n}$ of the initial IM $\mathcal{M}_{N_1}$ (where N_1 is the first N which satisfies the spectral gap condition (3-5) and $\varepsilon > 0$ is small enough) such that $\widetilde{\mathcal{M}}_{N_n}$ is μ-close to the IM $\mathcal{M}_{N_n}$ in the C^1_b-norm.

Proof for $n = 2$. Let N_1 be the first N for which the spectral gap condition (3-5) is satisfied with $L := \|F'\|_{C_b(H,\mathcal{L}(H,H))}$ and let the corresponding $\mathcal{M}_1$ be the $C^{1,\varepsilon}$-smooth IM which exists due to Theorem 3.2 and Corollary 3.8. Recall that for any $p \in H$, we have a solution $V(p,t)$ of problem (3-6) (where p is replaced by $\boldsymbol{P}_{N_1} p$) and its Fréchet derivative $V'_\xi(t) := V'(p,t)\xi$ in p satisfies the equation of variations

(3-21), such that both functions $V(p,\cdot)$ and $V'_\xi(\cdot)$ and belong to the space $L^2_{e^{\theta_1 t}}(\mathbb{R}_-, H)$ for any θ_1 satisfying (3-17), i.e., $\lambda_{N_1}+L<\theta_1<\lambda_{N_1}-L$. Moreover, for any other $p_1\in H$, we have the estimate

$$\|V(p_1,t)-V(p,t)-V'_\xi(t)\|_{L^2_{e^{\theta_1(1+\varepsilon)t}}(\mathbb{R}_-,H)}\le C\|\boldsymbol{P}_{N_1}(p-p_1)\|^{1+\varepsilon}, \tag{4-3}$$

where $\varepsilon>0$, $\xi:=p_1-p$ and C is independent of p and p_1.

Let now $N_2>N_1$ be the first N which satisfies

$$\lambda_{N_2+1}-\lambda_{N_2}-\lambda_{N_1}>3L \tag{4-4}$$

(such N exists due to condition (4-2)). Then, we have the corresponding $C^{1,\varepsilon}$-smooth IM $\mathcal{M}_{N_2}$. Let us denote by $W(p,t)$, $p\in H$, the corresponding solution of (3-6) (where N is replaced by N_2 and p is replaced by $\boldsymbol{P}_{N_2}p$). This solution belongs to $L^2_{e^{\theta_2 t}}(\mathbb{R}_-, H)$ with θ_2 satisfying (3-17) (with N replaced by N_2). Moreover, analogously to (4-3), we have

$$\|W(p_1,t)-W(p,t)-W'_\xi(t)\|_{L^2_{e^{\theta_2(1+\varepsilon)t}}(\mathbb{R}_-,H)}\le C\|\boldsymbol{P}_{N_2}(p-p_1)\|^{1+\varepsilon}, \tag{4-5}$$

where $W'_\xi(t)=W'(p,t)\xi$ solves (3-21) with N replaced by N_2. We also know that $V(p,t)=W(p,t)$ if $p\in\mathcal{M}_{N_1}$ and, therefore, due to (4-3) and (4-5),

$$\|V'(p,\cdot)\xi-W'(p,\cdot)\xi\|_{L^2_{e^{\theta_2(1+\varepsilon)t}}(\mathbb{R}_-,H)}\le C\|\boldsymbol{P}_{N_2}\xi\|^{1+\varepsilon},\quad \xi=p_1-p,\ \ p,p_1\in\mathcal{M}_{N_1}. \tag{4-6}$$

Let us define for every $p\in\mathcal{M}_{N_1}$ and every $\xi\in H$ the "second derivative" $W''_\xi=W''(p,t)[\xi,\xi]$ of the trajectory $u(t)=W(p,t)=V(p,t)$ as a solution of the problem

$$\partial_t W''_\xi+AW''_\xi-F'(V(p,t))W''_\xi=2F''(V(p,t))[V'_\xi,W'_\xi]-F''(V(p,t))[V'_\xi,V'_\xi],\quad \boldsymbol{P}_{N_2}W''_\xi|_{t=0}=0. \tag{4-7}$$

Note that the right-hand side of this equation belongs to the weighted space $L^2_{e^{(\theta_1+\theta_2)t}}(\mathbb{R}_-, H)$, where the exponents θ_1 and θ_2 satisfy assumption (3-17) with $N=N_1$ and $N=N_2$ respectively, i.e.,

$$\lambda_{N_1}+L<\theta_1<\lambda_{N_1}-L,\quad \lambda_{N_2}+L<\theta_2<\lambda_{N_2}-L.$$

Moreover, due to assumption (4-4), it is possible to fix θ_1 and θ_2 in such a way that the exponent $\theta_1+\theta_2$ still satisfies (3-17) with $N=N_2$. Thus, by Proposition 3.7, there exists a unique solution of (4-7) belonging to the space $L^2_{e^{(\theta_1+\theta_2)t}}(\mathbb{R}_-, H)$ and the function W''_ξ is well-defined and satisfies

$$\|W''_\xi\|_{C_{e^{(\theta_1+\theta_2)t}}(\mathbb{R}_-,H)}\le C\|W''_\xi\|_{L^2_{e^{(\theta_1+\theta_2)t}}(\mathbb{R}_-,H)}\le C^2\|\xi\|^2,$$

where C is independent of p.

Let us define the desired quadratic polynomial $\xi\to J^2_\xi W(p,t)$, $p\in\mathcal{M}_{N_1}$, as

$$J^2_\xi W(p,t):=V(p,t)+W'(p,t)\xi+\tfrac12 W''(p,t)[\xi,\xi],\quad \xi\in H. \tag{4-8}$$

We need to verify the compatibility conditions for these "Taylor jets" on $p\in\mathcal{M}_{N_1}$. It is straightforward to check using $F\in C^{2,\varepsilon}$, $V,W\in C^{1,\varepsilon}$ and Proposition 3.7 that

$$\|W''(p_1,\cdot)[\xi,\xi]-W''(p,\cdot)[\xi,\xi]\|_{L^2_{e^{(\theta_1+\theta_2+\varepsilon)t}}(\mathbb{R}_-,H)}\le C\|\xi\|^2\|p-p_1\|^{\varepsilon}$$

for $p, p_1 \in \mathcal{M}_{N_1}$. This gives us the desired compatibility condition for the second derivative; see (2-13) for $n = l = 2$.

Let us now verify the compatibility conditions for the first derivative ($l = 1$, $n = 2$ in (2-13)). To this end, we need to expand the difference $w(t) := W'(p_1, t)\xi - W'(p, t)\xi,\ p, p_1 \in \mathcal{M}_{N_1}$, in terms of $\delta = p - p_1$. By the definition of W', this function satisfies the equation

$$\begin{aligned}\partial_t w + Aw - F'(V(p,t))w &= (F'(V(p_1,t)) - F'(V(p,t)))W'(p_1,t)\xi \\ &= F''(V(p,t))[V'(p,t)\delta, W'(p,t)\xi] + h(t), \quad \boldsymbol{P}_{N_2} w|_{t=0} = 0, \end{aligned} \tag{4-9}$$

where the reminder h satisfies

$$\|h\|_{L^2_{e^{(\theta_1+\theta_2+\varepsilon)t}}(\mathbb{R}_-, H)} \le C\|\delta\|^{1+\varepsilon}\|\xi\|$$

for sufficiently small positive ε (this also follows from the fact that F is smooth and $V, W \in C^{1,\varepsilon}$). Thus, the remainder h in the right-hand side of (4-9) is of higher order in δ and, for this reason, is not essential, so we need to study the bilinear form (with respect to δ, ξ) in the right-hand side. Note that, in contrast to the case where the IM is C^2, this form is even not symmetric, so it should be corrected. Namely, we write the identity

$$\begin{aligned}&F''(V(p,t))[V'(p,t)\delta, W'(p,t)\xi] \\ &\quad = \big\{F''(V(p,t))[V'(p,t)\delta, W'(p,t)\xi] + F''(V(p,t))[V'(p,t)\xi, W'(p,t)\delta] \\ &\qquad\qquad - F''(V(p,t))[V'(p,t)\delta, V'(p,t)\xi]\big\} \\ &\qquad\qquad - F''(V(p,t))[V'(p,t)\xi, W'(p,t)\delta - V'(p,t)\delta] \end{aligned} \tag{4-10}$$

and note that the first term in the right-hand side is nothing more than the symmetric bilinear form which corresponds to the quadratic form

$$2F''(V(p,t))[V'(p,t)\xi, W'(p,t)\xi] - F''(V(p,t))[V'(p,t)\xi, V'(p,t)\xi]$$

used in (4-7) to define W'' and the second term is of order $\|\delta\|^{1+\varepsilon}\|\xi\|$ due to estimate (4-6) (where ξ is replaced by δ) and the growth rate of this term does not exceed $e^{-(\theta_1+\theta_2+\varepsilon)t}$ as $t \to -\infty$. Thus, by Proposition 3.7, we have

$$\|w - W''(p,\cdot)[\delta,\xi]\|_{L^2_{e^{(\theta_1+\theta_2+\varepsilon)t}}(\mathbb{R}_-, H)} \le C\|\delta\|^{1+\varepsilon}\|\xi\|$$

and the compatibility condition for $l = 1$ is verified.

Finally, let us check the zero-order compatibility condition ($l = 0$, $n = 2$ in (2-13)). Let

$$R(t) := V(p_1,t) - V(p,t) - W'(p,t)\delta - \frac{1}{2!}W''(p,t)[\delta,\delta].$$

Then, as elementary computations show, this function satisfies the equation

$$\begin{aligned}&\partial_t R + AR - F'(V(p,t))R \\ &\quad = \{F(V(p_1,t)) - F(V(p,t)) - F'(V(p,t))(V(p_1,t) - V(p,t))\} \\ &\qquad - \frac{1}{2!}\big(2F''(V(p,t))[V'(p,t)\delta, W'(p,t)\delta] \\ &\qquad\qquad - F''(V(p,t))[V'(p,t)\delta, V'(p,t)\delta]\big), \quad \boldsymbol{P}_{N_2}|_{t=0}R = 0. \end{aligned} \tag{4-11}$$

Since $F \in C^{2,\varepsilon}$ and $V \in C^{1,\varepsilon}$, the first term in the right-hand side equals

$$\frac{1}{2!}F''(V(p,t))[V'(p,t)\delta, V'(p,t)\delta] \tag{4-12}$$

up to the controllable in $L^2_{e^{(\theta_1+\theta_2+\varepsilon)t}}(\mathbb{R}_-, H)$-norm remainder of order $\|\delta\|^{2+\varepsilon}$. The second term can be simplified using (4-6) and also equals (4-12) up to higher-order terms. Thus, the right-hand side of (4-11) vanishes up to terms of order $\|\delta\|^{2+\varepsilon}$ and Proposition 3.7 gives us that

$$\|R\|_{L^2_{e^{(\theta_1+\theta_2+\varepsilon)t}}(\mathbb{R}_-, H)} \le C\|\delta\|^{2+\varepsilon} \tag{4-13}$$

for some positive ε. This finishes the verification of the compatibility conditions.

We are now ready to use the Whitney extension theorem. To this end, we first recall that the IM $\mathcal{M}_{N_2}$ is a graph of the $C^{1,\varepsilon}$-function $M_{N_2} : \boldsymbol{P}_{N_2}H \to \boldsymbol{Q}_{N_2}H$, which is defined via $M_{N_2}(p) := \boldsymbol{Q}_{N_2}W(p,0)$, $p \in \boldsymbol{P}_{N_2}H = H_{N_2}$ (all functions V, W, W', W'' defined above depend only on $\boldsymbol{P}_{N_2}$-component of $p \in H$, so without loss of generality we may assume that $p, \xi, \delta \in H_{N_2}$ (we took them from H in order to simplify the notation only). Thus, projecting the constructed Taylor jets to $t = 0$ and $\boldsymbol{Q}_{N_2}H$, we get the $C^{1,\varepsilon}$-function $M_{N_2}(p)$ restricted to the invariant set $p \in \boldsymbol{P}_{N_2}\mathcal{M}_{N_1}$ and a family of quadratic polynomials

$$J^2_\xi M_{N_2}(p) := \boldsymbol{Q}_{N_2} J^2_\xi W(p,0),$$

which satisfy the compatibility conditions on $p \in \boldsymbol{P}_{N_2}\mathcal{M}_{N_1}$. Therefore, since H_{N_2} is finite-dimensional, the Whitney extension theorem gives the existence of a $C^{2,\varepsilon}$-function $\widehat{M}_{N_2} : \boldsymbol{P}_{N_2}H \to \boldsymbol{Q}_{N_2}H$ such that

$$J^2_\xi \widehat{M}_{N_2}(p) = J^2_\xi M_{N_2}(p), \quad p \in \boldsymbol{P}_{N_2}\mathcal{M}_{N_1}.$$

Thus, the desired $C^{2+\varepsilon}$-extension of the IM $\mathcal{M}_{N_1}$ is "almost" constructed. It only remains to take care of the closeness in the C^1-norm. To this end, for any small $\nu > 0$, we introduce a cut-off function $\rho_\nu \in C^\infty(H_{N_2}, \mathbb{R})$ such that $\rho(p) \equiv 0$ if p belongs to the ν-neighbourhood $\mathcal{O}_\nu$ of $\boldsymbol{P}_{N_2}\mathcal{M}_{N_1}$ and $\rho(p) \equiv 1$ if $p \notin \mathcal{O}_{2\nu}$. Moreover, since $\boldsymbol{P}_{N_2}\mathcal{M}_{N_1}$ is $C^{1,\varepsilon}$-smooth, we may require also that

$$|\nabla_p \rho(p)| \le C\nu^{-1}, \tag{4-14}$$

where the constant C is independent of ν. Finally, we define

$$\widetilde{M}_{N_2}(p) := (1-\rho_\nu(p))\widehat{M}_{N_2}(p) + \rho_\nu(p)(\mathbb{S}_{\nu^2}M_{N_2})(p), \tag{4-15}$$

where $\mathbb{S}_\mu$ is a standard mollifying operator,

$$(\mathbb{S}_\mu f)(p) := \int_{\mathbb{R}^{N_2}} \beta_\mu(p-q) f(q)\, dq,$$

and the kernel $\beta_\mu(p)$ satisfies $\beta_\mu(p) = (1/\mu^{N_2})\beta_1(p/\mu)$ and $\beta_1(p)$ is a smooth, nonnegative function with compact support satisfying $\int_{\mathbb{R}^{N_2}} \beta_1(p)\, dp = 1$.

We claim that $\widetilde{M}_{N_2}$ is a desired extension. Indeed, $\widetilde{M}_{N_2}(p) \equiv \widehat{M}_{N_2}(p)$ in $\mathcal{O}_\nu$ and therefore $\widetilde{M}_{N_2}$ and M_{N_2} coincide on $\boldsymbol{P}_{N_2}\mathcal{M}_{N_1}$. Obviously, $\widetilde{M}_{N_2}$ is $C^{2,\varepsilon}$-smooth. To verify closeness, we note that

$$\widetilde{M}_{N_2}(p) - M_{N_2}(p) = (1-\rho_\nu(p))(\widehat{M}_{N_2}(p) - M_{N_2}(p)) + \rho_\nu(p)((\mathbb{S}_{\nu^2}M_{N_2})(p) - M_{N_2}(p)). \tag{4-16}$$

Using the fact that $M_{N_2} \in C^{1,\varepsilon}$ together with the standard estimates for the mollifying operator, we get

$$\|(\mathbb{S}_{\nu^2} M_{N_2})(p) - M_{N_2}(p)\| \le C\nu^2, \quad \|\nabla_p(\mathbb{S}_{\nu^2} M_{N_2})(p) - \nabla_p M_{N_2}(p)\| \le C\nu^{2\varepsilon},$$

which together with (4-14) shows that the C^1-norm of the second term in the right-hand side of (4-16) is of order $\nu^{2\varepsilon}$. To estimate the first term, we use that both functions $\widehat{M}_{N_2}(p)$ and $M_{N_2}(p)$ are at least $C^{1,\varepsilon}$-smooth and

$$\widehat{M}_{N_2}(p) = M_{N_2}(p), \quad \nabla_p \widehat{M}_{N_2}(p) = \nabla_p M_{N_2}(p), \quad p \in \boldsymbol{P}_{N_2}\mathcal{M}_{N_1}.$$

For this reason,

$$\|\widehat{M}_{N_2}(p) - M_{N_2}(p)\| \le C\nu^{1+\varepsilon}, \quad \|\nabla_p \widehat{M}_{N_2}(p) - \nabla_p M_{N_2}(p)\| \le C\nu^{\varepsilon}$$

for all $p \in \mathcal{O}_{2\nu}$. Thus, using (4-14) again, we see that

$$\|\widetilde{M}_{N_2}(\cdot) - M_{N_2}(\cdot)\|_{C_b^1(H_{N_2},H)} \le C\nu^{\varepsilon}.$$

This finishes the proof of the theorem for the case $n = 2$. □

Proof for general $n \in \mathbb{N}$. We will proceed by induction with respect to n. Assume that, for some $n \in \mathbb{N}$, we have already constructed the $C^{1,\varepsilon}$-smooth inertial manifold $\mathcal{M}_{N_n}$ which is a graph of a map $M_{N_n} : \boldsymbol{P}_{N_n} H \to \boldsymbol{Q}_{N_n} H$ and this map is constructed via the solution $V(p,t),\ t \le 0,\ p \in H$, of the backward problem (3-6), where N is replaced by N_n. Recall that this manifold is constructed using the n-th spectral gap. Assume also that, for every $p \in \boldsymbol{P}_{N_n}\mathcal{M}_{N_1}$, we have already constructed the n-th Taylor jet $J_\xi^n V(p,t)$ such that the compatibility conditions up to order n are satisfied. In contrast to the proof for the case $n = 2$, it is convenient for us to write these conditions in the form of (2-12):

$$\|J_\xi^n V(p_1,\cdot) - J_{\xi+\delta}^n V(p,\cdot)\|_{L^2_{e^{(\theta_n+(n-1)\theta_{n-1}+\varepsilon)t}}(\mathbb{R}_-,H)} \le C(\|\delta\| + \|\xi\|)^{n+\varepsilon}. \tag{4-17}$$

Here $\xi \in H$ is arbitrary, $\delta := p_1 - p,\ \varepsilon > 0$ and $\theta_1 < \theta_2 \cdots < \theta_n$ are the exponents which satisfy condition (3-17) for $N = N_1, \dots, N_n$. In order to simplify the notation, we will write below

$$J_\xi^n V(p_1) - J_{\xi+\delta}^n V(p) = O_{\theta_n+(n-1)\theta_{n-1}+\varepsilon}((\|\delta\| + \|\xi\|)^{n+\varepsilon}) \tag{4-18}$$

instead of (4-17) and likewise in similar situations. Rewriting (4-18) in terms of truncated jets with the help of (2-10) (where ξ is replaced by δ), we have

$$j_\xi^n V(p_1) + j_\delta^n V(p) - j_{\xi+\delta}^n V(p) = O_{n\theta_n+\varepsilon}((\|\delta\| + \|\xi\|)^{n+\varepsilon}), \tag{4-19}$$

where we have used that $\theta_{n-1} < \theta_n$. We also need the induction assumption that (4-19) holds for every $m \le n$, namely,

$$J_\xi^m V(p_1) - J_{\xi+\delta}^m V(p) = O_{m\theta_n+\varepsilon}((\|\delta\| + \|\xi\|)^{m+\varepsilon}). \tag{4-20}$$

Let us now consider the $(n+1)$-th spectral gap at $N = N_{n+1}$ which is the first N satisfying

$$\lambda_{N_{n+1}} + L + n(\lambda_{N_n+1} - L) < \lambda_{N_{n+1}+1} - L. \tag{4-21}$$

Let $\mathcal{M}_{N_{n+1}}$ be the corresponding IM which is generated by the backward solution $W(p,t)$ of problem (3-6) with N replaced by N_{n+1}. We need to define the $(n+1)$-th Taylor jet $J_\xi^{n+1}W(p,t)$ for the function $W(p,t)$,

$$J_\xi^{n+1}W(p,t) = W(p,t) + \sum_{k=1}^{n+1} \frac{1}{k!} W^{(k)}(p,t)[\{\xi\}^k], \tag{4-22}$$

$\xi \in H$ and $p \in \boldsymbol{P}_{N_{n+1}}\mathcal{M}_{N_1}$ and to verify the compatibility conditions of order $n+1$. Keeping in mind the already-considered cases $n=1$ and $n=2$, we introduce the required jet (4-22) as a backward solution of the equation

$$\partial_t J_\xi^{n+1}W(p) + AJ_\xi^{n+1}W(p) = F^{[n+1]}(p,\xi), \quad \boldsymbol{P}_{N_{n+1}} J_\xi^{n+1}(p)|_{t=0} = \boldsymbol{P}_{N_{n+1}}(p+\xi), \tag{4-23}$$

where

$$\begin{aligned} F^{[n+1]}(p,\xi,t) := {} & F(W(p,t)) + F'(W(p,t))j_\xi^{n+1}W(p,t) \\ & + \sum_{k=2}^{n+1} \frac{1}{k!}\big(kF^{(k)}(W(p,t))[\{j_\xi^n V(p,t)\}^{k-1}, j_\xi^n W(p,t)] \\ & \qquad - (k-1)F^{(k)}(W(p,t))[\{j_\xi^n V(p,t)\}^k]\big). \end{aligned} \tag{4-24}$$

The symbol "$[n+1]$" means that we have dropped out all terms of order greater than $n+1$ from the right-hand side, so $F^{[n+1]}$ is a polynomial of order $n+1$ in $\xi \in H$. Alternatively, the dropping out procedure means that we use the substitution

$$\{j_\xi^n V(p)\}^k \to \sum_{\substack{n_1+\cdots+n_k \le n+1 \\ n_i \in \mathbb{N}}} B_{n_1,\dots,n_k}\{j_\xi^{n_1}V(p),\dots,j_\xi^{n_k}V(p)\}, \tag{4-25}$$

where the numbers $B_{n_1,\dots,n_k} \in \mathbb{R}$ are chosen in such a way that polynomials in the left- and right-hand sides of (4-25) coincide up to order $\{\xi\}^{n+1}$ inclusively and the term $[\{j_\xi^n V(p,t)\}^{k-1}, j_\xi^n W(p,t)]$ is treated analogously. The explicit expressions for these coefficients can be found using the formulas for the higher-order chain rule (Faà di Bruno-type formulas; see, e.g., [Roman 1980; Hájek and Johanis 2014]), but these expressions are lengthy and not essential for what follows, so we omit them.

Note also that the truncated jets $j_\xi^n V(p,t)$ are taken from the induction assumption. We seek the solution of (4-23) belonging to $L^2_{e^{n\theta_n+\theta_{n+1}}}(\mathbb{R}_-, H)$ for some θ_{n+1} satisfying (3-17) with N replaced by N_{n+1}. Expanding (4-24) in series with respect to ξ, we get the recurrent equations for finding the "derivatives" $W_\xi^{(k)}(p,t) := W^{(k)}(p,t)[\{\xi\}^k]$:

$$\partial_t W_\xi^{(k)} + AW_\xi^{(k)} - F'(W(p))W_\xi^{(k)} = \Phi(j_\xi^{k-1}W, j_\xi^{k-1}V), \quad \boldsymbol{P}_{N_{n+1}} W_\xi^{(k)}|_{t=0} = 0, \tag{4-26}$$

for $k \ge 2$, where Φ is polynomial of order k in ξ which does not contain $W_\xi^{(l)}$, with $l \ge k$. Thus, the functions $W_\xi^{(k)}$ can be, indeed, found recursively. Moreover, the spectral gap assumption (4-21) guarantees that we can find θ_{n+1} satisfying (3-17) with $N = N_{n+1}$ such that $\theta_{n+1} + n\theta_n$ also satisfies this condition. Therefore, Proposition 3.7 guarantees the existence and uniqueness of the homogeneous polynomials $W_\xi^{(k)}(p)$ satisfying

$$\|W_\xi^{(k)}(p)\|_{L^2_{e^{(\theta_{n+1}+k\theta_n)t}}(\mathbb{R}_-,H)} \le C\|\xi\|^k \tag{4-27}$$

for $k = 1,\dots,n+1$.

To complete the proof of the theorem, we only need to verify that the jet $J_\xi W(p,t)$ satisfies the compatibility conditions of order $n+1$. If this is verified, the rest of the proof coincides with the one given above for the case $n=2$. We postpone this verification till the next section. Thus, the theorem is proved by modulo of compatibility conditions. □

Corollary 4.4. *Let the assumptions of Theorem 4.3 hold with $\mu > 0$ being small enough. Then the invariant manifold $\boldsymbol{P}_{N_n}\mathcal{M}_{N_1}$ of the extended IF (4-1) possesses an exponential tracking property in H_{N_n}; i.e., for every solution $u_{N_n}(t)$ of (4-1) there exists the corresponding solution $\bar{u}_{N_n}$ belonging to this manifold such that*

$$\|u_{N_n}(t) - \bar{u}_{N_n}(t)\| \le Ce^{-\theta_1 t} \tag{4-28}$$

for some positive C and θ_1.

Proof. As we have already mentioned, this is the standard corollary of the fact that $\mathcal{M}_{N_1}$ is normally hyperbolic and, therefore, persists under small C^1-perturbations; see [Bates et al. 1999; Fenichel 1972; Hirsch et al. 1977; Katok and Hasselblatt 1995]. Nevertheless, for the convenience of the reader, we now sketch a direct proof that does not use the normal hyperbolicity explicitly.

We first construct an invariant manifold $\overline{\mathcal{M}}_{N_1}$ with the base H_{N_1} in H_{N_n} for the extended IF. We do this exactly as in the proof of Theorem 3.2 by solving the backward problem

$$\partial_t u_{N_n} + Au_{N_n} - \boldsymbol{P}_{N_n} F(u_{N_n} + \widetilde{M}_{N_n}(u_{N_n})) = 0, \quad \boldsymbol{P}_{N_1} u_{N_n} = p \tag{4-29}$$

in the space $L^2_{e^{\theta t}}(\mathbb{R}_-, H_{N_n})$ with $\theta = (\lambda_{N_1} + \lambda_{N_1+1})/2$. This equation is $(C\mu)$-closed to

$$\partial_t \bar{u}_{N_n} + A\bar{u}_{N_n} - \boldsymbol{P}_{N_n} F(\bar{u}_{N_n} + M_{N_n}(\bar{u}_{N_n})) = 0, \quad \boldsymbol{P}_{N_1} \bar{u}_{N_n} = p \tag{4-30}$$

in the C^1-norm (since $\widetilde{M}_{N_n}$ is μ-closed to M_{N_n} due to Theorem 4.3). Thus, using Remark 3.13 and the Banach contraction theorem, we can construct a unique solution $u_{N_n}(t)$ of (4-29) in the $(C\mu)$-neighbourhood of the corresponding solution $\bar{u}_{N_n}$ of problem (4-30) and vice versa. This gives us the existence of the manifold $\overline{\mathcal{M}}_{N_1}$ which is generated by all backward solutions of (4-30) belonging to the space $L^2_{e^{\theta t}}(\mathbb{R}_-, H_{N_n})$. Since the solutions belonging to the invariant manifold $\boldsymbol{P}_{N_n}\mathcal{M}_{N_1}$ satisfy exactly the same property, we conclude that $\overline{\mathcal{M}}_{N_1} = \boldsymbol{P}_{N_n}\mathcal{M}_{N_1}$.

It remains to verify that the manifold $\overline{\mathcal{M}}_{N_1}$ possesses an exponential tracking property. This can be done as in the proof of Theorem 3.2 by considering the analogue of (3-14) for system (4-1) and using again that $\widetilde{M}_{N_n}$ is close to M_{N_n} in the C^1-norm. This finishes the proof of the corollary. □

Corollary 4.5. *Arguing as in Corollary 3.9, we check that the extended IM $\widetilde{\mathcal{M}}_{N_n}$ is also a $C^{n,\varepsilon}$-submanifold of $H^2 := D(A)$.*

5. Examples and concluding remarks

We now give several examples of our main theorem, as well as its reinterpretations, and state some interesting problems for further study. We start with the application to the 1-dimensional reaction-diffusion equation.

Example 5.1. Let us consider the following reaction-diffusion system in a 1-dimensional domain $\Omega = (-\pi, \pi)$:

$$\partial_t u = a\partial_x^2 u - f(u), \quad u|_\Omega = 0, \quad u|_{t=0} = u_0, \tag{5-1}$$

where u is an unknown function, $a > 0$ is a given viscosity parameter, and $f(u)$ is a given smooth function satisfying $f(0) = 0$ and some dissipativity conditions, for instance,

$$f(u)u \geq -C + \alpha|u|^2, \quad u \in \mathbb{R}.$$

for some C and $\alpha > 0$ (e.g., $f(u) = u^3 - u$ as in the case of real Ginzburg–Landau equation). Then, due to the maximum principle, we have the following dissipative estimate for the solutions of (5-1):

$$\|u(t)\|_{L^\infty} \leq \|u_0\|_{L^\infty} e^{-\alpha t} + C_*, \tag{5-2}$$

where the constant C_* is independent of u_0; see, e.g., [Babin and Vishik 1992; Chepyzhov and Vishik 2002; Temam 1988]. Thus, the associated solution semigroup $S(t)$ acting in the phase space $H := H_0^1(\Omega)$ possesses an absorbing set in $C(\bar{\Omega})$, and cutting-off the nonlinearity outside of this ball, we may assume without loss of generality that $f \in C_0^\infty(\mathbb{R})$.

After this transformation, (5-1) can be considered as an abstract parabolic equation (3-1) in the Sobolev space $H = H_0^1(\Omega)$. Since this space is an algebra with respect to pointwise multiplication (since we have only one spatial variable), the corresponding nonlinearity $F(u)(x) := f(u(x))$ is C^∞-smooth and all its derivatives are globally bounded.

Finally, the linear operator A in this example is $A = -a\partial_x^2$ endowed with the Dirichlet boundary conditions. Obviously, this operator is self-adjoint, positive definite and its inverse is compact. Moreover, its eigenvalues

$$\lambda_k = ak^2, \quad k \in \mathbb{N},$$

satisfy (4-2). Thus, our main Theorem 4.3 is applicable here and, therefore, problem (5-1) possesses an IM $\mathcal{M}_{N_1}$ of smoothness $C^{1,\varepsilon}$ for some $\varepsilon > 0$ and, for every $n \in \mathbb{N}$, this IM can be extended to a manifold $\widetilde{\mathcal{M}}_{N_n}$ of regularity C^{n,ε_n}, $\varepsilon_n > 0$, in the sense of Definition 4.1.

Remark 5.2. Our general theorem is applicable not only for a scalar reaction-diffusion equation (5-1), but also for systems where the analogue of (5-2) is known, for instance, for the case of 1-dimensional complex Ginzburg–Landau equation. However, one should be careful in the case where the diffusion matrix is not self-adjoint and especially when it contains nontrivial Jordan cells. In this case, even Lipschitz IM may not exist; see [Kostianko and Zelik 2022] for more details.

A bit unusual choice of the phase space $H = H_0^1(\Omega)$ (instead of the natural one $H = L^2(\Omega)$) is related to the fact that we need H to be an algebra in order to define Taylor jets for the nonlinearity F and to verify that it is C^∞. This, however, may be relaxed in applications since backward solutions of (3-4) and (3-18) are usually smooth in space and time if the nonlinearity f is smooth, so the Taylor jets for $V(p, t)$ will be well-defined even if we consider $L^2(\Omega)$ as a phase space and the theory works with minimal changes. This observation may be useful if we want to remove the assumption $f(0) = 0$ in (5-1), but in order to avoid technicalities, we prefer not to go further in this direction here.

The restriction to the 1-dimensional case is motivated by the fact that the spectral gap condition (4-2) is naturally satisfied by the Laplacian in 1-dimensional case only (it is an open problem already in the 2-dimensional case).

If we consider higher-order operators, say bi-Laplacian then the analogous result holds also in three dimensions. The typical example here is given by the Swift–Hohenberg equation in a bounded domain $\Omega \subset \mathbb{R}^3$:

$$\partial_t u = -(\Delta + 1)^2 u + u - u^2, \quad u|_{\partial\Omega} = \Delta u|_{\partial\Omega} = 0,$$

where the spectral gap condition (4-2) is also satisfied; see [Zelik 2014]. We also note that although our main theorem is stated and proved for the case where F maps H to H, it can be generalized in a very straightforward way to the case where the operator F decreases smoothness and maps H to $H^{-s} := D(A^{-s/2})$ for some $s \in (0, 2)$. The spectral gap assumption (4-2) should be replaced by

$$\limsup_{n\to\infty} \left\{ \frac{\lambda_{n+1} - \lambda_n}{\lambda_{n+1}^{s/2} + \lambda_n^{s/2}} \right\} = \infty.$$

After this extension, our theorem becomes applicable to equations which contain spatial derivatives in the nonlinearity. A typical example of such applications is the 1-dimensional Kuramoto–Sivashinsky equation

$$\partial_t u + a\partial_x^2 u + \partial_x^4 u + u\partial_x u = 0, \quad \Omega = (-\pi, \pi), \quad a > 0,$$

endowed with Dirichlet or periodic boundary conditions; see [Zelik 2014] for more details.

Remark 5.3. As we mentioned in the Introduction, there is some significant recent progress in constructing IMs for concrete classes of parabolic equations which do not satisfy the spectral gap conditions (such as scalar reaction-diffusion equations in higher dimensions, 3-dimensional Cahn–Hilliard or complex Ginzburg–Landau equations, various modifications of Navier–Stokes systems, 1-dimensional reaction-diffusion-advection systems, etc.). The techniques developed in the present paper are not directly applicable to such problems (in particular, our technique is strongly based on the Perron method of constructing the IMs and it is not clear how to use the Perron method here since we do not have the so-called absolute normal hyperbolicity in the most part of equations mentioned above; see [Kostianko 2018; Kostianko and Zelik 2015] for more details). However, we believe that the proper modification of our method would allow us to cover these cases as well. We return to this problem elsewhere.

We now give an alternative (probably more transparent and more elegant) formulation of Theorem 4.3. We recall that in Theorem 4.3, we have directly constructed a smooth extended IF (4-1) for the initial equation (3-1). This extended IF captures all nontrivial dynamics of (3-1), but the associated smooth extended IM $\mathcal{M}_n$ is not associated with the "true" IM of any system of the form (3-1). This drawback can be easily corrected in a more or less standard way which leads to the following reformulation of our main result.

Corollary 5.4. *Let the assumptions of Theorem 4.3 be satisfied and let $\mathcal{M}_{N_1}$ be the C^{1,ε_1}-smooth IM of (3-1) which corresponds to the first spectral gap. Then, for every $n \in \mathbb{N}$, $n > 1$, there exists a modified nonlinearity $\widetilde{F} : H \to H$ which belongs to $C_b^{n-1,\varepsilon_n}(H, H)$ for some $\varepsilon_n > 0$ such that:*

(1) *The initial IM $\mathcal{M}_{N_1}$ is simultaneously an IM for the modified equation*

$$\partial_t u + Au = \widetilde{F}_n(u). \tag{5-3}$$

(2) *Equation (5-3) possesses a C^{n,ε_n}-smooth IM $\widetilde{\mathcal{M}}_{N_n}$ of dimension N_n such that the initial IM $\mathcal{M}_1$ is a normally hyperbolic globally stable submanifold of $\widetilde{\mathcal{M}}_{N_n}$.*

(3) *The nonlinearity $\widetilde{F}_n(u)$ depends on the variable $u_{N_n} := \boldsymbol{P}_{N_n} u$ only and the IF associated with the IM $\widetilde{\mathcal{M}}_{N_n}$ is given by (4-1) where K_2 is replaced by N_n.*

Proof. Indeed, we take the manifold $\widetilde{\mathcal{M}}_{N_n}$ constructed in Theorem 4.3 and define the desired function $\widetilde{F}_n$ as

$$\boldsymbol{P}_{N_n}\widetilde{F}_n(u) := \boldsymbol{P}_{N_n} F(u_{N_n} + \widetilde{M}_{N_n}(u_{N_n})) \tag{5-4}$$

and

$$\boldsymbol{Q}_{N_n}\widetilde{F}_n(u) := \widetilde{M}'_{N_n}(u_{N_n})[-A\widetilde{M}_{N_n}(u_{N_n}) + \boldsymbol{P}_{N_n} F(u_{N_n} + \widetilde{M}_{N_n}(u_{N_n}))] + A\widetilde{M}_{N_n}(u_{N_n}). \tag{5-5}$$

Then, due to the choice of $\boldsymbol{P}_{N_n}$-component of $\widetilde{F}_n(u)$, the equation for u_{N_n} is decoupled from the equation for the $\boldsymbol{Q}_{N_n}$-component and coincides with the extended IF for (5-3) constructed in Theorem 4.3. On the other hand, the $\boldsymbol{Q}_{N_n}$-component of $\widetilde{F}_n$ is chosen in a form which guarantees that $\widetilde{\mathcal{M}}_{N_n}$ is an invariant manifold for (5-3). Moreover, if $u(t)$ solves (5-3) with such a nonlinearity and $v(t) := u(t) - \boldsymbol{P}_{N_n} u(t) - \widetilde{M}_{N_n}(u_{N_n}(t))$, then this function satisfies

$$\partial_t v + Av = 0, \qquad \boldsymbol{P}_{N_n} v(t) \equiv 0,$$

and, therefore,

$$\|v(t)\|_H \le \|v(0)\|_H e^{-\lambda_{N_n+1} t}.$$

Thus, $\widetilde{M}_{N_n}$ is indeed an IM for problem (5-3) and we only need to check the regularity of the modified function $\widetilde{F}_n$.

The $\boldsymbol{P}_{N_n}$ component (5-4) is clearly C^{n,ε_n}-smooth, but the situation with the $\boldsymbol{Q}_{N_n}$ is a bit more delicate due to the presence of terms $A\widetilde{M}_{N_n}(u_{N_n})$ and $\widetilde{M}'_{N_n}(u_{N_n})$. The first term is not dangerous since we know that $\widetilde{M}_{N_n}$ is C^{n,ε_n}-smooth as the map from H_{N_n} to H^2. The second term is worse and decreases the smoothness of the $\widetilde{F}_n$ till C^{n-1,ε_n}. Thus, the corollary is proved. □

Remark 5.5. The modified nonlinearity $\widetilde{F}_n(u)$ can be interpreted as a "clever" cut-off of the initial nonlinearity $F(u)$ outside of the global attractor (even outside of the IM of minimal dimension). In this sense we may say that all obstacles for the existence of $C^{n,\varepsilon}$-smooth IMs can be removed by appropriately cutting off the nonlinearity outside of the global attractor, which does not affect the dynamics of the initial problem. This demonstrates the importance of finding the proper cut off procedure in the theory of IMs.

Example 5.6. We now return to the model example of G. Sell introduced in Example 3.11 and show how the problem of smoothness of an invariant manifold can be resolved. Since the nonlinearity for this system is not *globally* Lipschitz continuous, the above-developed theory is formally not applicable and we need to cut-off the nonlinearity first. We overcome this problem by considering only *local* manifolds in a small neighbourhood of the origin.

Indeed, it is not difficult to see that system (3-31) has an explicit particular solution

$$u_1(t) = \pm e^{-t}, \quad u_{n+1}(t) = C_n e^{-2^n t} t^{2^n-1}, \quad n > 1,$$

where the coefficients C_n satisfy the recurrence relation

$$C_{n+1} = \frac{1}{2^n - 1} C_n^2, \quad C_0 = 1.$$

This solution determines a 1-dimensional local invariant manifold

$$\mathcal{M}_1 = \{p + M(p) : p \in H_1 = \mathbb{R},\ |p| < \beta\},$$

where $M : \mathbb{R} \to H$ is defined by $M = (0, M_1(p), M_2(p), \dots)$ and

$$M_{n+1}(p) = C_n p^{2^n} \left(\ln \frac{1}{|p|} \right)^{2^n-1}, \quad n \in \mathbb{N},$$

which is a 1-dimensional IM for system (3-31) and β is a sufficiently small positive number. Indeed, since $C_n \le 2^{-\alpha 2^n}$ for some positive α, this manifold is well-defined as a local submanifold of $H = l_2$ (if $\beta > 0$ is small enough) and is $C^{1,\varepsilon}$-smooth for any $\varepsilon \in (0, 1)$. Moreover, we see that $M_2(p)$ is only $C^{1,\varepsilon}$-smooth and higher components are more regular; in particular, $M_n(p)$ is $C^{2^{n-1}-1,\varepsilon}$-smooth. This shows us how to define the extended manifolds of an arbitrary finite smoothness. Namely, let us fix some $n \in \mathbb{N}$ and consider the manifold

$$\widetilde{\mathcal{M}}_n := \{p + \widetilde{M}_n(p) : p \in H_n,\ |p_1| < \beta\}, \quad \widetilde{M}_n(p) := (\{0\}^n, M_{n+1}(p_1), M_{n+2}(p_1), M_{n+3}(p_1), \dots). \tag{5-6}$$

Clearly $\widetilde{\mathcal{M}}_n$ is $C^{2^n-1,\varepsilon}$-smooth and $\mathcal{M}_1$ is a submanifold of $\widetilde{\mathcal{M}}_n$. Moreover, if we define the modified nonlinearity $\widetilde{F}_n(u)$ as

$$\widetilde{F}_n(u) = (0, u_1^2, u_2^2, \dots, u_{n-1}^2, M_{n+1}(u_1), M_{n+2}(u_1), \dots), \tag{5-7}$$

then it will be $C^{2^n-1,\varepsilon}$-smooth and the extended manifold $\widetilde{\mathcal{M}}_n$ will be an IM for the corresponding modified equation (5-3). Finally, the normal hyperbolicity of $\mathcal{M}_1$ in $\widetilde{\mathcal{M}}_n$ follows from the fact that any solution on $\mathcal{M}_1$ decays to zero no faster than e^{-t} due to the nonzero first component, if we look to the transversal directions, the smallest decay rate is determined by the second component and this decay is at least as $t^3 e^{-2t}$. Since our model system is explicitly solvable, we leave verifying this normal hyperbolicity to the reader. We also note that the extended IF in this case reads

$$\frac{d}{dt} u_1 + u_1 = 0, \quad \frac{d}{dt} u_k + 2^{k-1} u_k = u_{k-1}^2, \quad k = 2, \dots, n,$$

which is nothing more than the Galerkin approximation system to (3-31).

Remark 5.7. We see that, in the toy example of (3-31), we can find the desired extension of the initial IM explicitly without using the Whitney extension theorem (and even without assuming the global boundedness of F and its derivatives). Moreover, the dependence of smoothness of the extended IM on its dimension is very nice; namely, if we want to have a C^n-smooth IM, it is enough to take $\dim \widetilde{M} \sim \log_2 n$. Of course, this is partially related to good exponentially growing spectral gaps, but the main reason is

that we have an extra regularity property for the initial IM, namely, that the smoothness of projections $\boldsymbol{Q}_k M(p)$ grows with k. Unfortunately, this is not true in a more or less general case, which makes the extension construction much more involved. In particular, we do not know how to gain more than one unit of smoothness from one spectral gap and have to use n different spectral gaps to get n units of smoothness. This, in turn, leads to extremely fast growth of the dimension of the manifold with respect to the regularity (as not difficult to see, in Example 5.1, the dimension of $\widetilde{\mathcal{M}}_{N_n}$ grows as a double exponent with respect to n).

We believe that this problem is technical and the estimates for the dimension can be essentially improved. Indeed, if we would be able to get n units of extra regularity using one extra (sufficiently large) gap the above-mentioned growth of the dimension would become linear in n in Example 5.1. We expect that this linear growth is optimal, and we are even able to construct the corresponding Taylor jets. But these jets do not satisfy the compatibility conditions and we do not know how to correct them properly.

Appendix: Verifying the compatibility conditions

The aim of this appendix is to show that the jets $J_\xi^{n+1} W(p,t),\ p \in \boldsymbol{P}_{N_{n+1}} H$, constructed via (4-23), satisfy the compatibility conditions up to order $n+1$ and, thus, to complete the proof of Theorem 4.3. We will proceed by induction with respect to the order $m \le n+1$.

Indeed, the first-order compatibility conditions are trivially satisfied since the functions $W(p,t)$ are $C^{1,\varepsilon}$-smooth. Assume that the m-th order conditions are satisfied for some $m \le n+1$, and for all $m_1 \le m$

$$J_\xi^{m_1} W(p_1) - J_{\delta+\xi}^{m_1} W(p) = O_{\theta_{n+1}+(m_1-1)\theta_n}((\|\delta\| + \|\xi\|)^{m_1+\varepsilon}) \tag{A-1}$$

for all $\xi \in H,\ p_1, p \in \boldsymbol{P}_{N_{n+1}} \mathcal{M}_{N_1},\ \varepsilon > 0,\ \delta := p_1 - p$ and some constant C which is independent of p, p_1. Using the fact that $V(p,t) = W(p,t)$ for all $p \in \boldsymbol{P}_{N_{n+1}} \mathcal{M}_{N_1}$ together with the analogue of (A-1) for the already constructed jets $J_\xi^m V(p,t)$, we end up with

$$\begin{aligned} V(p_1) = W(p_1) &= V(p) + j_\delta^{m_1} V(p) + O_{m_1\theta_n+\varepsilon}(\|\delta\|^{m_1+\varepsilon}) \\ &= W(p) + j_\delta^{m_1} W(p) + O_{\theta_{n+1}+(m_1-1)\theta_n+\varepsilon}(\|\delta\|^{m_1+\varepsilon}) \end{aligned} \tag{A-2}$$

for all $p_1, p \in \boldsymbol{P}_{N_{n+1}} \mathcal{M}_{N_1},\ \delta := p_1 - p$ and, therefore $v(t) := V(p_1,t) - V(p,t)$ satisfies

$$\begin{gathered} v = j_\delta^{m_1} V(p) + O_{m_1\theta_n+\varepsilon}(\|\delta\|^{m_1+\varepsilon}) = j_\delta^{m_1} W(p) + O_{\theta_{n+1}+(m_1-1)\theta_n+\varepsilon}(\|\delta\|^{m_1+\varepsilon}), \\ j_\delta^{m_1} V(p) - j_\delta^{m_1} W(p) = O_{\theta_{n+1}+(m_1-1)\theta_n+\varepsilon}(\|\delta\|^{m_1+\varepsilon}). \end{gathered} \tag{A-3}$$

We now turn to the $(m+1)$-th jets and start with the following lemma which gives the compatibility conditions in the particular case $\xi = 0$.

Lemma A.1. *Let the above assumptions hold. Then*

$$v = W(p_1) - W(p) = j_\delta^{m+1} W(p) + O_{\theta_{n+1}+m\theta_n+\varepsilon}(\|\delta\|^{m+1+\varepsilon}) \tag{A-4}$$

for all $p_1, p \in \boldsymbol{P}_{N_{n+1}} \mathcal{M}_{N_1}$ and $\delta := p_1 - p$. Moreover,

$$F(V(p_1)) = F^{[m+1]}(p,\delta) + O_{\theta_{n+1}+m\theta_n+\varepsilon}(\|\delta\|^{m+1+\varepsilon}) \tag{A-5}$$

for some $\varepsilon > 0$.

Proof. Let $R := v - j_\delta^{m+1} W(p)$. Then, by the definition (4-23), this function solves

$$\partial_t R + AR = F(V(p_1)) - F^{[m+1]}(p, \delta), \quad \boldsymbol{P}_{N_{n+1}} R|_{t=0} = 0. \tag{A-6}$$

Let us study the term $F^{[m+1]}(p, \delta)$ at the right-hand side (which is defined by (4-24)). Using (A-3) and the trick (4-25), we may replace $j_\delta^m V(p)$ and $j_\delta^m W(p)$ by v in all terms in (4-24) which contain the second and higher derivatives of F (the error will be of order $\|\delta\|^{m+1+\varepsilon}$). Actually, we cannot do this in the term with the first derivative at the moment since this requires (A-3) for W of order $m+1$, which we are now verifying. This, gives

$$F^{[m+1]}(p, \delta) = F(V(p)) + F'(V(p)) j_\delta^{m+1} W(p) + \sum_{k=2}^{m+1} \frac{1}{k!} F^{(k)}(V(p))[\{v\}^k] + O_{\theta_{n+1}+m\theta_n+\varepsilon}(\|\delta\|^{m+1+\varepsilon}). \tag{A-7}$$

Indeed, let us consider the terms in (A-7) containing $j_\delta^m W$ only (the terms without it are analogous, but simpler). Using the analogue of (4-25),

$$[\{j_\delta^m V(p)\}^{k-1}, j_\delta^m W(p)] \to \sum_{\substack{n_1+\cdots+n_k \le m+1 \\ n_i \in \mathbb{N}}} B'_{n_1,\dots,n_k}\{j_\delta^{n_1} V(p), \dots, j_\delta^{n_{k-1}} V(p), j_\delta^{n_k} W(p)\}, \tag{A-8}$$

the growth exponent of the remainder does not exceed

$$(n_1 + \cdots + n_{k-1})\theta_n + \theta_{n+1} + (n_k - 1)\theta_n + \varepsilon \le \theta_{n+1} + m\theta_n + \varepsilon,$$

where we have implicitly used our induction assumptions (A-3) and decreased the exponent ε if necessary.

Using now the Taylor theorem for $F \in C^{m+1,\varepsilon}$ together with estimate (3-23) for v, we infer that

$$F(V(p_1)) - F^{[m+1]}(p, \delta) = F'(V(p))R + O_{\theta_{n+1}+m\theta_n+\varepsilon}(\|\delta\|^{m+1+\varepsilon})$$

and, therefore, the function R solves

$$\partial_t R + AR - F'(V(p))R = O_{\theta_{n+1}+m\theta_n+\varepsilon}(\|\delta\|^{m+1+\varepsilon}), \quad \boldsymbol{P}_{N_{n+1}} R|_{t=0} = 0. \tag{A-9}$$

Since by the induction assumption $\theta_n < \lambda_{N_n+1} - L$, assumption (4-21) guarantees the existence of θ_{n+1} and $\varepsilon > 0$ such that $\theta_{n+1} + m\theta_n + \varepsilon$ satisfies (3-17) with N replaced by N_{n+1}. Thus, Proposition 3.7 gives the estimate

$$\|R\|_{L^2_{e^{(\theta_{n+1}+m\theta_n+\varepsilon)t}}(\mathbb{R}_-, H)} \le C\|\delta\|^{m+1+\varepsilon}$$

and (A-4) is proved. Estimate (A-5) is now a straightforward corollary of (A-7) and the Taylor theorem (since we are now allowed to replace $j_\delta^{m+1} W$ by v). Thus, the lemma is proved. □

We now turn to the general case $\xi \neq 0$. To this end we need the following key lemma.

Lemma A.2. *Let the above assumptions hold. Then, the following formula is satisfied*:

$$F^{[m+1]}(p_1, \xi) - F^{[m+1]}(p, \xi + \delta) = F'(V(p))\big(j_\delta^{m+1} W(p) + j_\xi^{m+1} W(p_1) - j_{\xi+\delta}^{m+1} W(p)\big) + O_{\theta_{n+1}+m\theta_n+\varepsilon}((\|\delta\| + \|\xi\|)^{m+1+\varepsilon}), \tag{A-10}$$

where $\xi \in H$, $p_1, p \in \boldsymbol{P}_{N_{n+1}} \mathcal{M}_{N_1}$ *and* $\delta = p_1 - p$.

Proof. Indeed, according to the definition (4-24) and formula (A-5), we have

$$\begin{aligned} F^{[m+1]}(p_1,\xi) &= F^{[m+1]}(p,\delta) + F'(V(p_1))j_\xi^{m+1}W(p_1) \\ &\quad + \sum_{l=2}^{m+1}\frac{1}{l!}\big(lF^{(l)}(V(p_1))[j_\xi^m W(p_1), \{j_\xi^m V(p_1)\}^{l-1}] - (l-1)F^{(l)}(V(p_1))[\{j_\xi^m V(p_1)\}^l]\big) \\ &\quad + O_{\theta_{n+1}+m\theta_n+\varepsilon}((\|\xi\|+\|\delta\|)^{m+1+\varepsilon}). \end{aligned} \tag{A-11}$$

We recall that, according to our agreement and formula (4-25), the right-hand side does not contain the terms of order larger than $m+1$. Expanding now the derivatives $F^{(l)}(V(p_1))$ into Taylor series around $V(p)$ and using (A-3), we get

$$\begin{aligned} F^{[m+1]}(p_1,\xi) &= F^{[m+1]}(p,\delta) + F'(V(p))(j_\xi^{m+1}W(p_1) - j_\xi^m W(p_1)) \\ &\quad + \sum_{l=1}^{m+1}\sum_{k=l}^{m+1}\frac{1}{l!(k-l)!}\big(lF^{(k)}(V(p))[\{j_\delta^m V(p)\}^{k-l}, j_\xi^m W(p_1), \{j_\xi^m V(p_1)\}^{l-1}] \\ &\qquad\qquad - (l-1)F^{(k)}(V(p))[\{j_\delta^m V(p)\}^{k-l}, \{j_\xi^m V(p_1)\}^l]\big) \\ &\quad + O_{\theta_{n+1}+m\theta_n+\varepsilon}((\|\xi\|+\|\delta\|)^{m+1+\varepsilon}). \end{aligned} \tag{A-12}$$

Finally, changing the order of summation, we arrive at

$$\begin{aligned} F^{[m+1]}(p_1,\xi) &= F^{[m+1]}(p,\delta) + F'(V(p))j_\xi^{m+1}W(p_1) \\ &\quad + \sum_{k=2}^{m+1}\frac{1}{k!}\sum_{l=1}^{k}C_k^l\big(lF^{(k)}(V(p))[\{j_\delta^m V(p)\}^{k-l}, j_\xi^m W(p_1), \{j_\xi^m V(p_1)\}^{l-1}] \\ &\qquad\qquad - (l-1)F^{(k)}(V(p))[\{j_\delta^m V(p)\}^{k-l}, \{j_\xi^m V(p_1)\}^l]\big) \\ &\quad + O_{\theta_{n+1}+m\theta_n+\varepsilon}((\|\xi\|+\|\delta\|)^{m+1+\varepsilon}). \end{aligned} \tag{A-13}$$

Let us now look to the term $F^{[m+1]}(p,\xi+\delta)$. According to (4-24), we have

$$\begin{aligned} &F^{[m+1]}(p,\xi+\delta) \\ &\quad = F(V(p)) + F'(V(p))j_{\xi+\delta}^{m+1}W(p) \\ &\quad + \sum_{k=2}^{m+1}\frac{1}{k!}\big(kF^{(k)}(V(p))[j_{\xi+\delta}^m W(p), \{j_{\xi+\delta}^m V(p)\}^{k-1}] - (k-1)F^{(k)}(V(p))[\{j_{\xi+\delta}^m V(p)\}^k]\big). \end{aligned} \tag{A-14}$$

From the induction assumption, the compatibility assumptions (A-1) hold for $j_{\xi+\delta}^{m_1}W$ and give

$$j_{\xi+\delta}^{m_1}W(p) = j_\delta^{m_1}W(p) + j_\xi^{m_1}W(p_1) + O_{\theta_{n+1}+(m_1-1)\theta_n+\varepsilon}((\|\delta\|+\|\xi\|)^{m_1+\varepsilon})$$

for all $m_1 \le m$ and the analogous identities hold also for $j_{\xi+\delta}^{m_1}V$:

$$j_{\xi+\delta}^{m_1}V(p) = j_\delta^{m_1}V(p) + j_\xi^{m_1}V(p_1) + O_{m_1\theta_n+\varepsilon}((\|\delta\|+\|\xi\|)^{m_1+\varepsilon}).$$

Moreover, using (A-3), we may also get

$$j_{\xi+\delta}^{m_1}W(p) = j_\delta^{m_1}V(p) + j_\xi^{m_1}W(p_1) + O_{\theta_{n+1}+(m_1-1)\theta_n+\varepsilon}((\|\delta\|+\|\xi\|)^{m_1+\varepsilon})$$

for all $m_1 \le m$. Inserting these formulas to (A-14), we arrive at

$$\begin{aligned} F^{[m+1]}(p,\xi+\delta) &= F(V(p)) + F'(V(p))j^{m+1}_{\xi+\delta}W(p) \\ &\quad + \sum_{k=2}^{m+1}\frac{1}{k!}\big(kF^{(k)}(V(p))[j^m_\delta V(p) + j^m_\xi W(p_1), \{j^m_\delta V(p) + j^m_\xi V(p_1)\}^{k-1}] \\ &\qquad\qquad - (k-1)F^{(k)}(V(p))[\{j^m_\delta V(p) + j^m_\xi V(p_1)\}^k]\big) \\ &\quad + O_{\theta_{n+1}+m\theta_n+\varepsilon}((\|\delta\|+\|\xi\|)^{m+1+\varepsilon}). \end{aligned} \tag{A-15}$$

Using the binomial formula (2-1), we arrive at

$$\begin{aligned} F^{[m+1]}(p,\xi+\delta) &= F(V(p)) + F'(V(p))j^{m+1}_{\xi+\delta}W(p) \\ &\quad + \sum_{k=2}^{m+1}\frac{1}{k!}\Bigg(\sum_{l=1}^{k} kC^{l-1}_{k-1}F^{(k)}(V(p))[j^m_\xi W(p_1), \{j^m_\delta V(p)\}^{k-l}, \{j^m_\xi V(p_1)\}^{l-1}] \\ &\qquad + \sum_{l=0}^{k-1} kC^{l}_{k-1}F^{(k)}(V(p))[j^m_\delta V(p), \{j^m_\delta V(p)\}^{k-l-1}, \{j^m_\xi V(p_1)\}^{l}] \\ &\qquad\qquad - \sum_{l=0}^{k}(k-1)C^l_k F^{(k)}(V(p))[\{j^m_\delta V(p)\}^{k-l}, \{j^m_\xi V(p_1)\}^l]\Bigg) \\ &\quad + O_{\theta_{n+1}+m\theta_n+\varepsilon}.((\|\delta\|+\|\xi\|)^{m+1+\varepsilon}). \end{aligned} \tag{A-16}$$

We need to compare (A-13) and (A-16). To this end, we first note that

$$lC^l_k = kC^{l-1}_{k-1}$$

and, therefore, the terms containing the jets of W in these two formulas coincide. Thus, we only need to look at the terms without jets of W. In the case $l = k$, we have only one term in the right-hand side of (A-16), which obviously coincides with the analogous term in (A-13). Let us now look at the terms with $l = 1, \dots, k-1$. Due to the obvious identity

$$-(l-1)C^l_k = kC^l_{k-1} - (k-1)C^l_k,$$

these terms again coincide. Thus, it remains to look at the extra terms which correspond to $l = 0$ in (A-16) and which are absent in the sums of (A-13). Finally, using (A-3) and (A-5), we get the following identity involving these extra terms:

$$\begin{aligned} F(V(p)) &+ \sum_{k=2}^{m+1}\frac{1}{k!}F^{(k)}(V(p))[\{j^m_\delta V(p)\}^k] \\ &= F^{[m+1]}(p,\delta) - F'(V(p))j^{m+1}_\delta W(p) + O_{\theta_{n+1}+m\theta_n+\varepsilon}((\|\delta\|+\|\xi\|)^{m+1+\varepsilon}). \end{aligned} \tag{A-17}$$

This gives the identity

$$\begin{aligned} &F^{[m+1]}(p_1,\xi) - F'(V(p))j^{m+1}_\xi W(p_1) \\ &= F^{[m+1]}(p,\xi+\delta) - F'(V(p))(j^{m+1}_{\xi+\delta}W(p) - j^{m+1}_\delta W(p)) + O_{\theta_{n+1}+m\theta_n+\varepsilon}((\|\delta\|+\|\xi\|)^{m+1+\varepsilon}) \end{aligned} \tag{A-18}$$

and finishes the proof of the lemma. □

We are now ready to finish the check of the compatibility conditions. Note that, due to (A-4), we have

$$\begin{aligned} J_\xi^{m+1}W(p_1) - J_{\xi+\delta}^{m+1}W(p) \\ = j_\delta^{m+1}W(p) + j_\xi^{m+1}W(p_1) - j_{\xi+\delta}^{m+1}W(p) + O_{\theta_{n+1}+m\theta_n+\varepsilon}((\|\delta\|+\|\xi\|)^{m+1+\varepsilon}). \end{aligned} \tag{A-19}$$

Let finally $U(t) := J_\xi^{m+1}W(p_1) - J_{\xi+\delta}^{m+1}W(p)$. Then, according to definition (4-22), Lemma A.2 and the fact that $\delta = p_1 - p$, this function solves the equation

$$\partial_t U + AU - F'(V(p))U = O_{\theta_{n+1}+m\theta_n+\varepsilon}(\|\delta\|+\|\xi\|)^{m+1+\varepsilon}), \quad \boldsymbol{P}_{N_{n+1}}U|_{t=0} = 0, \tag{A-20}$$

and by Proposition 3.7, we arrive at

$$J_\xi^{m+1}W(p_1) - J_{\xi+\delta}^{m+1}W(p) = O_{\theta_{n+1}+m\theta_n+\varepsilon}((\|\delta\|+\|\xi\|)^{m+1+\varepsilon}). \tag{A-21}$$

Thus, the $(m+1)$-th order compatibility conditions for $J_\xi^{m+1}W(p)$ are verified. The induction with respect to m gives us that $J_\xi^{n+1}W(p)$ also satisfies the compatibility conditions (of course, we cannot take $m > n$ since we need the compatibility conditions of order m for $J_\xi^m V(p)$ to proceed). This completes the proof of our main Theorem 4.3.

References

[Babin and Vishik 1992] A. V. Babin and M. I. Vishik, *Attractors of evolution equations*, Stud. Math. Appl. **25**, North-Holland, Amsterdam, 1992. MR Zbl

[Bates et al. 1999] P. W. Bates, K. Lu, and C. Zeng, "Persistence of overflowing manifolds for semiflow", *Comm. Pure Appl. Math.* **52**:8 (1999), 983–1046. MR Zbl

[Ben-Artzi et al. 1993] A. Ben-Artzi, A. Eden, C. Foias, and B. Nicolaenko, "Hölder continuity for the inverse of Mañé's projection", *J. Math. Anal. Appl.* **178**:1 (1993), 22–29. MR Zbl

[Chepyzhov and Vishik 2002] V. V. Chepyzhov and M. I. Vishik, *Attractors for equations of mathematical physics*, Amer. Math. Soc. Colloq. Publ. **49**, Amer. Math. Soc., Providence, RI, 2002. MR Zbl

[Chow et al. 1992] S.-N. Chow, K. Lu, and G. R. Sell, "Smoothness of inertial manifolds", *J. Math. Anal. Appl.* **169**:1 (1992), 283–312. MR Zbl

[Constantin et al. 1989] P. Constantin, C. Foias, B. Nicolaenko, and R. Temam, *Integral manifolds and inertial manifolds for dissipative partial differential equations*, Appl. Math. Sci. **70**, Springer, 1989. MR Zbl

[Eden et al. 2013] A. Eden, S. V. Zelik, and V. K. Kalantarov, "Counterexamples to the regularity of Mañé's projections in the theory of attractors", *Uspekhi Mat. Nauk* **68**:2(410) (2013), 3–32. In Russian; translated in *Russ. Math. Surv.* **68**:2 (2013), 199–226. MR Zbl

[Fefferman 2005] C. L. Fefferman, "A sharp form of Whitney's extension theorem", *Ann. of Math.* (2) **161**:1 (2005), 509–577. MR Zbl

[Fenichel 1972] N. Fenichel, "Persistence and smoothness of invariant manifolds for flows", *Indiana Univ. Math. J.* **21** (1972), 193–226. MR Zbl

[Foias et al. 1988] C. Foias, G. R. Sell, and R. Temam, "Inertial manifolds for nonlinear evolutionary equations", *J. Differential Equations* **73**:2 (1988), 309–353. MR Zbl

[Gal and Guo 2018] C. G. Gal and Y. Guo, "Inertial manifolds for the hyperviscous Navier–Stokes equations", *J. Differential Equations* **265**:9 (2018), 4335–4374. MR Zbl

[Hájek and Johanis 2014] P. Hájek and M. Johanis, *Smooth analysis in Banach spaces*, de Gruyter Ser. Nonlinear Anal. Appl. **19**, de Gruyter, Berlin, 2014. MR Zbl

[Hale 1988] J. K. Hale, *Asymptotic behavior of dissipative systems*, Math. Surv. Monogr. **25**, Amer. Math. Soc., Providence, RI, 1988. MR Zbl

[Henry 1981] D. Henry, *Geometric theory of semilinear parabolic equations*, Lecture Notes in Math. **840**, Springer, 1981. MR Zbl

[Hirsch et al. 1977] M. W. Hirsch, C. C. Pugh, and M. Shub, *Invariant manifolds*, Lecture Notes in Math. **583**, Springer, 1977. MR Zbl

[Hunt and Kaloshin 1999] B. R. Hunt and V. Y. Kaloshin, "Regularity of embeddings of infinite-dimensional fractal sets into finite-dimensional spaces", *Nonlinearity* **12**:5 (1999), 1263–1275. MR Zbl

[Katok and Hasselblatt 1995] A. Katok and B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Encycl. Math. Appl. **54**, Cambridge Univ. Press, 1995. MR Zbl

[Kielhöfer 2004] H. Kielhöfer, *Bifurcation theory: an introduction with applications to PDEs*, Appl. Math. Sci. **156**, Springer, 2004. MR Zbl

[Koksch 1998] N. Koksch, "Almost sharp conditions for the existence of smooth inertial manifolds", pp. 139–166 in *Equadiff IX*: *Conference on Differential Equations and Their Applications* (Brno, Czech Republic, 1997), edited by Z. Došlá et al., Masaryk Univ., Brno, 1998.

[Kostianko 2018] A. Kostianko, "Inertial manifolds for the 3D modified-Leray-α model with periodic boundary conditions", *J. Dynam. Differential Equations* **30**:1 (2018), 1–24. MR Zbl

[Kostianko 2020] A. Kostianko, "Bi-Lipschitz Mané projectors and finite-dimensional reduction for complex Ginzburg–Landau equation", *Proc. A.* **476**:2239 (2020), art. id. 20200144. MR Zbl

[Kostianko and Zelik 2015] A. Kostianko and S. Zelik, "Inertial manifolds for the 3D Cahn–Hilliard equations with periodic boundary conditions", *Commun. Pure Appl. Anal.* **14**:5 (2015), 2069–2094. MR Zbl

[Kostianko and Zelik 2017] A. Kostianko and S. Zelik, "Inertial manifolds for 1D reaction-diffusion-advection systems, I: Dirichlet and Neumann boundary conditions", *Commun. Pure Appl. Anal.* **16**:6 (2017), 2357–2376. MR Zbl

[Kostianko and Zelik 2018] A. Kostianko and S. Zelik, "Inertial manifolds for 1D reaction-diffusion-advection systems, II: Periodic boundary conditions", *Commun. Pure Appl. Anal.* **17**:1 (2018), 285–317. MR Zbl

[Kostianko and Zelik 2022] A. Kostianko and S. Zelik, "Kwak transform and inertial manifolds revisited", *J. Dynam. Differential Equations* **34**:4 (2022), 2975–2995. MR Zbl

[Li and Sun 2020] X. Li and C. Sun, "Inertial manifolds for the 3D modified-Leray-α model", *J. Differential Equations* **268**:4 (2020), 1532–1569. MR Zbl

[Mallet-Paret and Sell 1988] J. Mallet-Paret and G. R. Sell, "Inertial manifolds for reaction diffusion equations in higher space dimensions", *J. Amer. Math. Soc.* **1**:4 (1988), 805–866. MR Zbl

[Mallet-Paret et al. 1993] J. Mallet-Paret, G. R. Sell, and Z. D. Shao, "Obstructions to the existence of normally hyperbolic inertial manifolds", *Indiana Univ. Math. J.* **42**:3 (1993), 1027–1055. MR Zbl

[Miklavčič 1991] M. Miklavčič, "A sharp condition for existence of an inertial manifold", *J. Dynam. Differential Equations* **3**:3 (1991), 437–456. MR Zbl

[Miranville and Zelik 2008] A. Miranville and S. Zelik, "Attractors for dissipative partial differential equations in bounded and unbounded domains", pp. 103–200 in *Handbook of differential equations*: *evolutionary equations*, *IV*, edited by C. M. Dafermos and M. Pokorný, Elsevier, Amsterdam, 2008. MR Zbl

[Robinson 1999] J. C. Robinson, "Global attractors: topology and finite-dimensional dynamics", *J. Dynam. Differential Equations* **11**:3 (1999), 557–581. MR Zbl

[Robinson 2001] J. C. Robinson, *Infinite-dimensional dynamical systems: an introduction to dissipative parabolic PDEs and the theory of global attractors*, Cambridge Univ. Press, 2001. MR Zbl

[Robinson 2011] J. C. Robinson, *Dimensions, embeddings, and attractors*, Cambridge Tracts in Math. **186**, Cambridge Univ. Press, 2011. MR Zbl

[Roman 1980] S. Roman, "The formula of Faà di Bruno", *Amer. Math. Monthly* **87**:10 (1980), 805–809. MR Zbl

[Romanov 1993] A. V. Romanov, "Sharp estimates for the dimension of inertial manifolds for nonlinear parabolic equations", *Izv. Ross. Akad. Nauk Ser. Mat.* **57**:4 (1993), 36–54. In Russian; translated in *Izv. Math.* **43**:1 (1994), 31–47. MR Zbl

[Romanov 2000] A. V. Romanov, "Three counterexamples in the theory of inertial manifolds", *Mat. Zametki* **68**:3 (2000), 439–447. In Russian; translated in *Math. Notes* **68**:3-4 (2000), 378–385. MR Zbl

[Rosa and Temam 1996] R. Rosa and R. Temam, "Inertial manifolds and normal hyperbolicity", *Acta Appl. Math.* **45**:1 (1996), 1–50. MR Zbl

[Sell and You 2002] G. R. Sell and Y. You, *Dynamics of evolutionary equations*, Appl. Math. Sci. **143**, Springer, 2002. MR Zbl

[Stein 1970] E. M. Stein, *Singular integrals and differentiability properties of functions*, Princeton Math. Series **30**, Princeton Univ. Press, 1970. MR Zbl

[Temam 1988] R. Temam, *Infinite-dimensional dynamical systems in mechanics and physics*, Appl. Math. Sci. **68**, Springer, 1988. MR Zbl

[Wells 1973] J. C. Wells, "Differentiable functions on Banach spaces with Lipschitz derivatives", *J. Differential Geom.* **8** (1973), 135–152. MR Zbl

[Zelik 2014] S. Zelik, "Inertial manifolds and finite-dimensional reduction for dissipative PDEs", *Proc. Roy. Soc. Edinburgh Sect. A* **144**:6 (2014), 1245–1327. MR Zbl

Received 20 Jun 2021. Accepted 26 Jul 2022.

ANNA KOSTIANKO: a.kostianko@imperial.ac.uk
Department of Mathematics, Zhejiang Normal University, Zhejiang, China

SERGEY ZELIK: s.zelik@surrey.ac.uk
Department of Mathematics, Zhejiang Normal University, Zhejiang, China

and

Department of Mathematics, University of Surrey, Guildford, United Kingdom

and

Keldysh Institute of Applied Mathematics, Moscow, Russia

ANALYSIS AND PDE
Vol. 17 (2024), No. 2, pp. 535–585
DOI: 10.2140/apde.2024.17.535

SEMICLASSICAL EIGENVALUE ESTIMATES UNDER MAGNETIC STEPS

WAFAA ASSAAD, BERNARD HELFFER AND AYMAN KACHMAR

We establish accurate eigenvalue asymptotics and, as a by-product, sharp estimates of the splitting between two consecutive eigenvalues for the Dirichlet magnetic Laplacian with a nonuniform magnetic field having a jump discontinuity along a smooth curve. The asymptotics hold in the semiclassical limit, which also corresponds to a large magnetic field limit and is valid under a geometric assumption on the curvature of the discontinuity curve.

1. Introduction

The paper studies a semiclassical Schrödinger operator with a step magnetic field and Dirichlet boundary conditions, in a smooth bounded domain. The aim is to give accurate estimates of the lower eigenvalues in the semiclassical limit.

Let Ω be an open, bounded, and simply connected subset of $\mathbb{R}^2$ with smooth C^1 boundary. We consider a simple smooth curve $\Gamma \subset \mathbb{R}^2$ that splits $\mathbb{R}^2$ into two disjoint unbounded open sets, P_1 and P_2, and such that Γ is a semistraight line when $|x|$ tends to $+\infty$. We assume that Γ decomposes Ω into two sets Ω_1 and Ω_2 as follows (see Figure 1):

(1) Γ intersects $\partial\Omega$ transversally at two distinct points.

(2) $\Omega_1 := \Omega \cap P_1 \neq \varnothing$ and $\Omega_2 := \Omega \cap P_2 \neq \varnothing$.

Let $h > 0$ and $\boldsymbol{F} = (F_1, F_2) \in H^1_{\mathrm{loc}}(\mathbb{R}^2)$ be a magnetic potential whose associated magnetic field is

$$\operatorname{curl} \boldsymbol{F} = a_1 \mathbb{1}_{P_1} + a_2 \mathbb{1}_{P_2}, \quad \boldsymbol{a} := (a_1, a_2) \in \mathbb{R}^2,\ a_1 \neq a_2. \tag{1-1}$$

When restricted to Ω, the vector field $\boldsymbol{F}$ satisfies

$$\operatorname{curl} \boldsymbol{F} = a_1 \mathbb{1}_{\Omega_1} + a_2 \mathbb{1}_{\Omega_2}, \quad \boldsymbol{a} := (a_1, a_2) \in \mathbb{R}^2,\ a_1 \neq a_2 \text{ and } \boldsymbol{F} \in L^4(\Omega). \tag{1-2}$$

Note that the curve Γ separates the two regions Ω_1 and Ω_2 which are assigned with different values of the magnetic field. For this reason, we refer to Γ as the *magnetic edge*. We consider the quadratic form on $H^1_0(\Omega)$

$$u \mapsto \mathcal{Q}_h(u) = \int_\Omega |(h\nabla - i\boldsymbol{F})u|^2\, dx. \tag{1-3}$$

MSC2020: 35P15, 35P20, 81Q20.
Keywords: semiclassical analysis, magnetic Laplacian, magnetic steps.

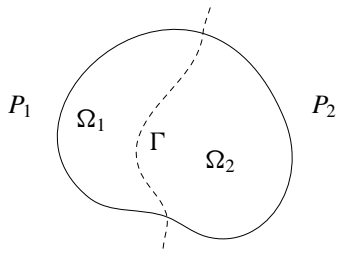

Figure 1. The curve Γ transversally cuts $\partial\Omega$ at two points and splits Ω into two regions, Ω_1 and Ω_2.

This quadratic form is closed on the form domain $H_0^1(\Omega)$. By the Friedrichs extension procedure, we can associate its Dirichlet realization in Ω

$$\mathcal{P}_h := -(h\nabla - i\boldsymbol{F})^2 = -\sum_{j=1}^{2}(h\partial_{x_j} - iF_j)^2, \tag{1-4}$$

whose domain is

$$\mathrm{Dom}(\mathcal{P}_h) = \{u \in L^2(\Omega) : (h\nabla - i\boldsymbol{F})^j u \in L^2(\Omega),\ j \in \{1,2\},\ u|_{\partial\Omega} = 0\}. \tag{1-5}$$

The operator $\mathcal{P}_h$ is self-adjoint, has compact resolvent, and its spectrum is an increasing sequence, $(\lambda_n(h))_{n\in\mathbb{N}}$, of real eigenvalues listed with multiplicities.

In this contribution, we aim at giving the asymptotic expansion of the low-lying eigenvalues of $\mathcal{P}_h$, in the semiclassical limit, i.e., when h tends to 0.

Schrödinger operators with a discontinuous magnetic field, like $\mathcal{P}_h$, appear in many models in nanophysics such as in quantum transport while studying the transport properties of a bidimensional electron gas [Reijniers and Peeters 2000; Peeters and Matulis 1993]. In that context, the magnetic edge is *straight* and bound states interestingly feature currents flowing along the magnetic edge.

The present contribution addresses another appealing question on the influence of the shape of the magnetic edge on the energy of the bound states. We give an affirmative answer by providing sharp semiclassical eigenvalue asymptotics under a single "well" hypothesis on the curvature of the magnetic edge (see Assumption 1.1 and Theorem 1.2 below). Loosely speaking, our hypothesis says that we perform a local deformation of the magnetic edge so that its curvature has a unique nondegenerate maximum.

Another important occurrence of magnetic Laplace operators is in the Ginzburg–Landau model of superconductivity [Saint-James and de Gennes 1963]. In bounded domains, the spectral properties of these operators can describe interesting physical situations. In the context of superconductivity, accurate information about the lowest eigenvalues is important for giving a precise description of the concentration of superconductivity in a type-II superconductor. Moreover, it improves the estimates of the third critical field, H_{C_3}, that marks the onset of superconductivity in the domain. We refer the reader to [Assaad and Kachmar 2022; Assaad 2021] for discontinuous field cases, and to [Fournais and Helffer 2006; Helffer and Pan 2003; Lu and Pan 1999a; 1999b; 2000; Bonnaillie-Noël and Fournais 2007; Bonnaillie-Noël and Dauge 2006; Bernoff and Sternberg 1998; Tilley and Tilley 1990] for a further discussion in smooth

fields cases. In the present paper, the Dirichlet realization of $\mathcal{P}_h$ in the bounded domain Ω can physically correspond to a superconductor which is set in the normal (nonsuperconducting) state at its boundary.

Using symmetry and scaling arguments, one can reduce the problem to the study of cases of $\boldsymbol{a} = (a_1, a_2)$, where $a_1 = 1$ and $a_2 = a \in [-1, 1)$. Moreover, we will soon make a more restrictive choice of cases of $\boldsymbol{a}$ (see (1-11) below). Towards justifying the upcoming choice of $\boldsymbol{a}$, we introduce the effective operator $\mathfrak{h}_a[\xi]$ with a discontinuous field, defined on $\mathbb{R}$ and parametrized by $\xi \in \mathbb{R}$:

$$\mathfrak{h}_a[\xi] = -\frac{d^2}{d\tau^2} + (\xi + b_a(\tau)\tau)^2, \tag{1-6}$$

where

$$b_a(\tau) = \mathbf{1}_{\mathbb{R}_+}(\tau) + a\mathbf{1}_{\mathbb{R}_-}(\tau). \tag{1-7}$$

This operator arises from the approximation by the case where $\Omega = \mathbb{R}^2$ and $\Gamma = \{x_2 = 0\}$, τ corresponding to the variable x_2 and ξ being the dual variable of x_1. The known spectral properties of $\mathfrak{h}_a[\xi]$, obtained earlier in [Hislop et al. 2016; Assaad et al. 2019; Assaad and Kachmar 2022], are recalled in Section 2A. Here, we only present some features of this operator that are useful to this introduction. The bottom of the spectrum of $\mathfrak{h}_a[\xi]$, denoted by $\mu_a(\xi)$, is a simple eigenvalue for $a \neq 0$, usually called the *band function* in the literature. Minimizing the band function leads us to introduce

$$\beta_a = \inf_{\xi \in \mathbb{R}} \mu_a(\xi). \tag{1-8}$$

We list the following properties of β_a, depending on the values of a:

Case $a = -1$: In the case where $\Omega = \mathbb{R}^2$ and $\Gamma = \{x_2 = 0\}$, this case is called the "symmetric trapping magnetic steps" and is well-understood in the literature (see, e.g., [Hislop et al. 2016]). In this case, the study of $\mathfrak{h}_a[\xi]$ can be reduced to that of the de Gennes operator (a harmonic oscillator on the half-axis with Neumann condition at the origin). We refer the reader to [Fournais and Helffer 2010] for the spectral properties of this operator. Here,

$$\Theta_0 := \beta_{-1} \approxeq 0.59 \tag{1-9}$$

is attained by $\mu_{-1}(\cdot)$ at a unique and nondegenerate minimum $\xi_0 = -\sqrt{\Theta_0}$. Moreover, $\beta_{-1} = \mu_{-1}(\xi_0)$ is a simple eigenvalue of $\mathfrak{h}_{-1}[\xi_0]$.

Case $-1 < a < 0$: This case is called the "asymmetric trapping magnetic steps" and is studied in many works (see [Assaad and Kachmar 2022; Assaad et al. 2019; Hislop et al. 2016]). We have $|a|\Theta_0 < \beta_a < \min(|a|, \Theta_0)$ and β_a is attained by $\mu_a(\cdot)$ at a unique $\zeta_a < 0$ [Assaad and Kachmar 2022]

$$\mu_a(\zeta_a) = \beta_a. \tag{1-10}$$

Moreover, the minimum is nondegenerate, i.e., $\mu_a''(\zeta_a) > 0$.

Case $a = 0$: This corresponds to the "magnetic wall" case studied for instance in [Reijniers and Peeters 2000; Hislop et al. 2016]. We refer to [Hislop et al. 2016, Section 2] for this case.

For $\xi \leq 0$, we have

$$\sigma(h_a[\xi]) = \sigma_{\mathrm{ess}}(h_a[\xi]) = [\xi^2, +\infty),$$

where σ and σ_{ess} respectively denote the spectrum and essential spectrum.

For $\xi > 0$,

$$\sigma_{\mathrm{ess}}(h_a[\xi]) = [\xi^2, +\infty)$$

and $h_a[\xi]$ may have positive eigenvalues $\lambda < \xi^2$. Consequently, $\beta_0 = \mu_0(0) = \inf \sigma_{\mathrm{ess}} \mathfrak{h}_0[0] = 0$, and β_0 is not an eigenvalue of $\mathfrak{h}_a[\xi]$ for all $\xi \in \mathbb{R}$.

Case $0 < a < 1$: This corresponds with the "nontrapping magnetic steps" case; see [Assaad et al. 2019; Hislop and Soccorsi 2015; Iwatsuka 1985]. Here, $\beta_a = a$ and $\mu_a(\cdot)$ doesn't achieve a minimum; the infimum is attained at $+\infty$.

A key ingredient in establishing the asymptotics of the eigenvalues $\lambda_n(h)$ is that β_a is an eigenvalue of $\mathfrak{h}_a[\xi]$ for some $\xi \in \mathbb{R}$. We will use the corresponding eigenfunction in constructing quasimodes of the operator $\mathcal{P}_h$. The above discussion shows that β_a is an eigenvalue only when $a \in [-1, 0)$. The case $a = -1$ is excluded from our study, despite the fact that β_{-1} is an eigenvalue of $\mathfrak{h}_{-1}[\xi_0]$. Except when Γ is an axis of symmetry of Ω as in [Hislop et al. 2016], the situation is more difficult and the curvature will play a more important role. We hope to treat this case in a future work. This explains our choice to work under the following assumption on $\boldsymbol{a}$ (thus on the magnetic field curl $\boldsymbol{F}$) throughout the paper:

$$\boldsymbol{a} = (1, a), \quad \text{with } -1 < a < 0. \tag{1-11}$$

Under assumption (1-11), we introduce two spectral invariants

$$c_2(a) = \frac{1}{2}\mu_a''(\zeta_a) > 0 \quad \text{and} \quad M_3(a) = \frac{1}{3}\Big(\frac{1}{a} - 1\Big)\zeta_a \phi_a(0)\phi_a'(0) < 0, \tag{1-12}$$

where μ_a and ζ_a are introduced in (1-8) and (1-10), and ϕ_a is the positive L^2-normalized eigenfunction of $\mathfrak{h}_a[\zeta_a]$ corresponding to β_a.

Furthermore, we work under the following assumption:

Assumption 1.1. The curvature $\Gamma \ni s \mapsto k(s)$ at the magnetic edge has a unique maximum

$$k(s) < k(s_0) =: k_{\max} \quad \text{for } s \neq s_0.$$

This maximum is attained in $\Gamma \cap \Omega$ and is nondegenerate:

$$k_2 := k''(s_0) < 0.$$

The goal of this paper is to prove the following theorem:

Theorem 1.2. *Let $n \in \mathbb{N}^*$ and $\boldsymbol{a} = (1, a)$, with $-1 < a < 0$. Under Assumption 1.1, the n-th eigenvalue $\lambda_n(h)$ of $\mathcal{P}_h$, defined in (1-4), satisfies, as $h \to 0$,*

$$\lambda_n(h) = h\beta_a + h^{3/2} k_{\max} M_3(a) + h^{7/4}(2n-1)\sqrt{\frac{k_2 M_3(a) c_2(a)}{2}} + \mathcal{O}(h^{15/8}),$$

where β_a, $c_2(a)$ and $M_3(a)$ are the spectral quantities introduced in (1-8) and (1-12).

Remark 1.3. This theorem extends [Assaad and Kachmar 2022, Theorem 4.5], where the first two terms in the expansion of the first eigenvalue were determined with a remainder in $\mathcal{O}(h^{5/3})$. The proof of Theorem 1.2 partially relies on decay estimates of the eigenfunctions with the right scale; see Section 6

and [Assaad and Kachmar 2022]. In fact, away from the edge Γ, the eigenfunctions decay exponentially at the scale $h^{-1/2}$ of the distance to Γ, while, along Γ, they decay exponentially with a scale of $h^{-1/8}$ of the tangential distance on Γ to the point with maximum curvature.

Comparison with earlier situations. It is useful to compare the asymptotics of $\lambda_n(h)$ in Theorem 1.2 with those obtained in the literature for regular domains submitted to uniform magnetic fields. In bounded planar domains with smooth boundary, subject to unit magnetic fields and when the *Neumann* boundary condition is imposed, the low-lying eigenvalues of the linear operator, analogous to $\mathcal{P}_h$, admit the following asymptotics as h tends to 0 (see, e.g., [Fournais and Helffer 2006]):

$$\lambda_n(h) = h\Theta_0 - h^{3/2}\tilde{k}_{\max}C_1 + h^{7/4}C_1\Theta_0^{1/4}(2n-1)\sqrt{\tfrac{3}{2}\tilde{k}_2} + \mathcal{O}(h^{15/8}),$$

where Θ_0 is as in (1-9), $C_1 > 0$ is some spectral value, and $\tilde{k}_{\max}$ and $\tilde{k}_2$ are positive constants introduced in what follows. In this uniform field/Neumann condition situation, the eigenstates localize near the boundary of the domain. More precisely, they localize near the point $\tilde{s}$ with maximum curvature $k(\tilde{s})$ of this boundary, assuming the uniqueness and nondegeneracy of this point. We define $\tilde{k}_{\max} = k(\tilde{s})$ and $\tilde{k}_2 = -k''(\tilde{s}) > 0$. In [Fournais and Helffer 2006], the foregoing localization of eigenstates restricted the study to the boundary, involving a family of one-dimensional effective operators which act in the normal direction to the boundary. These are the de Gennes operators

$$\mathfrak{h}^N[\xi] = -\frac{d^2}{d\tau^2} + (\xi+\tau)^2,$$

defined on $\mathbb{R}_+$ with Neumann boundary condition at $\tau = 0$, and parametrized by $\xi \in \mathbb{R}$. We recover the value Θ_0 as an *effective energy* associated to $(\mathfrak{h}^N[\xi])_\xi$,

$$\Theta_0 = \inf_{\xi\in\mathbb{R}} \mu^N(\xi),$$

where $\mu^N(\xi)$ is the bottom of the spectrum $\sigma(\mathfrak{h}^N[\xi])$ of $\mathfrak{h}^N[\xi]$, for $\xi \in \mathbb{R}$.

Back to our discontinuous field case with *Dirichlet* boundary condition, we prove that our eigenstates are localized near the magnetic edge Γ, and more particularly, near the point with maximum curvature of this edge (see Section 6). Analogously to the aforementioned uniform field/Neumann condition situation, our study near Γ involves the family of one-dimensional effective operators $(\mathfrak{h}_a[\xi])_{\xi\in\mathbb{R}}$ which act in the normal direction to the edge Γ, along with the associated effective energy β_a.

At this stage, it is natural to discuss our problem when the Dirichlet boundary conditions are replaced by Neumann boundary ones. In this situation, one can prove the concentration of the eigenstates of the operator $\mathcal{P}_h$ near the points of intersection between the edge Γ and the boundary $\partial\Omega$. This was shown in [Assaad 2021, Theorem 6.1] at least for the lowest eigenstate. In such settings, a geometric condition is usually imposed related to the angles formed at the intersection $\Gamma\cap\partial\Omega$; see [Assaad 2021, Assumption 1.3 and Remark 1.4]. The localization of the eigenstates near $\Gamma\cap\partial\Omega$ will involve effective models that are genuinely two-dimensional, i.e., they cannot be fibered to one-dimensional operators; see [Assaad 2021, Section 3]. Studying this case may show similarity features with the case of piecewise smooth bounded domains with corners submitted to uniform magnetic fields, treated in [Bonnaillie-Noël and Dauge 2006];

see also [Bonnaillie-Noël et al. 2007; Bonnaillie-Noël 2005; Bonnaillie-Noël 2003; 2007] for studies on corner domains. Such similarities were first revealed in [Assaad 2021, Section 1.3]. More precisely, one expects the result in the discontinuous field/Neumann condition situation to be similar to that in [Bonnaillie-Noël and Dauge 2006, Theorem 7.1]. Such a result is worth establishing in a future work.

Theorem 1.2 permits us to deduce the splitting between the ground-state energy (lowest eigenvalue) and the energy of the first excited state of $\mathcal{P}_h$. More precisely, introducing the spectral gap

$$\Delta(h) := \lambda_2(h) - \lambda_1(h),$$

we get by Theorem 1.2:

Corollary 1.4. *Under the conditions in Theorem 1.2, we have as $h \to 0$*

$$\Delta(h) = h^{7/4}\sqrt{2k_2 M_3(a)c_2(a)} + \mathcal{O}(h^{15/8}).$$

Apart from its own interest, estimating the foregoing spectral gap has potential applications in nonlinear bifurcation problems, for instance, in the context of the Ginzburg–Landau model of superconductivity (see [Fournais and Helffer 2010, Section 13.5.1]).

Remark 1.5. Altering the regularity/geometry of the edge Γ may lead to radical changes in Theorem 1.2.

- If Γ is a piecewise smooth curve (a broken edge) then we have to analyze a new model in the full plane (reminiscent of a model in [Assaad 2021]). We expect analogies with domains with corners in a uniform magnetic field [Bonnaillie-Noël 2003].
- If we relax Assumption 1.1 by allowing the curvature k to have two symmetric maxima, then a tunnel effect may occur and the splitting in Theorem 1.2 becomes of exponential order. This was recently analyzed in [Fournais et al. 2022] based on the analysis of this paper and [Bonnaillie-Noël et al. 2022].
- If the curvature along Γ or a part of Γ is constant, then we expect that the magnitude of the splitting in Theorem 1.2 will change too, probably leading to multiple eigenvalues. It would be desirable to get accurate estimates in this setting. We expect analogies with disc domains in a nonuniform magnetic field [Fournais and Persson-Sundqvist 2015].

Heuristics of the proofs. Our proof of Theorem 1.2 is purely variational. The derivation of the eigenvalue upper bound is rather standard. It is obtained by computing the energy of a well-chosen trial state, $v_{h,n}^{\mathrm{app}}$, constructed by expressing the operator in a Frenet frame near the point of maximum curvature and doing WKB like expansions (for the operator and the trial state).

Proving the eigenvalue lower bound is more involved. The idea is to project the actual bound state, $v_{h,n}$, on the trial state $v_{h,n}^{\mathrm{app}}$, and to prove that this provides us with a well-chosen trial state for a one-dimensional effective operator, $H_a^{\mathrm{harm}} = -c_2(a)\partial_\sigma^2 - \frac{1}{2}k_2 M_3(a)\sigma^2$. To validate this method, we need sharp estimates of the tangential derivative of the actual bound state, which we derive via a simple, but lengthy and quite technical method involving Agmon estimates and other implementations from one-dimensional model operators. At this stage, one advantage of our approach seems its applicability with weaker regularity assumptions on the magnetic edge or the magnetic field, which could be useful in other situations as well, like the study of the three-dimensional problem in [Helffer and Morame 2004].

Outline of the paper. The paper is organized as follows. Sections 2 and 3 contain the necessary material on the model one-dimensional problems for flat and curved magnetic edges, respectively. Section 4 is devoted to the eigenvalue upper bounds matching with the asymptotics of Theorem 1.2. Here, we give the construction of the aforementioned trial state $v_{h,n}^{\mathrm{app}}$.

In Sections 5 and 6, we estimate the tangential derivative of the actual bound states, after being truncated and properly expressed in rescaled variables. The tangential derivative estimate of the L^2 norm will follow straightforwardly from the main result of Section 5. However, a higher-regularity estimate will require additional work in Section 6.

In Section 7, using the actual bound states, we construct trial states for the effective one-dimensional operator, and eventually prove the eigenvalue lower bounds of Theorem 1.2. Finally, we give two appendices, Appendix A on the Frenet coordinates near the magnetic edge, and Appendix B on the control of a remainder term that we meet in Section 7.

2. Fiber operators

2A. ***Band functions.*** Let $a \in [-1, 0)$. We first introduce some constants whose definition involves the following family of fiber operators in $L^2(\mathbb{R})$:

$$\mathfrak{h}_a[\xi] = -\frac{d^2}{d\tau^2} + V_a(\xi, \tau), \tag{2-1}$$

where $\xi \in \mathbb{R}$ is a parameter,

$$V_a(\xi, \tau) = (\xi + b_a(\tau)\tau)^2, \quad b_a(\tau) = \mathbf{1}_{\mathbb{R}_+}(\tau) + a\mathbf{1}_{\mathbb{R}_-}(\tau), \tag{2-2}$$

and the domain of $\mathfrak{h}_a[\xi]$ is given by

$$\mathrm{Dom}(\mathfrak{h}_a[\xi]) = B^2(\mathbb{R}).$$

Here the space $B^n(I)$ is defined for a positive integer n and an open interval $I \subset \mathbb{R}$ as

$$B^n(I) = \left\{ u \in L^2(I) : \tau^i \frac{d^j u}{d\tau^j} \in L^2(I) \text{ for all } i, j \in \mathbb{N} \text{ such that } i + j \le n \right\}. \tag{2-3}$$

The operator $\mathfrak{h}_a[\xi]$ is essentially self-adjoint and has compact resolvent. Actually, it can also be defined as the Friedrichs realization starting from the closed quadratic form

$$u \mapsto q_a[\xi](u) = \int_{\mathbb{R}} (|u'(\tau)|^2 + V_a(\xi, \tau)|u(\tau)|^2)\, d\tau \tag{2-4}$$

defined on $B^1(\mathbb{R})$.

For $(a, \xi) \in [-1, 0) \times \mathbb{R}$, the ground-state energy (bottom of the spectrum) $\mu_a(\xi)$ of $\mathfrak{h}_a[\xi]$ can be characterized by

$$\mu_a(\xi) = \inf_{u \in B^1(\mathbb{R}), u \neq 0} \frac{q_a[\xi](u)}{\|u\|^2_{L^2(\mathbb{R})}}, \tag{2-5}$$

and $\xi \mapsto \mu_a(\xi)$ will be called the band function.

We then introduce the *step constant* at a by

$$\beta_a := \inf_{\xi\in\mathbb{R}} \mu_a(\xi). \tag{2-6}$$

For $a=-1$, it is easy to identify by symmetrization $\mu_{-1}(\xi)$ with the ground-state energy of the Neumann realization of $-(d^2/d\tau^2)+(\tau+\xi)^2$ in $\mathbb{R}_+$ and therefore

$$\beta_{-1} = \Theta_0, \tag{2-7}$$

where Θ_0 is the celebrated de Gennes constant.

By the general theory for the Schrödinger operator, $\mu_a(\xi)$ is, for each $\xi\in\mathbb{R}$, a simple eigenvalue, that we associate with a unique *positive* L^2-normalized eigenfunction denoted by $\varphi_{a,\xi}$, i.e., satisfying

$$\varphi_{a,\xi} > 0, \quad (\mathfrak{h}_a[\xi]-\mu_a(\xi))\varphi_{a,\xi} = 0 \quad \text{and} \quad \int_{\mathbb{R}} |\varphi_{a,\xi}(\tau)|^2\, d\tau = 1. \tag{2-8}$$

By Kato's theory, the band function μ_a is an analytic function on $\mathbb{R}$. Its derivative was computed in [Hislop and Soccorsi 2015] (see also [Assaad et al. 2019, Proposition A.4]),

$$\mu_a'(\xi) = \Big(1-\frac{1}{a}\Big)\big(\varphi_{a,\xi}'(0)^2 + (\mu_a(\xi)-\xi^2)\varphi_{a,\xi}(0)^2\big), \tag{2-9}$$

which results from the following Feynman–Hellmann formula (see [Assaad et al. 2019, equation (A.9); Bolley and Helffer 1993; Dauge and Helffer 1993]):

$$\mu_a'(\xi) = 2\int_{\mathbb{R}} (\xi + b_a(\tau)\tau)|\varphi_{a,\xi}(\tau)|^2\, d\tau. \tag{2-10}$$

2B. ***Properties of band functions/states.*** For $a\in(-1,0)$, the following results were recently established in [Assaad and Kachmar 2022; Assaad et al. 2019; Hislop et al. 2016]:

(1) $|a|\Theta_0 < \beta_a < \min(|a|, \Theta_0)$.

(2) There exists a unique $\zeta_a\in\mathbb{R}$ such that $\beta_a = \mu_a(\zeta_a)$.

(3) $\zeta_a < 0$, $\mu_a''(\zeta_a) > 0$ and the ground state $\phi_a := \varphi_{a,\zeta_a}$ satisfies

$$\phi_a'(0) < 0 \text{ and } \zeta_a = -\sqrt{\beta_a + (\phi_a'^2(0)/\phi_a^2(0))}.$$

In particular, using (2-10) for $\xi=\zeta_a$, we observe that the functions ϕ_a and $(\zeta_a + b_a(\tau)\tau)\phi_a$ are orthogonal

$$\int_{\mathbb{R}} (\zeta_a + b_a(\tau)\tau)|\phi_a(\tau)|^2\, d\tau = 0. \tag{2-11}$$

Moreover, the ground-state ϕ_a satisfies the following decay estimates:

Proposition 2.1. *Let $a\in[-1,0)$. For any $\gamma>0$, there exists a positive constant C_γ such that*

$$\int_{\mathbb{R}} e^{\gamma|\tau|}(|\phi_a(\tau)|^2 + |\phi_a'(\tau)|^2)\, d\tau \le C_\gamma.$$

Consequently, for all $n \in \mathbb{N}^*$ there exists $C_n > 0$ such that

$$\int_{\mathbb{R}} |\tau|^n |\phi_a(\tau)|^2 \, d\tau \le C_n. \tag{2-12}$$

The proof is classical by using Agmon's approach for proving decay estimates. We omit it and refer the reader to [Fournais and Helffer 2010, Theorem 7.2.2] or to the proof of Lemma 2.4 below.

2C. ***Moments.*** Later in the paper, we will encounter the *moments*

$$M_n(a) = \int_{-\infty}^{+\infty} \frac{1}{b_a(\tau)} (\zeta_a + b_a(\tau)\tau)^n |\phi_a(\tau)|^2 \, d\tau, \tag{2-13}$$

which are finite according to (2-12).

For $n \in \{1, 2, 3\}$, they were computed in [Assaad and Kachmar 2022] and we have

$$M_1(a) = 0, \tag{2-14}$$

$$M_2(a) = -\frac{1}{2}\beta_a \int_{-\infty}^{+\infty} \frac{1}{b_a(\tau)} |\phi_a(\tau)|^2 \, d\tau + \frac{1}{4}\Big(\frac{1}{a} - 1\Big)\zeta_a \phi_a(0)\phi_a'(0), \tag{2-15}$$

$$M_3(a) = \frac{1}{3}\Big(\frac{1}{a} - 1\Big)\zeta_a \phi_a(0)\phi_a'(0). \tag{2-16}$$

Remark 2.2. From the properties of the band function recalled in Section 2B, we get that $M_3(a)$ is negative for $-1 < a < 0$ and vanishes for $a = -1$.

Remark 2.3. The next identities follow in a straightforward manner from the foregoing formulas of the moments:

$$\int_{-\infty}^{+\infty} \tau(\zeta_a + b_a(\tau)\tau)|\phi_a(\tau)|^2 \, d\tau = M_2(a),$$

$$\int_{-\infty}^{+\infty} \tau(\zeta_a + b_a(\tau)\tau)^2|\phi_a(\tau)|^2 \, d\tau = M_3(a) - \zeta_a M_2(a),$$

$$\int_{-\infty}^{+\infty} b_a(\tau)\tau^2(\zeta_a + b_a(\tau)\tau)|\phi_a(\tau)|^2 \, d\tau = M_3(a) - 2\zeta_a M_2(a),$$

$$\int_{-\infty}^{+\infty} \tau|\phi_a(\tau)|^2 \, d\tau = -\zeta_a \int_{-\infty}^{+\infty} \frac{1}{b_a(\tau)} |\phi_a(\tau)|^2 \, d\tau,$$

$$\int_{-\infty}^{+\infty} \tau|\phi_a'(\tau)|^2 \, d\tau = \beta_a\zeta_a \int_{-\infty}^{+\infty} \frac{1}{b_a(\tau)} |\phi_a(\tau)|^2 \, d\tau + 2M_3(a) - 2\zeta_a M_2(a).$$

We will also encounter the moment

$$I_2(a) := \int_{\mathbb{R}} (\zeta_a + b_a(\tau)\tau)\phi_a \mathfrak{R}_a[(\zeta_a + b_a(\tau)\tau)\phi_a] \, d\tau, \tag{2-17}$$

involving the resolvent $\mathfrak{R}_a$, which is an operator defined on $L^2(\mathbb{R})$ by means of the following lemma:

Lemma 2.4. *If $u \in L^2(\mathbb{R})$ is orthogonal to ϕ_a, we define $(\mathfrak{h}_a[\zeta_a] - \beta_a)^{-1}u$ in $L^2(\mathbb{R})$ as the unique solution v orthogonal to ϕ_a to*

$$(\mathfrak{h}_a[\zeta_a] - \beta_a)v = u.$$

We introduce the regularized resolvent $\mathfrak{R}_a$ *in* $\mathcal{L}(L^2(\mathbb{R}))$ *by*

$$\mathfrak{R}_a(u) = \begin{cases} 0 & \text{if } u \parallel \phi_a, \\ (\mathfrak{h}_a[\zeta_a] - \beta_a)^{-1} u & \text{if } u \perp \phi_a \end{cases} \tag{2-18}$$

(*extended by linearity*). *Then, for any* $\gamma \geq 0$, $\mathfrak{R}_a$ *and* $(d/d\tau) \circ \mathfrak{R}_a$ *are two bounded operators on* $L^2(\mathbb{R}, \exp(\gamma|\tau|)\, d\tau)$.

Proof. We follow Agmon's approach. Consider $v \in \mathrm{Dom}(\mathfrak{h}_a[\zeta_a])$ and $u \in L^2(\mathbb{R}, \exp(\gamma|\tau|)\, d\tau)$ such that

$$(\mathfrak{h}_a[\zeta_a] - \beta_a)v = u.$$

For all $\gamma > 0$ and $N > 1$, consider the continuous function on $\mathbb{R}$

$$\Phi_{\gamma,N}(\tau) = \min(\gamma|\tau|, N).$$

Observe that $\Phi_{\gamma,N} \in H^1_{\mathrm{loc}}(\mathbb{R})$ and

$$|\Phi'_{\gamma,N}(\tau)| = \begin{cases} \gamma & \text{if } \gamma|\tau| < N, \\ 0 & \text{if } \gamma|\tau| > N. \end{cases}$$

Integration by parts yields

$$\begin{aligned} \langle u, e^{2\Phi_{\gamma,N}} v\rangle &= \langle (\mathfrak{h}_a[\zeta_a] - \beta_a)v, e^{2\Phi_{\gamma,N}} v\rangle \\ &= \|(e^{\Phi_{\gamma,N}} v)'\|^2 + \int_{\mathbb{R}} ((\zeta_a + b\tau)^2 - \beta_a)|e^{\Phi_{\gamma,N}} v|^2\, d\tau - \|\Phi'_{\gamma,N} e^{\Phi_{\varepsilon,N}} v\|^2 \\ &\geq \|(e^{\Phi_{\gamma,N}} v)'\|^2 + \int_{\mathbb{R}} ((\zeta_a + b\tau)^2 - \beta_a - \gamma^2)|e^{\Phi_{\gamma,N}} v|^2\, d\tau. \end{aligned}$$

Choose $A_\gamma > 1$ so that, for $|\tau| \geq A_\gamma$, we have $(\zeta_a + b\tau)^2 - \beta_a - \gamma^2 \geq 1$; consequently, for $N \geq \gamma A_\gamma$,

$$\langle u, e^{2\Phi_{\gamma,N}} v\rangle \geq \|(e^{\Phi_{\gamma,N}} v)'\|^2 + \int_{\{|\tau| \geq A_\gamma\}} |e^{\Phi_{\gamma,N}} v|^2\, d\tau - (\beta_a + \gamma^2) e^{2\gamma A_\gamma} \|v\|^2.$$

Using the Cauchy–Schwarz inequality, we get further

$$\|e^{\Phi_{\gamma,N}} u\|\, \|e^{\Phi_{\gamma,N}} v\| \geq \|(e^{\Phi_{\gamma,N}} v)'\|^2 + \int_{\{|\tau| \geq A_\gamma\}} |e^{\Phi_{\gamma,N}} v|^2\, d\tau - (\beta_a + \gamma^2) e^{2\gamma A_\gamma} \|v\|^2.$$

Rearranging the terms in (2-19) and using Cauchy's inequality

$$\|e^{\Phi_{\gamma,N}} u\|\, \|e^{\Phi_{\gamma,N}} v\| \leq 2\|e^{\Phi_{\gamma,N}} u\|^2 + \tfrac{1}{2}\|e^{\Phi_{\gamma,N}} v\|^2,$$

we get

$$\|(e^{\Phi_{\gamma,N}} v)'\|^2 + \frac{1}{2} \int_{\{|\tau| \geq A_\gamma\}} |e^{\Phi_{\gamma,N}} v|^2\, d\tau \leq (\beta_a + \gamma^2 + 1) e^{2\gamma A_\gamma} \|v\|^2 + 2\|e^{\Phi_{\gamma,N}} u\|^2.$$

We end up with the estimate

$$\int |e^{\Phi_{\gamma,N}} v'|^2\, d\tau + \int |e^{\Phi_{\gamma,N}} v|^2\, d\tau \leq C_\gamma(\|v\|^2 + \|e^{\Phi_\gamma} u\|^2),$$

where we note that the right-hand side is independent of N.

Since $\Phi_{\gamma,N}$ is nonnegative and monotone increasing with respect to N, we get by monotone convergence that $e^{\Phi_\gamma} v$ and $e^{\Phi_\gamma} v'$ belong to $L^2(\mathbb{R})$ and satisfy

$$\int |e^{\Phi_\gamma} v'|^2 \, d\tau + \int |e^{\Phi_\gamma} v|^2 \, d\tau \le C_\gamma (\|v\|^2 + \|e^{\Phi_\gamma} u\|^2), \tag{2-19}$$

where

$$\Phi_\gamma(\tau) = \lim_{N\to+\infty} \Phi_{\gamma,N}(\tau) = \gamma|\tau|.$$

To finish the proof, we note that, since the regularized resolvent is bounded and $\Phi_\gamma \ge 0$,

$$\|v\|^2 = \|\mathfrak{R}_a u\|^2 \le \|\mathfrak{R}_a\|^2 \|u\|^2 \le \|\mathfrak{R}_a\|^2 \|e^{\Phi_\gamma} u\|^2. \qquad \square$$

Proposition 2.5. *For any $a \in (-1, 0)$, it holds*

$$\mu_a''(\zeta_a) = 2(1 - 4I_2(a)) > 0. \tag{2-20}$$

Proof. First we notice $(\zeta_a + b_a(\tau)\tau)\phi_a$ is orthogonal to ϕ_a in $L^2(\mathbb{R})$ (see (2-10)). Thus $\mathfrak{R}_a[(\zeta_a + b_a(\tau)\tau)\phi_a]$ is well-defined as $(\mathfrak{h}_a[\zeta_a] - \beta_a)^{-1}(\zeta_a + b_a(\tau)\tau)\phi_a$. Let $z \in \mathbb{R}$, and $E_a(z)$ be the lowest eigenvalue of the operator $H_a(z)$, defined on $L^2(\mathbb{R})$ as

$$H_a(z) := \mathfrak{h}_a[\zeta_a + z] = -\frac{d^2}{d\tau^2} + (\zeta_a + z + b_a(\tau)\tau)^2.$$

We adopt the same proof of [Fournais and Helffer 2006, Proposition A.3] (replacing P_0 by $H_a(0) - \beta_a$ there) to get the identity in (2-20). Finally, by [Assaad and Kachmar 2022], $\mu''(\zeta_a) > 0$. $\square$

3. One-dimensional model involving the curvature

We consider a new family of fiber operators which are obtained by adding to the fiber operators in Section 2 new terms that will be related to the geometry of the magnetic edge. This family was introduced earlier in [Assaad and Kachmar 2022] and their definition is reminiscent of the weighted operators introduced in the context of the Neumann Laplacian with a uniform magnetic field [Helffer and Morame 2001].

We introduce the parameters

$$a \in (-1, 0), \quad \delta \in \left(0, \tfrac{1}{12}\right), \quad M > 0, \quad h_0 > 0 \quad \text{and} \quad \kappa \in [-M, M],$$

which satisfy

$$M h_0^{1/2-\delta} < \tfrac{1}{3},$$

and will be fixed throughout this section.

Consider on $(-h^{-\delta}, h^{-\delta})$ the positive function $a_{\kappa,h}(\tau) = (1 - \kappa h^{1/2}\tau)$, the Hilbert space $L^2((-h^{-\delta}, h^{-\delta}); a_{\kappa,h}\, d\tau)$ with the inner product

$$\langle u, v\rangle = \int_{-h^{-\delta}}^{h^{-\delta}} u(\tau)\overline{v(\tau)}\,(1 - \kappa h^{1/2}\tau)\, d\tau,$$

and, for $\xi \in \mathbb{R}$, the operator

$$\mathcal{H}_{a,\xi,\kappa,h} = -\frac{d^2}{d\tau^2} + (b_a(\tau)\tau + \xi)^2 + \kappa h^{1/2}(1 - \kappa h^{1/2}\tau)^{-1}\partial_\tau + 2\kappa h^{1/2}\tau\left(b_a(\tau)\tau + \xi - \kappa h^{1/2} b_a(\tau)\frac{\tau^2}{2}\right)^2$$
$$- \kappa h^{1/2} b_a(\tau)\tau^2(b_a(\tau)\tau + \xi) + \kappa^2 h b_a(\tau)^2 \frac{\tau^4}{4}, \tag{3-1}$$

where b_a is the function in (2-2) and

$$\mathrm{Dom}(\mathcal{H}_{a,\xi,\kappa,h}) = \{u \in H^2(-h^{-\delta}, h^{-\delta}) : u(\pm h^{-\delta}) = 0\}. \tag{3-2}$$

The operator $\mathcal{H}_{a,\xi,\kappa,h}$ is a self-adjoint operator in $L^2((-h^{-\delta}, h^{-\delta}); a_{\kappa,h}\, d\tau)$ with compact resolvent. We denote by $(\lambda_n(\mathcal{H}_{a,\xi,\kappa,h}))_{n\geq 1}$ its sequence of min-max eigenvalues. The first eigenvalue can be expressed as

$$\lambda_1(\mathcal{H}_{a,\xi,\kappa,h}) = \inf\{q_{a,\xi,\kappa,h}(u) : u \in H_0^1(-h^{-\delta}, h^{-\delta}) \text{ and } \|u\|_{L^2((-h^{-\delta},h^{-\delta});a_{\kappa,h}\,d\tau)} = 1\}, \tag{3-3}$$

where

$$q_{a,\xi,\kappa,h}(u) = \int_{-h^{-\delta}}^{h^{-\delta}} \Big(|u'(\tau)|^2 + (1+2\kappa h^{1/2}\tau)\Big(b_a(\tau)\tau + \xi - \kappa h^{1/2} b_a(\tau)\frac{\tau^2}{2}\Big)^2 u^2(\tau)\Big)(1-\kappa h^{1/2}\tau)\, d\tau. \tag{3-4}$$

By Cauchy's inequality, we write, for any $\varepsilon \in (0, 1)$,

$$\Big(b_a(\tau)\tau + \xi - \kappa h^{1/2} b_a(\tau)\frac{\tau^2}{2}\Big)^2 \geq (1-\varepsilon)(b_a(\tau)\tau + \xi)^2 - \varepsilon^{-1}\kappa^2 h b_a(\tau)^2 \frac{\tau^4}{4}.$$

Noticing that $h\tau^4 \leq h^{1-4\delta}$ for $\tau \in (h^{-\delta}, h^{\delta})$ and optimizing with respect to ε, we choose $\varepsilon = h^{1/2-2\delta}$ and get

$$\Big(b_a(\tau)\tau + \xi - \kappa h^{1/2} b_a(\tau)\frac{\tau^2}{2}\Big)^2 \geq (1-h^{1/2-2\delta})(b_a(\tau)\tau + \xi)^2 - \kappa^2 b_a(\tau)^2 h^{1/2-2\delta}. \tag{3-5}$$

We plug (3-5) in (3-4) to get, for some $C_0 > 0$,

$$q_{a,\xi,\kappa,h}(u) \geq (1 - C_0 h^{1/2-2\delta}) q_a[\xi](u) - C_0 h^{1/2-2\delta} \|u\|^2_{L^2(-h^{-\delta},h^{-\delta})}, \tag{3-6}$$

where $q_a[\xi]$ is the quadratic form in (2-4). The min-max principle ensures that

$$q_a[\xi](u) \geq \beta_a \|u\|^2_{L^2(-h^{-\delta},h^{-\delta})} \quad \text{for all } u \in H_0^1(-h^{-\delta}, h^{-\delta}). \tag{3-7}$$

Since $\beta_a > 0$, (3-6) and (3-7) imply

$$q_{a,\xi,\kappa,h}(u) \geq (1 - C h^{1/2-2\delta}) q_a[\xi](u), \tag{3-8}$$

with $C = (1+\beta_a^{-1})C_0$. From (3-8) and the min-max principle we deduce the lower bounds in Lemma 3.1 below (see [Assaad and Kachmar 2022, Section 4.2] for details).

Lemma 3.1. *Given $a \in (-1, 0)$, there exist positive constants $\varepsilon_0(a), \varepsilon_1(a), \varepsilon_2(a), c_0(a), h_0(a), C_0(a)$ such that, for all $h \in (0, h_0(a))$,*

- *For $|\xi - \zeta_a| \geq \varepsilon_0(a)$, we have*
$$\lambda_1(\mathcal{H}_{a,\xi,\kappa,h}) \geq \beta_a + c_0(a).$$
- *For $\varepsilon_2(a)h^{1/4-\delta} \leq |\xi - \zeta_a| \leq \varepsilon_0(a)$, we have*
$$\lambda_1(\mathcal{H}_{a,\xi,\kappa,h}) \geq \beta_a + \varepsilon_1(a)(\xi - \zeta_a)^2.$$
- *For $|\xi - \zeta_a| \leq \varepsilon_2(a)h^{1/4-\delta}$, we have*
$$\lambda_1(\mathcal{H}_{a,\xi,\kappa,h}) \geq \beta_a + c_2(a)|\xi - \zeta_a|^2 + \kappa M_3(a)h^{1/2} - C_0(a)\max(h^{1/2}|\xi - \zeta_a|, |\xi - \zeta_a|^3, h),$$

where

$$c_2(a) = \tfrac{1}{2}\mu_a''(\zeta_a) > 0. \tag{3-9}$$

We can now state the following:

Proposition 3.2. *There exists $\hat{c}_0(a) > 0$ and, for all $\varepsilon \in (0, 1)$, there exist $C_\varepsilon, h_\varepsilon > 0$ such that, for all $h \in (0, h_\varepsilon)$ and $\xi \in \mathbb{R}$, the following inequality holds*:

$$\lambda_1(\mathcal{H}_{a,\xi,\kappa,h}) \geq \beta_a + \hat{c}_0(a)\min((\xi-\zeta_a)^2, \varepsilon) + \kappa M_3(a)h^{1/2} - C_\varepsilon h.$$

Proof. In the third item of Lemma 3.1, we estimate the remainder term

$$\max(h^{1/2}|\xi-\zeta_a|, |\xi-\zeta_a|^3, h) \leq (\eta^{-1}+1)h + \eta|\xi-\zeta_a|^2 + |\xi-\zeta_a|^3$$

for all $\eta \in (0, 1)$. Choosing $\eta = c_2(a)/(4C_0(a))$, where $C_0(a)$ is the constant in Lemma 3.1, we deduce from Lemma 3.1 the lower bound for the eigenvalue $\lambda_1(\mathcal{H}_{a,\xi,\kappa,h})$, with

$$\hat{c}_0(a) = \frac{1}{2}\min\left(\varepsilon_1(a), \frac{c_0(a)}{\varepsilon_0(a)^2}, c_0(a)\right). \qquad \square$$

4. Upper bound

We establish an upper bound of the n-th eigenvalue $\lambda_n(h)$ of $\mathcal{P}_h$, which was defined in (1-4). This will involve the spectral value β_a introduced in (2-6), the moment $M_3(a) < 0$ introduced in (2-16), and $c_2(a) > 0$ the value defined in (3-9). In this section, we consider two parameters $\eta \in \left(0, \frac{1}{8}\right)$ and $\delta \in \left(0, \frac{1}{2}\right)$.

Theorem 4.1. *Let $n \in \mathbb{N}^*$ and $\boldsymbol{a} = (1, a)$, with $-1 < a < 0$. Under Assumption 1.1, there exist $h_0 > 0$ and $C_0 > 0$ such that, for all $h \in (0, h_0)$, the n-th eigenvalue $\lambda_n(h)$ of the operator $\mathcal{P}_h$ defined in (1-4) satisfies*

$$\lambda_n(h) \leq h\beta_a + h^{3/2}k_{\max}M_3(a) + h^{7/4}(2n-1)\sqrt{\frac{k_2 M_3(a)c_2(a)}{2}} + C_0 h^{15/8}, \tag{4-1}$$

where $c_2(a)$ and $M_3(a)$ were introduced in (1-12).

Proof. The approach is similar to the one used in the literature in establishing upper bounds for the low-lying eigenvalues of operators defined on smooth bounded domains, like Schrödinger operators with uniform magnetic fields (and Neumann boundary conditions) or the Laplacian (with Robin boundary conditions). For instance, one can see [Bernoff and Sternberg 1998; Fournais and Helffer 2006; Helffer and Kachmar 2017]. The proof relies on the construction of quasimodes localized near the point of maximal curvature on Γ.

Let $h \in (0, 1)$. Working near Γ, we start by expressing the operator $\mathcal{P}_h$ in the adapted (s, t)-coordinates there (see Appendix A):

$$\widetilde{\mathcal{P}}_h = -\mathfrak{a}^{-1}(h\partial_s - i\widetilde{F}_1)\mathfrak{a}^{-1}(h\partial_s - i\widetilde{F}_1) - \mathfrak{a}^{-1}(h\partial_t - i\widetilde{F}_2)\mathfrak{a}(h\partial_t - i\widetilde{F}_2). \tag{4-2}$$

Recall that we assume that the maximum is attained for $s = 0$, hence $k_{\max} = k(0)$, and having Lemma A.1, we perform a global change of gauge ω such that the magnetic potential $\boldsymbol{F}$ satisfies in Ω near the edge Γ,

when expressed in the (s,t)-coordinates,

$$\widetilde{\boldsymbol{F}}(s,t)=\begin{pmatrix}-b_a(t)\big(t-\frac{1}{2}t^2k(s)\big)\\ 0\end{pmatrix}, \tag{4-3}$$

where $t\mapsto b_a(t)$ is defined by

$$b_a(t)=\mathbb{1}_{\mathbb{R}_+}(t)+a\mathbb{1}_{\mathbb{R}_-}(t),\quad t\in\mathbb{R}.$$

Performing the change of variables

$$\sigma=h^{-1/8}s\quad\text{and}\quad\tau=h^{-1/2}t,$$

the operator $\widetilde{\mathcal{P}}_h$ becomes in the (σ,τ)-coordinates

$$\check{\mathcal{P}}_h=-\check{\mathfrak{a}}^{-1}(h^{7/8}\partial_\sigma+ih^{1/2}b_a(\tau)\tau\check{\mathfrak{a}}_2)\check{\mathfrak{a}}^{-1}(h^{7/8}\partial_\sigma+ih^{1/2}b_a(\tau)\tau\check{\mathfrak{a}}_2)-h\check{\mathfrak{a}}^{-1}\partial_\tau\check{\mathfrak{a}}\partial_\tau, \tag{4-4}$$

with

$$\check{\mathfrak{a}}(\sigma,\tau;h)=1-h^{1/2}\tau k(h^{1/8}\sigma)\quad\text{and}\quad\check{\mathfrak{a}}_2(\sigma,\tau;h)=1-\tfrac{1}{2}h^{1/2}\tau k(h^{1/8}\sigma). \tag{4-5}$$

It is convenient to introduce the operator

$$\mathcal{P}_h^{\text{new}}=e^{-i\sigma\zeta_a/h^{3/8}}h^{-1}\check{\mathcal{P}}_he^{i\sigma\zeta_a/h^{3/8}}-\beta_a, \tag{4-6}$$

where ζ_a is introduced in Section 2B and we get

$$\begin{aligned}\mathcal{P}_h^{\text{new}}=-\check{\mathfrak{a}}^{-1}\partial_\tau\check{\mathfrak{a}}\partial_\tau-\beta_a-\check{\mathfrak{a}}^{-1}\big(h^{3/8}\partial_\sigma+i(\zeta_a+b_a(\tau)\tau)-ib_a(\tau)\tau(1-\check{\mathfrak{a}}_2)\big)\\ \times\check{\mathfrak{a}}^{-1}\big(h^{3/8}\partial_\sigma+i(\zeta_a+b_a(\tau)\tau)-ib_a(\tau)\tau(1-\check{\mathfrak{a}}_2)\big).\end{aligned}$$

Using the boundedness and the smoothness of k, and the fact that $k'(0)=0$ and $k''(0)<0$, we write

$$\begin{aligned}\check{\mathfrak{a}}(\sigma,\tau;h)&=1-h^{1/2}\tau k(0)-h^{3/4}\tau\sigma^2\frac{k''(0)}{2}+h^{7/8}e_{1,h}(\sigma,\tau),\\ \check{\mathfrak{a}}_2(\sigma,\tau;h)&=1-h^{1/2}\tau\frac{k(0)}{2}-h^{3/4}\tau\sigma^2\frac{k''(0)}{4}+h^{7/8}e_{2,h}(\sigma,\tau),\\ \check{\mathfrak{a}}^{-1}(\sigma,\tau;h)&=1+h^{1/2}\tau k(0)+h^{3/4}\tau\sigma^2\frac{k''(0)}{2}+h^{7/8}e_{3,h}(\sigma,\tau),\\ \check{\mathfrak{a}}^{-2}(\sigma,\tau;h)&=1+2h^{1/2}\tau k(0)+h^{3/4}\tau\sigma^2k''(0)+h^{7/8}e_{4,h}(\sigma,\tau),\end{aligned}$$

where $(e_{i,h})_{i=1,\dots,4}$ are functions of σ and τ having the property that there exist C and h_0 such that,[1] for $h\in(0,h_0)$, $\sigma\in(-h^{-\eta},h^{-\eta})$ and $\tau\in(-h^{-\rho},h^{-\rho})$ we have

$$|e_{1,h}(\sigma,\tau)|+|e_{2,h}(\sigma,\tau)|\le C|\tau\sigma^3|,\quad |e_{3,h}(\tau,\sigma)|+|e_{4,h}(\tau,\sigma)|\le C(\sigma^6+\tau^4+1), \tag{4-7}$$

and

$$\sum_{i=1}^4\bigg(\sum_{j=1}^2\big(|\partial_\tau^je_{i,h}(\sigma,\tau)|+|\partial_\sigma^je_{i,h}(\sigma,\tau)|\big)+|\partial_{\sigma\tau}^2e_{i,h}(\sigma,\tau)|\bigg)\le C(|\sigma|^5+|\tau|^3+1). \tag{4-8}$$

[1]The following conditions on the length scales of τ and σ (namely that $\sigma\in(-h^{-\delta},h^{-\delta})$ and $\tau\in(-h^{-\rho},h^{-\rho})$), as well as (4-7) and (4-8) below, are set for a later use in the paper.

Hence,

$$\mathcal{P}_h^{\text{new}} = P_0 + h^{3/8}P_1 + h^{1/2}P_2 + h^{3/4}P_3 + h^{7/8}Q_h, \tag{4-9}$$

where

$$\begin{aligned}
P_0 &= -\partial_\tau^2 + (\zeta_a + b_a(\tau)\tau)^2 - \beta_a,\\
P_1 &= -2i(\zeta_a + b_a(\tau)\tau)\partial_\sigma,\\
P_2 &= k(0)[2\tau(\zeta_a + b_a(\tau)\tau)^2 - b_a(\tau)\tau^2(\zeta_a + b_a(\tau)\tau)] + k(0)\partial_\tau,\\
P_3 &= -\partial_\sigma^2 + \frac{k''(0)}{2}\sigma^2[2\tau(\zeta_a + b_a(\tau)\tau)^2 - b_a(\tau)\tau^2(\zeta_a + b_a(\tau)\tau)] + \frac{k''(0)}{2}\sigma^2\partial_\tau,
\end{aligned} \tag{4-10}$$

and

$$Q_h = \mathcal{E}_{1,h}(\sigma,\tau)\partial_\sigma^2 + \mathcal{E}_{2,h}(\sigma,\tau)\partial_\sigma + \mathcal{E}_{3,h}(\sigma,\tau)\partial_\tau + \mathcal{E}_{4,h}(\sigma,\tau). \tag{4-11}$$

Here the terms $(\mathcal{E}_{i,h})_{i=1,\dots,4}$ are functions in σ and τ having the property that there exist C and h_0 such that, for $h \in (0, h_0)$, $\sigma \in (-h^{-\eta}, h^{-\eta})$ and $\tau \in (-h^{-\rho}, h^{-\rho})$, we have

$$|\mathcal{E}_{i,h}(\sigma,\tau)| + |\partial_\sigma \mathcal{E}_{i,h}(\sigma,\tau)| + |\partial_\tau \mathcal{E}_{i,h}\sigma,\tau)| \le C\,(|\sigma|^6 + |\tau|^6 + 1). \tag{4-12}$$

In what follows, we will construct, for each $n \in \mathbb{N}^*$, a trial function $\phi_n \in \operatorname{Dom}\mathcal{P}_h^{\text{new}}$ satisfying

$$\begin{aligned}
\left\| \mathcal{P}_h^{\text{new}}\phi_n - \left(h^{1/2}k_{\max}M_3(a) + h^{3/4}(2n-1)\sqrt{\frac{k_2 M_3(a)c_2(a)}{2}}\right)\phi_n \right\|_{L^2(\mathbb{R}^2, h^{5/8}\tilde{\mathfrak{a}}\,d\sigma\,d\tau)}\\
= \mathcal{O}(h^{7/8})\|\phi_n\|_{L^2(\mathbb{R}^2, h^{5/8}\tilde{\mathfrak{a}}\,d\sigma\,d\tau)}
\end{aligned} \tag{4-13}$$

(recall $k_2 = k''(0)$).

The result in (4-13), once established, will imply by the spectral theorem the existence of an eigenvalue $\lambda_n^{\text{new}}(h)$ of $\mathcal{P}_h^{\text{new}}$ such that

$$\lambda_n^{\text{new}}(h) = h^{1/2}k_{\max}M_3(a) + h^{3/4}(2n-1)\sqrt{\frac{k_2 M_3(a)c_2(a)}{2}} + \mathcal{O}(h^{7/8}). \tag{4-14}$$

Furthermore, by the definition of $\mathcal{P}_h^{\text{new}}$ in (4-6) we have

$$\sigma(\mathcal{P}_h) = h\,\sigma(\mathcal{P}_h^{\text{new}}).$$

Thus, (4-14) will yield the result in (4-1). Hence, the discussion above shows that establishing (4-13) is sufficient to complete the proof of the theorem.

We construct the trial functions in the form

$$\phi_h(\sigma,\tau) = h^{-5/16}\chi(h^\eta\sigma)\chi(h^\rho\tau)g(\sigma,\tau), \tag{4-15}$$

where χ is a smooth cut-off function supported in $(-1, 1)$ and $g = g[h]$ will be determined in $L^2(\mathbb{R}^2)$ with rapid decay at infinity. First we set

$$g[h] = g_0 + h^{3/8}g_1 + h^{1/2}g_2 + h^{3/4}g_3, \tag{4-16}$$

with $g_i \in L^2(\mathbb{R}^2)$ for $i = 0, \dots, 3$, and

$$\mu = \mu(h) = \mu_0 + h^{3/8}\mu_1 + h^{1/2}\mu_2 + h^{3/4}\mu_3, \tag{4-17}$$

with $\mu_i \in \mathbb{R}$ for $i = 0, \dots, 3$. We will search for μ and g satisfying on $\mathbb{R}^2$

$$(\mathcal{P}_h^{\text{new}} - \mu)g = \mathcal{O}(h^{7/8}). \tag{4-18}$$

More precisely, using the expansion of $\mathcal{P}_h^{\text{new}}$ in (4-9), we will search for μ_i and g_i satisfying the system of equations

$$\begin{cases} (e_0)\colon\ (P_0 - \mu_0)g_0 = 0, \\ (e_1)\colon\ (P_0 - \mu_0)g_1 + (P_1 - \mu_1)g_0 = 0, \\ (e_2)\colon\ (P_0 - \mu_0)g_2 + (P_2 - \mu_2)g_0 = 0, \\ (e_3)\colon\ (P_0 - \mu_0)g_3 + (P_1 - \mu_1)g_1 + (P_3 - \mu_3)g_0 = 0. \end{cases}$$

Let $u_0 = \phi_a$ be the positive normalized eigenfunction of the operator $\mathfrak{h}_a[\zeta_a]$ (in (2-1)) corresponding to the lowest eigenvalue β_a.

Obviously, the pair

$$(\mu_0, g_0) = (0, u_0 f) \tag{4-19}$$

is a solution of (e_0) for any $f \in \mathcal{S}(\mathbb{R}_\sigma)$.

We implement this choice of (μ_0, g_0) in (e_1) and write

$$P_0 g_1 = -(P_1 - \mu_1)g_0 = [2i(\zeta_a + b_a(\tau)\tau)\partial_\sigma + \mu_1]u_0 f.$$

Noticing that $(\zeta_a + b_a(\tau)\tau)u_0$ is orthogonal to u_0 in $L^2(\mathbb{R})$, $\mathfrak{R}_a[(\zeta_a + b_a(\tau)\tau)u_0]$ is well-defined with $\mathfrak{R}_a$ in (2-18) (see (2-11) and Remark 2.2), and the pair

$$(\mu_1, g_1) = \big(0, 2i\mathfrak{R}_a[(\zeta_a + b_a(\tau)\tau)u_0]\partial_\sigma f\big) \tag{4-20}$$

is a solution of (e_1).

Similarly,

$$P_0 g_2 = -(P_2 - \mu_2)g_0 = \big[-k_{\max}\big(2\tau(\zeta_a + b_a(\tau)\tau)^2 - b_a(\tau)\tau^2(\zeta_a + b_a(\tau)\tau)\big) + \mu_2\big]u_0 f - k_{\max} f \partial_\tau u_0.$$

From Remark 2.3, we observe that $[2\tau(\zeta_a + b_a(\tau)\tau)^2 - b_a(\tau)\tau^2(\zeta_a + b_a(\tau)\tau) - M_3(a)]u_0$ is orthogonal to u_0 in $L^2(\mathbb{R})$. Moreover, the normalization of u_0 in $L^2(\mathbb{R})$ yields $\partial_\tau u_0 \perp u_0$. Hence, the pair

$$\begin{aligned}(\mu_2, g_2) = \big(&k_{\max} M_3(a), \\ &-k_{\max}\mathfrak{R}_a([2\tau(\zeta_a + b_a(\tau)\tau)^2 - b_a(\tau)\tau^2(\zeta_a + b_a(\tau)\tau) - M_3(a)]u_0 + \partial_\tau u_0) f\big)\end{aligned} \tag{4-21}$$

is a solution of equation (e_2).

Finally, we consider equation (e_3):

$$P_0 g_3 = -P_1 g_1 - (P_3 - \mu_3)g_0.$$

We will search for μ_3 and f satisfying

$$\big(P_1 g_1(\sigma, \cdot) + (P_3 - \mu_3)g_0(\sigma, \cdot)\big) \perp u_0(\cdot) \tag{4-22}$$

for every fixed σ. This orthogonality result will allow us to choose

$$g_3(\sigma,\cdot) = -\mathfrak{R}_a[P_1 g_1(\sigma,\cdot) + (P_3-\mu_3)g_0(\sigma,\cdot)] \tag{4-23}$$

in order to satisfy (e_3). To that end, the aforementioned choice of g_0, g_1 and g_2 gives for any fixed σ

$$\begin{aligned}
&\langle P_1 g_1(\sigma,\cdot) + (P_3-\mu_3)g_0(\sigma,\cdot), u_0(\cdot)\rangle_{L^2(\mathbb{R})}\\
&= 4\partial_\sigma^2 f(\sigma)\int_{\mathbb{R}} (\zeta_a + b_a(\tau)\tau)u_0\mathfrak{R}_a[(\zeta_a+b_a(\tau)\tau)u_0]\,d\tau + \frac{k_2}{2}\sigma^2 f(\sigma)\int_{\mathbb{R}} u_0\partial_\tau u_0\,d\tau\\
&\qquad + \int_{\mathbb{R}}\Big(-\partial_\sigma^2 f(\sigma) + \frac{k_2}{2}\sigma^2 f(\sigma)[2\tau(\zeta_a+b_a(\tau)\tau)^2 - b_a(\tau)\tau^2(\zeta_a+b_a(\tau)\tau)] - \mu_3 f(\sigma)\Big)u_0^2\,d\tau\\
&= -(1-4I_2(a))\partial_\sigma^2 f(\sigma) + \frac{k_2 M_3(a)}{2}\sigma^2 f(\sigma) - \mu_3 f(\sigma) \quad (\text{using } \|u_0\|_{L^2(\mathbb{R})}=1)\\
&= -c_2(a)\partial_\sigma^2 f(\sigma) + \frac{k_2 M_3(a)}{2}\sigma^2 f(\sigma) - \mu_3 f(\sigma),
\end{aligned} \tag{4-24}$$

where $I_2(a)$ is introduced in (2-17) and (2-20), and $c_2(a)$ is introduced in (1-12).

We consider the harmonic oscillator on $\mathbb{R}$

$$H_a^{\text{harm}} := -c_2(a)\frac{d^2}{d\sigma^2} + \frac{1}{2}k_2 M_3(a)\sigma^2. \tag{4-25}$$

For each $n\in\mathbb{N}^*$, let $f_n\in\mathcal{S}(\mathbb{R})$ be the n-th normalized eigenfunction of H_a^{harm} corresponding to the eigenvalue $(2n-1)\sqrt{k_2M_3(a)c_2(a)/2}$. The choice

$$f = f_n \quad\text{and}\quad \mu_3 = (2n-1)\sqrt{\frac{k_2M_3(a)c_2(a)}{2}} \tag{4-26}$$

makes the expression in (4-24) equal to zero, hence realizing the orthogonality result in (4-22).

We can now gather the above results. For each $n\in\mathbb{N}^*$, we choose μ in (4-17) and $g=g_{(n)}$ in (4-16) such that μ_i, g_i and f are as in (4-19)–(4-21), (4-23) and (4-26).

For h sufficiently small, using the properties of Q_h in (4-11) and (4-12), the fact that $f\in\mathcal{S}(\mathbb{R})$, the decay properties of ϕ_a in Proposition 2.1 and those of the resolvent $\mathfrak{R}_a$ in (2-18), the foregoing choice of g and μ implies (4-18).

Now, we consider the trial function (see (4-15)) associated with $g_{(n)}$. Using again the decay properties of u_0 and f, and Lemma 2.4 for getting the same properties for the g_j, one can neglect the effect of the cut-off functions in the computation while concluding from (4-18) the desired result in (4-13). We omit further details of the computation, and refer the reader to [Fournais and Helffer 2006, Sections 2–3]. □

Remark 4.2. The formal construction of the pairs $(\mu_i, g_i)_{i=0,\dots,3}$ in the proof of Theorem 4.1 can be pushed to any order, assuming that the curve Γ is C^∞ smooth. Using the same approach we can construct pairs $(\mu_i, g_i)_{i\in\mathbb{N}^*}$ for defining quasimodes yielding an accurate upper bound of the eigenvalue $\lambda_n(h)$, which is an infinite expansion of powers of $h^{1/8}$. This upper bound will agree with the one in Theorem 4.1 up to the order $h^{7/4}$; see [Bernoff and Sternberg 1998; Fournais and Helffer 2006; Helffer and Kachmar 2017].

Remark 4.3. In the derivation of the lower bound in Section 7, the operator H_a^{harm} introduced in (4-25) plays the role of an effective operator in the tangential variable. In light of (4-16), (4-19), (4-20), (4-21) and (4-26), the quasimode

$$v_{h,n}^{\mathrm{app}} = \phi_a(\tau) f_n(\sigma) + 2ih^{3/8}\mathfrak{R}_a\big((\zeta_a + b_a(\tau)\tau)\phi_a(\tau)\big)\partial_\sigma f_n(\sigma) + h^{1/2} g_2(\sigma, \tau)$$

is a candidate for the profile of an actual eigenfunction of the operator $\mathcal{P}_h$, after rescaling and a gauge transformation.

5. Functions localized near the magnetic edge

In this section, we consider functions satisfying the energy bound[2] in (5-1), which are consequently localized near the maximum of the curvature of the magnetic edge Γ. We will be able to estimate the tangential derivative of such functions.

As we shall see in Section 5A, bound states and their first-order tangential derivatives are examples of the functions we discuss in this section.

5A. ***Localization hypotheses.*** We fix $t_0 > 0$ so that the Frenet coordinates recalled in Appendix A are valid in $\{d(x, \Gamma) < t_0\}$. We recall our assumption that the curvature of Γ attains its maximum at a unique point defined by the tangential coordinate $s = 0$.

Let $\theta \in \big(0, \frac{3}{8}\big)$ be a fixed constant. Consider a family of functions $(g_h)_{h\in(0,h_0]}$ in $H^1(\Omega)$ for which there exist positive constants C_1, C_2 such that, for $h \in (0, h_0]$,

$$\mathcal{Q}_h(g_h) \le (h\beta_a + h^{3/2} M_3(a) k_{\max} + C_1 h^{7/4}) \|g_h\|^2_{L^2(\Omega)} + C_2 h^{5/2-\theta}, \tag{5-1}$$

where $\mathcal{Q}_h$ is the quadratic form introduced in (1-3).

Suppose also that there exist constants α, $C > 0$ and a family $(r_h)_{h\in(0,h_0]} \subset \mathbb{R}_+$ such that

$$\limsup_{h\to 0_+} r_h < +\infty, \tag{5-2}$$

and the following two estimates hold:

$$\int_\Omega \big(|g_h|^2 + h^{-1}|(h\nabla - i\boldsymbol{F})g_h|^2\big) \exp(\alpha h^{-1/2} d(x, \Gamma))\, dx \le C r_h, \tag{5-3}$$

$$\int_{d(x,\Gamma)\le t_0} \big(|g_h(x)|^2 + h^{-1}|(h\nabla - i\boldsymbol{F})g_h|^2\big) \exp(\alpha h^{-1/8}|s(x)|)\, dx \le C r_h. \tag{5-4}$$

We can derive from the decay estimates in (5-3) and (5-4) four estimates.

The two first estimates follow from the inequality $e^z \ge z^N/N!$ for $z \ge 0$ and read: for $N \ge 1$, there exist C_N, $h_N > 0$ such that, for all $h \in (0, h_N]$, we have

$$A_N(g_h) := \int_\Omega (d(x, \Gamma))^N \big(|g_h(x)|^2 + h^{-1}|(h\nabla - i\boldsymbol{F})g_h(x)|^2\big)\, dx \le C_N h^{N/2} r_h, \tag{5-5}$$

[2]This is coherent with (4-1) if we consider the function a normalized bound state.

and, for $\rho \in \left(0, \frac{1}{2}\right)$, there exist $C_{N,\rho}, h_{N,\rho} > 0$ such that, for all $h \in (0, h_{N,\rho}]$,

$$B_N(g_h) := \int_{d(x,\Gamma)\le h^\rho} |s(x)|^N \left(|g_h(x)|^2 + h^{-1}|(h\nabla - i\boldsymbol{F})g_h(x)|^2\right) dx \le C_N h^{N/8} r_h. \tag{5-6}$$

The two last estimates imply that, for a fixed $\rho \in \left(0, \frac{1}{2}\right)$, and $N \ge 1$, there exist $C_{N,\rho}, h_{N,\rho} > 0$ such that, for all $h \in (0, h_{N,\rho}]$, we have

$$\int_{d(x,\Gamma)\ge h^\rho} \left(|g_h(x)|^2 + h^{-1}|(h\nabla - i\boldsymbol{F})g_h(x)|^2\right) dx \le C_{N,\rho}\, h^N r_h, \tag{5-7}$$

and for $\eta \in \left(0, \frac{1}{8}\right)$, there exist $C_{N,\rho,\eta}, h_{N,\rho,\eta} > 0$ such that, for all $h \in (0, h_{N,\rho,\eta}]$, we have

$$\int_{\substack{d(x,\Gamma)\le h^\rho \\ |s(x)|\ge h^\eta}} \left(|g_h(x)|^2 + h^{-1}|(h\nabla - i\boldsymbol{F})g_h|^2\right) dx \le C_{N,\rho,\eta}\, h^N r_h. \tag{5-8}$$

In fact, (5-7) and (5-8) follow in a straightforward manner from (5-3) and (5-4) after noticing that

$$\int_{d(x,\Gamma)\ge h^\rho} \left(|g_h(x)|^2 + h^{-1}|(h\nabla - i\boldsymbol{F})g_h|^2\right) dx \le C r_h \exp(-\alpha h^{\rho-1/2}),$$
$$\int_{\substack{d(x,\Gamma)\le h^\rho \\ |s(x)|\ge h^\eta}} \left(|g_h(x)|^2 + h^{-1}|(h\nabla - i\boldsymbol{F})g_h|^2\right) dx \le C r_h \exp(-\alpha h^{\eta-1/8}).$$

5B. ***Rescaled functions and tangential estimates.*** Let $\delta \in \left(0, \frac{1}{12}\right)$ and $\eta \in \left(0, \frac{1}{8}\right)$ be two fixed constants. Consider the function w_h defined as

$$w_h(\sigma, \tau) = h^{5/16}\chi(h^\eta\sigma)\chi(h^\delta\tau)\tilde{g}_h(h^{1/8}\sigma, h^{1/2}\tau), \tag{5-9}$$

where $\tilde{g}_h$ is the function assigned to g_h by the Frenet coordinates as in (A-3), namely

$$\tilde{g}_h(s, t) = g_h(x),$$

and $\chi \in C_c^\infty(\mathbb{R})$, $\operatorname{supp}\chi \subset [-1, 1]$, $0 \le \chi \le 1$ and $\chi = 1$ on $\left[-\frac{1}{2}, \frac{1}{2}\right]$.

Note that, due to our conditions on δ and η, w_h can be seen as a function on $\mathbb{R}^2$, and its L^2-norm can be estimated by using (A-7) and (5-5) as follows:

$$\|w_h\|^2_{L^2(\mathbb{R}^2)} = (1 + \mathcal{O}(h^{1/2}))\|g_h\|^2_{L^2(\Omega)}. \tag{5-10}$$

Under our hypotheses on the function g_h (particularly (5-1) for $\theta \in \left(0, \frac{3}{8}\right)$ and (5-3)–(5-4)), we can estimate the tangential derivative of the function w_h.

Proposition 5.1. *For all $\theta \in \left(0, \frac{3}{8}\right)$, there exist constants $C_\theta, h_\theta > 0$ such that, if $h \in (0, h_\theta]$, and g_h satisfies (5-1)$_\theta$, (5-3) and (5-4), then the function w_h introduced in (5-9) satisfies the estimate*

$$\|(h^{3/8}\partial_\sigma - i\zeta_a)w_h\|_{L^2(\mathbb{R}^2)} \le C h^{3/8-\theta/2}(\|w_h\|_{L^2(\mathbb{R}^2)} + \sqrt{r_h} + h^{3/8-3\theta/4}). \tag{5-11}$$

Proof. The proof is split into four steps.

Step 1: We localize the integrals defining the L^2-norm and the quadratic form of g_h to the neighborhood, $\mathcal{N}_h = \{x \in \Omega : d(x,\Gamma) \le h^{1/2-\delta},\ |s(x)| \le h^{\eta}\}$, of the point of maximal curvature, $s=0$. In fact, by the decay estimates in (5-7) and (5-8),

$$\|g_h\|^2_{L^2(\Omega)} = \int_{\mathcal{N}_h} |g_h(x)|^2\,dx + \mathcal{O}(h^\infty) \quad \text{and} \quad \mathcal{Q}_h(g_h) = \int_{\mathcal{N}_h} |(h\nabla - i\boldsymbol{F})g_h|^2\,dx + \mathcal{O}(h^\infty).$$

We refine the localization of these integrals by using the decay estimates in (5-5) and (5-6), the change of variable formulas in (A-7) and the expansions

$$k(s) = \kappa + \mathcal{O}(s^2), \quad \mathfrak{a}(s,t) = 1 - t\kappa + \mathcal{O}(s^2 t), \quad \mathfrak{a}^{-2} = 1 + 2t\kappa + \mathcal{O}(s^2 t),$$

where we set $\kappa = k_{\max}$. More precisely,

$$\|g_h\|^2_{L^2(\Omega)} = \int_{\mathbb{R}} \int_{-h^{1/2-\delta}}^{h^{1/2-\delta}} |\tilde{g}_h|^2 (1-t\kappa)\,ds\,dt + \int_{\mathbb{R}} \int_{-h^{1/2-\delta}}^{h^{1/2-\delta}} \mathcal{O}(s^2 t)|\tilde{g}_h|^2\,ds\,dt + \mathcal{O}(h^\infty).$$

To estimate the second term in the right-hand side we use the Cauchy–Schwarz inequality to obtain

$$\int_{\mathbb{R}} \int_{-h^{1/2-\delta}}^{h^{1/2-\delta}} s^2|t|\,|\tilde{g}_h|^2\,ds\,dt \le \left(\int_{\mathbb{R}} \int_{-h^{1/2-\delta}}^{h^{1/2-\delta}} t^2|\tilde{g}_h|^2\,ds\,dt\right)^{1/2} \left(\int_{\mathbb{R}} \int_{-h^{1/2-\delta}}^{h^{1/2-\delta}} s^4|\tilde{g}_h|^2\,ds\,dt\right)^{1/2}.$$

Hence by (5-5) (with $N=2$) and (5-6) (with $N=4$) we get

$$\int_{\mathbb{R}} \int_{-h^{1/2-\delta}}^{h^{1/2-\delta}} s^2|t|\,|\tilde{g}_h(s,t)|^2\,ds\,dt = \mathcal{O}(h^{3/4})r_h.$$

Implementing the above, we have

$$\|g_h\|^2_{L^2(\Omega)} \le \int_{\mathbb{R}} \int_{-h^{-\delta}}^{h^{-\delta}} |w_h|^2 (1 - h^{1/2}\tau\kappa)\,d\sigma\,d\tau + \mathcal{O}(h^{3/4})r_h + \mathcal{O}(h^\infty) \tag{5-12}$$

and

$$\mathcal{Q}_h(g_h) = \int_{\mathbb{R}} \int_{-h^{1/2-\delta}}^{h^{1/2-\delta}} \left(|h\partial_t \tilde{g}_h|^2 + (1+2\kappa t)\left|\left(h\partial_s + ib_a(t)\left(t - \frac{\kappa t^2}{2}\right)\right)\tilde{g}_h\right|^2\right)(1-\kappa t)\,ds\,dt + \mathcal{O}(h^\infty) + \mathcal{O}(R_h), \tag{5-13}$$

where

$$R_h = \int_{\mathbb{R}^2} s^2|t|\left(|h\partial_t \tilde{g}_h|^2 + \left|\left(h\partial_s + ib_a(t)\left(t - \frac{k(s)t^2}{2}\right)\right)\tilde{g}_h\right|^2\right)ds\,dt + \int_{\mathbb{R}^2} s^4 t^4 |\tilde{g}_h|^2\,ds\,dt + \left(\int_{\mathbb{R}^2} s^4 t^4 |\tilde{g}_h|^2\,ds\,dt\right)^{1/2} \|(h\nabla - i\boldsymbol{F})g_h\|_{L^2(\Omega)}.$$

Proceeding as above for the treatment of $\int_{\mathbb{R}^2} s^4 t^4 |\tilde{g}_h|^2\,ds\,dt$, we infer from (5-1), (5-5) and (5-6) that

$$R_h \le C\big((A_2(g_h)B_4(g_h))^{1/2}h + (A_8(g_h)B_8(g_h))^{1/2} + (A_8(g_h)B_8(g_h))^{1/4}h^{1/2}\big) = \mathcal{O}(h^{7/4}r_h).$$

Now, coming back to (5-1), we get after performing a change of variable and dividing by h that[3]

$$\int_{\mathbb{R}}\int_{-h^{-\delta}}^{h^{-\delta}}\Big(|\partial_\tau w_h|^2+(1+2\kappa h^{1/2}\tau)\Big|\Big(h^{3/8}\partial_\sigma+i\Big(b_a(\tau)\tau-\kappa h^{1/2}b_a(\tau)\frac{\tau^2}{2}\Big)\Big)w_h\Big|^2\Big)(1-\kappa h^{1/2}\tau)\,d\sigma\,d\tau$$
$$\leq(\beta_a+h^{1/2}M_3(a)\kappa+\mathcal{O}(h^{3/4}))m_h+\mathcal{O}(h^{3/4}r_h)+\mathcal{O}(h^{3/2-\theta}),\quad(5\text{-}14)$$

where

$$m_h:=\int_{\mathbb{R}}\int_{-h^{-\delta}}^{h^{-\delta}}|w_h|^2(1-\kappa h^{1/2}\tau)\,d\sigma\,d\tau=(1+o(1))\|w_h\|^2_{L^2(\mathbb{R}^2)}.\quad(5\text{-}15)$$

In the sequel, we set

$$M_h=m_h+r_h.\quad(5\text{-}16)$$

Next we perform a Fourier transform with respect to σ and denote the transform of w_h by

$$\hat{w}_h(\xi,t)=\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}w_h(\sigma,t)e^{-i\sigma\xi}\,d\sigma.$$

Then it is immediate from (5-14) and (5-15) that we have

$$\int_{\mathbb{R}}\int_{-h^{-\delta}}^{h^{-\delta}}\Big(|\partial_\tau\hat{w}_h|^2+(1+2\kappa h^{1/2}\tau)\Big|\Big(h^{3/8}\xi+b_a(\tau)\tau-\kappa h^{1/2}b_a(\tau)\frac{\tau^2}{2}\Big)\hat{w}_h\Big|^2\Big)(1-\kappa h^{1/2}\tau)\,d\xi\,d\tau$$
$$\leq(\beta_a+h^{1/2}M_3(a)\kappa)m_h+\mathcal{O}(h^{3/4}M_h)+\mathcal{O}(h^{3/2-\theta}),\quad(5\text{-}17)$$

and m_h introduced in (5-15) now satisfies

$$m_h=\int_{\mathbb{R}}\int_{-h^{-\delta}}^{h^{-\delta}}|\hat{w}_h|^2(1-\kappa h^{1/2}\tau)\,d\xi\,d\tau.\quad(5\text{-}18)$$

<u>Step 2</u>: We introduce

$$f_h(\xi)=q_{a,\zeta,\kappa,h}(\hat{w}_h)\big|_{\zeta=h^{3/8}\xi},\quad(5\text{-}19)$$

where $q_{a,\zeta,\kappa,h}$ is the quadratic form introduced in (3-4). We rewrite (5-17) as

$$\int_{\mathbb{R}}f_h(\xi)\,d\xi\leq(\beta_a+h^{1/2}M_3(a)\kappa)m_h+\mathcal{O}(h^{3/4}M_h)+\mathcal{O}(h^{3/2-\theta}).\quad(5\text{-}20)$$

Fix a positive constant $\varepsilon<1$. Then by Proposition 3.2,

$$f_h(\xi)\geq\int_{-h^{-\delta}}^{h^{-\delta}}\big(\beta_a+\hat{c}_0(a)\min((h^{3/8}\xi-\zeta_a)^2,\varepsilon)+h^{1/2}M_3(a)\kappa-C_\varepsilon h\big)|\hat{w}_h|^2(1-h^{1/2}\kappa\tau)\,d\tau.\quad(5\text{-}21)$$

Inserting this into (5-20) we get

$$\int_{\mathbb{R}}\int_{-h^{-\delta}}^{h^{-\delta}}\hat{c}_0(a)\min((h^{3/8}\xi-\zeta_a)^2,\varepsilon)|\hat{w}_h|^2(1-h^{1/2}\kappa\tau)\,d\xi\,d\tau=\mathcal{O}(h^{3/4}M_h)+\mathcal{O}(h^{3/2-\theta}),$$

from which we infer the two estimates

$$\int_{|h^{3/8}\xi-\zeta_a|^2<\varepsilon}\int_{-h^{-\delta}}^{h^{-\delta}}|h^{3/8}\xi-\zeta_a|^2|\hat{w}_h|^2(1-h^{1/2}\kappa\tau)\,d\xi\,d\tau=\mathcal{O}(h^{3/4}M_h)+\mathcal{O}(h^{3/2-\theta}),\quad(5\text{-}22)$$

$$\int_{|h^{3/8}\xi-\zeta_a|^2\geq\varepsilon}\int_{-h^{-\delta}}^{h^{-\delta}}|\hat{w}_h|^2(1-h^{1/2}\kappa\tau)\,d\xi\,d\tau=\mathcal{O}(h^{3/4}M_h)+\mathcal{O}(h^{3/2-\theta}).\quad(5\text{-}23)$$

[3]Replacing the cut-off functions in (5-9) by 1 in the integrals produces $\mathcal{O}(h^\infty)$ errors by (5-7) and (5-8).

Step 3: Noticing the simple decomposition

$$\int_{\mathbb{R}}\int_{-h^{-\delta}}^{h^{-\delta}} |\hat{w}_h|^2(1-h^{1/2}\kappa\tau)\,d\xi\,d\tau$$
$$= \int_{|h^{3/8}\xi-\zeta_a|^2<\varepsilon}\int_{-h^{-\delta}}^{h^{-\delta}} |\hat{w}_h|^2(1-h^{1/2}\kappa\tau)\,d\xi\,d\tau + \int_{|h^{3/8}\xi-\zeta_a|^2\geq\varepsilon}\int_{-h^{-\delta}}^{h^{-\delta}} |\hat{w}_h|^2(1-h^{1/2}\kappa\tau)\,d\xi\,d\tau, \tag{5-24}$$

we get from (5-23) and (5-18)

$$\int_{|h^{3/8}\xi-\zeta_a|^2<\varepsilon}\int_{-h^{-\delta}}^{h^{-\delta}} |\hat{w}_h|^2(1-h^{1/2}\kappa\tau)\,d\xi\,d\tau = m_h + \mathcal{O}(h^{3/4}M_h) + \mathcal{O}(h^{3/2-\theta}). \tag{5-25}$$

Similarly, we decompose the integral in (5-20) as

$$\int_{\mathbb{R}} f_h(\xi)\,d\xi = \int_{|h^{3/8}\xi-\zeta_a|^2<\varepsilon} f_h(\xi)\,d\xi + \int_{|h^{3/8}\xi-\zeta_a|^2\geq\varepsilon} f_h(\xi)\,d\xi. \tag{5-26}$$

We write a lower bound of the integral on $\{|h^{3/8}\xi-\zeta_a|^2\geq\varepsilon\}$ by using (5-21). Noting that $\hat{c}_0(a)>0$, we get, by (5-25),

$$\int_{|h^{3/8}\xi-\zeta_a|^2<\varepsilon} f_h(\xi)\,d\xi \geq \big(\beta_a + h^{1/2}M_3(a)\kappa + \mathcal{O}(h)\big)m_h + \mathcal{O}(h^{3/4}M_h) + \mathcal{O}(h^{3/2-\theta}).$$

Inserting this into (5-26) and using (5-20), we get

$$\int_{|h^{3/8}\xi-\zeta_a|^2\geq\varepsilon} f_h(\xi)\,d\xi = \mathcal{O}(h^{3/4}M_h) + \mathcal{O}(h^{3/2-\theta}). \tag{5-27}$$

Step 4: We write a lower bound for $f_h(\xi)$ by gathering (5-19) and (3-8), thereby obtaining

$$\int_{|h^{3/8}\xi-\zeta_a|^2\geq\varepsilon} f_h(\xi)\,d\xi \geq (1-Ch^{1/2-2\delta})\int_{|h^{3/8}\xi-\zeta_a|^2\geq\varepsilon}\int_{\mathbb{R}}\big(|\partial_\tau\hat{w}_h|^2 + |(b_a(\tau)\tau+h^{3/8}\xi)\hat{w}_h|^2\big)\,d\xi\,d\tau.$$

Using (5-27) and the inequality (note that $|b_a|\leq 1$ since $|a|<1$)

$$(b_a(\tau)\tau+h^{3/8}\xi)^2 \geq \tfrac{1}{2}(h^{3/8}\xi)^2 - 2\tau^2,$$

we get

$$\frac{1}{2}\int_{|h^{3/8}\xi-\zeta_a|^2\geq\varepsilon}\int_{\mathbb{R}} |h^{3/8}\xi\,\hat{w}_h|^2\,d\xi\,d\tau \leq 2\int_{|h^{3/8}\xi-\zeta_a|^2\geq\varepsilon}\int_{\mathbb{R}} \tau^2|\hat{w}_h|^2\,d\xi\,d\tau + \mathcal{O}(h^{3/4}M_h) + \mathcal{O}(h^{3/2-\theta}). \tag{5-28}$$

Let $p=1/\theta$ and $q=1/(1-\theta)$. By the Hölder inequality, (5-5) and (5-23), we write

$$\begin{aligned}
\int_{|h^{3/8}\xi-\zeta_a|^2\geq\varepsilon}\int_{\mathbb{R}} \underbrace{\tau^2|\hat{w}_h|^2}_{=\tau^2|\hat{w}_h|^{2\theta}|\hat{w}_h|^{2-2\theta}}\,d\xi\,d\tau
&\leq \left(\int_{|\xi_h-\zeta_a|^2\geq\varepsilon}\int_{\mathbb{R}} \tau^{2p}|\hat{w}_h|^{2p\theta}\,d\xi\,d\tau\right)^{1/p}\left(\int_{|\xi_h-\zeta_a|^2\geq\varepsilon}\int_{\mathbb{R}} |\hat{w}_h|^{q(2-2\theta)}\,d\xi\,d\tau\right)^{1/q}\\
&\leq \left(\int_{\mathbb{R}^2} \tau^{2p}|w_h|^2\,d\tau\,ds\right)^{1/p}\left(\int_{|\xi_h-\zeta_a|^2\geq\varepsilon}\int_{\mathbb{R}} |\hat{w}_h|^2\,d\xi\,d\tau\right)^{1/q}\\
&= \mathcal{O}(h^{3/4(1-\theta)}M_h) + \mathcal{O}(M_h^{\theta}\,h^{(1-\theta)(3/2-\theta)})\\
&= \mathcal{O}(h^{3/4(1-\theta)}M_h) + \mathcal{O}(h^{3/2-5\theta/2}),
\end{aligned}$$

where, in the last step, we used Young's inequality,

$$\begin{aligned} M_h^\theta h^{(1-\theta)(3/2-\theta)} &= M_h \theta h^{\theta(3/4-\theta)} h^{(1-\theta)(3/2-\theta)-\theta(3/4-\theta)} \\ &\le \theta M_h h^{3/4-\theta} + (1-\theta) h^{3/2-\theta} h^{-(3/4-\theta)\theta/(1-\theta)} \\ &\le \theta M_h h^{3/4-\theta} + (1-\theta) h^{3/2-5\theta/2} \quad \text{for } 0<\theta<\tfrac{3}{8}. \end{aligned}$$

Inserting this estimate into (5-28), we get

$$\int_{|h^{3/8}\xi-\zeta_a|^2\ge\varepsilon} \int_{\mathbb{R}} |h^{3/8}\xi \hat{w}_h|^2 \, d\xi \, d\tau = \mathcal{O}(h^{3/4-\theta} M_h) + \mathcal{O}(h^{3/2-5\theta/2}).$$

Collecting the foregoing estimate and those in (5-22) and (5-23), we deduce that

$$\int_{\mathbb{R}^2} |(h^{3/8}\partial_\sigma - i\zeta_a) w_h|^2 \, d\sigma \, d\tau = \int_{\mathbb{R}} \int_{-h^{-\delta}}^{h^{-\delta}} |h^{3/8}\xi - \zeta_a|^2 |\hat{w}_h|^2 \, d\xi \, d\tau = \mathcal{O}(h^{3/4-\theta} M_h) + \mathcal{O}(h^{3/2-5\theta/2}).$$

With (5-15) and (5-16) in mind, this implies (5-11) as stated in the proposition. □

6. Localization of bound states

In this section, we fix a labeling $n \ge 1$ and denote by $\psi_{h,n}$ a normalized eigenfunction of the operator $\mathcal{P}_h$ with eigenvalue $\lambda_n(h)$. By Theorem 4.1, it holds

$$\mathcal{Q}_h(\psi_{h,n}) \le (h\beta_a + h^{3/2} M_3(a) k_{\max} + C_1 h^{7/4}) \|\psi_{h,n}\|^2_{L^2(\Omega)}, \tag{6-1}$$

where $\mathcal{Q}_h$ is the quadratic form introduced in (1-3).

The decay estimates in Sections 6A and 6B follow by standard semiclassical Agmon estimates. We refer to [Helffer and Morame 2001; Fournais and Helffer 2006] for details in the case of the Laplacian with a smooth magnetic field, and to [Assaad and Kachmar 2022] for adaptations in the piecewise constant field discussed here.

Using the aforementioned decay estimates, the bound state $\psi_{h,n}$ satisfies the hypotheses in Section 5. Namely the estimates in $(5\text{-}1)_\theta$, (5-3) and (5-4) hold with $g_h = \psi_{h,n}$, $r_h = 1$ and for any $\theta \in \left(0, \frac{3}{8}\right)$. Consequently, we will be able to estimate its tangential derivative (see Proposition 6.2). Estimating the second-order tangential derivative of $\psi_{h,n}$ (as in Proposition 6.3) requires the analysis of the decay of its first-order tangential derivative in order to verify the hypotheses of Section 5.

6A. ***Decay away from the edge.*** The derivation of an Agmon decay estimate relies on the following useful lower bound of the quadratic form [Assaad and Kachmar 2022, Section 4.3]. For every $R_0 > 1$, there exists a positive constant C_0 and $h_0 > 0$ such that, for $h \in (0, h_0]$,

$$\mathcal{Q}_h(u) \ge \int_\Omega (U_{h,a}(x) - C_0 R_0^{-2} h) |u(x)|^2 \, dx \quad (u \in H_0^1(\Omega)), \tag{6-2}$$

where $\mathcal{Q}_h$ is introduced in (1-3) and

$$U_{h,a}(x) = \begin{cases} |a|h & \text{if } \operatorname{dist}(x,\Gamma) > R_0 h^{1/2}, \\ \beta_a h & \text{if } \operatorname{dist}(x,\Gamma) < R_0 h^{1/2}. \end{cases}$$

Note that the decay property is a consequence of $\beta_a < |a|$. Following [Fournais and Helffer 2010, Theorem 8.2.4], it results from the foregoing lower bound that the eigenfunction $\psi_{h,n}$ decays roughly like $\exp(-\alpha_0 h^{-1/2} d(x,\Gamma))$ for some constant $\alpha_0 > 0$. More precisely, the following holds:

$$\int_\Omega \left(|\psi_{h,n}|^2 + h^{-1}|(h\nabla - i\boldsymbol{F})\psi_{h,n}|^2\right) \exp(2\alpha_0 h^{-1/2}\, d(x,\Gamma))\, dx \le C. \tag{6-3}$$

6B. ***Decay along the edge.*** Here we discuss tangential estimates along the edge Γ. Recall that $s = 0$ corresponds to the (unique) point of maximal curvature.

The starting point is the following refined lower bound of the quadratic form [Assaad and Kachmar 2022, Section 4.3]:

$$\mathcal{Q}_h(u) \ge \int_\Omega (U^\Gamma_{h,a}(x) - C_0 h^{7/4})|u|^2\, dx \quad (u \in H^1_0(\Omega)), \tag{6-4}$$

where, with $x = \Phi(s,t)$, $\kappa(s) = k_{\max} - \varepsilon_0 s^2$ and ε_0 a positive constant,

$$U^\Gamma_{h,a}(x) = \begin{cases} |a|h & \text{if } \operatorname{dist}(x,\Gamma) \ge 2h^{1/6}, \\ \beta_a h + M_3(a)\kappa(s)h^{3/2} & \text{if } \operatorname{dist}(x,\Gamma) < 2h^{1/6}. \end{cases}$$

Here we recall that $M_3(a)$ is negative so the potential in the second zone is minimal at the point of maximal curvature. The lower bound (6-4) can be derived along the same arguments in [Fournais and Helffer 2010, Proposition 8.3.3, Remark 8.3.6] and by using Proposition 3.2.

The eigenfunction $\psi_{n,h}$ decays exponentially roughly like $\exp(-\alpha_1 h^{-1/8} s(x))$ for some constant $\alpha_1 > 0$. More precisely, picking t_0 sufficiently small so that the Frenet coordinates recalled in Appendix A are valid in $\{d(x,\Gamma) < t_0\}$, we have

$$\int_{d(x,\Gamma)\le t_0} \left(|\psi_{h,n}(x)|^2 + h^{-1}|(h\nabla - i\boldsymbol{F})\psi_{h,n}|^2\right) \exp(2\alpha_1 h^{-1/8}|s(x)|)\, dx \le C. \tag{6-5}$$

Remark 6.1. We observe, by collecting (6-1), (6-3) and (6-5), that the eigenfunction $g_h = \psi_{h,n}$ satisfies the hypotheses of Proposition 5.1, namely

- (5-1) holds for any $\theta \in \left(0, \frac{3}{8}\right)$,
- (5-3) and (5-4) hold with $0 < \alpha \le \min(2\alpha_1, 2\alpha_2)$ and $r_h = 1$.

6C. ***Estimating tangential frequency.*** The localization of the eigenfunction $\psi_{h,n}$ is to be measured by two parameters $\rho \in \left(0, \frac{1}{2}\right)$ and $\eta \in \left(0, \frac{1}{8}\right)$. We will choose $\rho = \frac{1}{2} - \delta$ with $\delta \in \left(0, \frac{1}{12}\right)$; i.e., we are assuming

$$\tfrac{5}{12} < \rho < \tfrac{1}{2}.$$

We introduce the function

$$u_{h,n}(\sigma,\tau) = h^{5/16}\chi(h^\eta\sigma)\chi(h^\delta\tau)\tilde{\psi}_{h,n}(h^{1/8}\sigma, h^{1/2}\tau), \tag{6-6}$$

where $\tilde{\psi}_{h,n}$ is the function assigned to $\psi_{h,n}$ by the Frenet coordinates as in (A-3), $\chi \in C^\infty_c(\mathbb{R})$, $\operatorname{supp}\chi \subset [-1,1]$, $0 \le \chi \le 1$ and $\chi = 1$ on $\left[-\frac{1}{2}, \frac{1}{2}\right]$. Note that $u_{h,n}$ can be seen as a function on $\mathbb{R}^2$, and by (5-10)

(applied with $g_h = \psi_{h,n}$), its L^2-norm satisfies

$$\|u_{h,n}\|^2_{L^2(\mathbb{R}^2)} = \|\psi_{h,n}\|^2_{L^2(\Omega)}(1+\mathcal{O}(h^{1/2})) = 1+\mathcal{O}(h^{1/2}), \tag{6-7}$$

since $\psi_{h,n}$ is normalized in $L^2(\Omega)$.

Using Proposition 5.1, we can estimate the tangential derivative of $u_{h,n}$. More precisely, we apply this proposition with $g_h = \psi_{h,n}$, $r_h = 1$ and *any* $0 < \theta < \frac{3}{8}$ (see Remark 6.1). In this case, the function introduced in (5-9) is given by $w_h = u_{h,n}$.

Proposition 6.2. *For all $\theta \in \left(0, \frac{3}{8}\right)$, there exist constants $C_\theta, h_\theta > 0$ such that, for all $h \in (0, h_\theta]$,*

$$\|(h^{3/8}\partial_\sigma - i\zeta_a)u_{h,n}\|_{L^2(\mathbb{R}^2)} \le C_\theta\, h^{3/8-\theta}.$$

We can estimate higher-order tangential derivatives of $u_{h,n}$.

Proposition 6.3. *For all $\theta \in \left(0, \frac{3}{4}\right)$, there exist constants $C_\theta, h_\theta > 0$ such that, for all $h \in (0, h_\theta]$,*

$$\|(h^{3/8}\partial_\sigma - i\zeta_a)^2 u_{h,n}\|_{L^2(\mathbb{R}^2)} \le C_\theta\, h^{3/4-\theta}, \tag{6-8}$$

where $u_{h,n}$ is introduced in (6-6).

Before proceeding with the proof of Proposition 6.3, we introduce the notation, $r_h = \widetilde{\mathcal{O}}(h^\gamma)$ for a positive number γ, to mean

$$\text{for all } \theta \in (0,\gamma), \text{ there exists } C_\theta, h_\theta > 0 \text{ such that, for all } h \in (0, h_\theta), \quad |r_h| \le C_\theta h^{\gamma-\theta}. \tag{6-9}$$

Proof of Proposition 6.3. We will apply Proposition 5.1 with an adequate choice of the function g_h defining the function w_h in (5-9).

We introduce the function φ_h on Ω as

$$\varphi_h(x) = f(x)\psi_{h,n}(x), \tag{6-10}$$

where $f(x) = (1-\chi(\mathrm{dist}(x,\partial\Omega)/t_1))\,\chi(\mathrm{dist}(x,\Gamma)/t_0)$, t_1 and t_0 are constants so that the set $\{x \in \Omega : \mathrm{dist}(x,\partial\Omega) > t_1\}$ contains the point of maximum curvature and the transformation in (A-1) is a diffeomorphism, $\chi \in C_c^\infty(\mathbb{R})$, $\mathrm{supp}\,\chi \subset [-1,1]$, $0 \le \chi \le 1$ and $\chi = 1$ on $\left[-\frac{1}{2}, \frac{1}{2}\right]$. Then we define

$$\tilde{g}_h(s,t) = (h^{1/2}\partial_s - i\zeta_a)\tilde{\varphi}_h(s,t), \tag{6-11}$$

where $\tilde{\varphi}_h$ is the function assigned to φ_h by (A-3). Notice that, using the notation in (6-9), the conclusion of Proposition 6.2 can be written as

$$\|g_h\|_{L^2(\Omega)} = \widetilde{\mathcal{O}}(h^{3/8}). \tag{6-12}$$

We will show that g_h satisfies $(5\text{-}1)_\theta$ for *any* $\theta \in \left(0, \frac{3}{8}\right)$, and that (5-3) and (5-4) hold with

$$r_h = \|g_h\|^2_{L^2(\Omega)} + h^{3/4}. \tag{6-13}$$

This will be done in several steps outlined below.

- In Step 1, we establish rough decay estimates for g_h in the normal and tangential directions (see (6-20)). These estimates are nevertheless weaker than the estimates in (5-3) and (5-4) that we wish to prove.
- In Step 2, we show that g_h is in the domain of the operator $\mathcal{P}_h$ introduced in (1-4).
- In Step 3, using the rough estimates obtained in Steps 1 and 2, we can verify that (5-1) holds for any $\theta \in \left(0, \frac{3}{8}\right)$.
- In Step 4, using the estimates obtained in Steps 1 and 3, and the Agmon method, we derive the decay estimates for g_h as in (5-3) and (5-4) with r_h given in (6-13).
- In Step 5, we can apply the conclusion of Proposition 5.1 and conclude the proof of Proposition 6.3.

Step 1: We show that the function g_h decays exponentially in the normal and tangential directions. We select the constant t_0 so that the two functions

$$x \mapsto \operatorname{dist}(x, \Gamma) \quad \text{and} \quad x \mapsto s(x)$$

are smooth in the neighborhood, Γ_{2t_0}, of the edge Γ. Consequently, the transformation in (A-1) is valid in Γ_{2t_0}. Since we encounter integrals of the function g_h, which is supported in $\Gamma_{t_0} \cap \Omega$, we select the gauge given in Lemma A.1. In particular, by (A-4), we have

$$|\boldsymbol{F}(x)| = \mathcal{O}(\operatorname{dist}(x, \Gamma)) \quad \text{on } \Omega \cap \Gamma_{t_0}. \tag{6-14}$$

Let $\alpha_2 \in \left(0, \frac{1}{2}\min(\alpha_0, \alpha_1)\right)$, where α_0, α_1 are the positive constants in (6-3) and (6-5). We introduce on Ω the weight functions

$$\Phi_{\text{norm}}(x) = \exp\left(\frac{\alpha_2 \operatorname{dist}(x, \Gamma)}{h^{1/2}}\right) \quad \text{and} \quad \Phi_{\tan}(x) = \exp\left(\frac{\alpha_2\, s(x)}{h^{1/8}}\right). \tag{6-15}$$

By Remark 6.1, we can use (5-5) for $\psi_{h,n}$. It results from (6-5), (6-14), the Hölder inequality, and our choice of α_2, that, for $j \in \{1, 2\}$,

$$\begin{aligned}\int_\Omega |\boldsymbol{F}|^{2j} |\psi_{h,n}|^2 \Phi_{\tan}^2\, dx &= \int_{\Omega\cap\Gamma_{t_0}} |\boldsymbol{F}|^{2j} |\psi_{h,n}|^2 \Phi_{\tan}^2\, dx + \mathcal{O}(h^\infty) \\ &\le A_{4j}(\psi_{h,n})^{1/2} \|\Phi_{\tan}^2 \psi_{h,n}\|_{L^2(\Omega)} + \mathcal{O}(h^\infty) = \mathcal{O}(h^j),\end{aligned} \tag{6-16}$$

where $A_{4j}(\,\cdot\,)$ is defined in (5-5) and

$$\begin{aligned}\int_\Omega |\boldsymbol{F}\cdot(h\nabla - i\boldsymbol{F})\psi_{h,n}|^2 \Phi_{\tan}^2\, dx &= \int_{\Omega\cap\Gamma_{t_0}} |\boldsymbol{F}\cdot(h\nabla - i\boldsymbol{F})\psi_{h,n}|^2 \Phi_{\tan}^2\, dx + \mathcal{O}(h^\infty) \\ &\le A_4(\psi_{h,n})^{1/2} \|\Phi_{\tan}^2 (h\nabla - i\boldsymbol{F})\psi_{h,n}\|_{L^2(\Omega)} + \mathcal{O}(h^\infty) = \mathcal{O}(h^2).\end{aligned}$$

Similarly, we estimate the $L^2(\Omega)$-norms of $\boldsymbol{F}\psi_{h,n}\Phi_{\text{norm}}$, $(\boldsymbol{F}\cdot\boldsymbol{F})\psi_{h,n}\Phi_{\text{norm}}$ and $\Phi_{\text{norm}}\boldsymbol{F}\cdot(h\nabla - i\boldsymbol{F})\psi_{h,n}$ using (6-3). Eventually, we get the estimates

$$\begin{aligned}&\|\boldsymbol{F}\psi_{h,n}\Phi_{\text{norm}}\|_{L^2(\Omega\cap\Gamma_{2t_0};\mathbb{R}^2)} + \|\boldsymbol{F}\psi_{h,n}\Phi_{\tan}\|_{L^2(\Omega\cap\Gamma_{2t_0};\mathbb{R}^2)} \\ &\qquad \le Ch^{1/2}\|\boldsymbol{F}\cdot\nabla(\psi_{h,n}\Phi_{\text{norm}})\|_{L^2(\Omega\cap\Gamma_{2t_0};\mathbb{R}^2)} + \|\boldsymbol{F}\cdot\nabla(\psi_{h,n}\Phi_{\tan})\|_{L^2(\Omega\cap\Gamma_{2t_0})} \le C.\end{aligned} \tag{6-17}$$

Furthermore, the following two estimates hold:

$$\begin{aligned}\|\psi_{h,n}\Phi_{\text{norm}}\|_{L^2(\Omega\cap\Gamma_{2t_0})}+\|\psi_{h,n}\Phi_{\tan}\|_{L^2(\Omega\cap\Gamma_{2t_0})}\le C,\\ \|\psi_{h,n}\Phi_{\text{norm}}\|_{H^1(\Omega\cap\Gamma_{2t_0})}+\|\psi_{h,n}\Phi_{\tan}\|_{H^1(\Omega\cap\Gamma_{2t_0})}\le Ch^{-1/2}.\end{aligned}\tag{6-18}$$

Notice that for $w_\# := \psi_{h,n}\Phi_\#$, ($\# \in \{\text{norm}, \tan\}$), we have, with $\mathcal{P}_h$ the operator introduced in (1-4),

$$\mathcal{P}_h w_\# = \lambda_n(h) w_\# - 2h\nabla\Phi_\#\cdot(h\nabla - i\boldsymbol{F})\psi_{h,n} - h^2\Delta\Phi_\#\,\psi_{h,n}.$$

Hence, noting that $\mathcal{P}_h = -h^2\Delta + 2ih\boldsymbol{F}\cdot\nabla + ih\,\mathrm{div}\boldsymbol{F} + |\boldsymbol{F}|^2$, we find by (4-1), (6-16) and (6-17),

$$\begin{aligned}h^2\|\Delta w_\#\|_{L^2(\Omega\cap\Gamma_{2t_0})}\le\big(\|\mathcal{P}_h w_\#\|_{L^2(\Omega)}+\|(h\nabla-i\boldsymbol{F})w_\#\|_{L^2(\Omega\cap\Gamma_{2t_0})}+h\|\mathrm{div}\,\boldsymbol{F}\,w_\#\|_{L^2(\Omega\cap\Gamma_{2t_0})}\\ +2h\|\boldsymbol{F}\cdot\nabla w_\#\|_{L^2(\Omega\cap\Gamma_{t_0})}+\||\boldsymbol{F}|^2 w_\#\|_{L^2(\Omega\cap\Gamma_{2t_0})}\big)=\mathcal{O}(h).\end{aligned}$$

By the L^2-elliptic estimates for the Dirichlet problem in $\Gamma_{2t_0}\cap\Omega$, and noting that $w_\#$ satisfies the Dirichlet condition,

$$\|w_\#\|_{H^2(\Omega\cap\Gamma_{t_0})}\le C(t_0,\Omega)(\|\Delta w_\#\|_{L^2(\Omega\cap\Gamma_{2t_0})}+\|w_\#\|_{L^2(\Omega\cap\Gamma_{2t_0})}).$$

Consequently, we get the estimate

$$\|\psi_{h,n}\Phi_{\text{norm}}\|_{H^2(\Omega\cap\Gamma_{t_0})}+\|\psi_{h,n}\Phi_{\tan}\|_{H^2(\Omega\cap\Gamma_{t_0})}\le Ch^{-1}.\tag{6-19}$$

Now we can derive decay estimates of the function g_h introduced in (6-11). Controlling the decay of the magnetic gradient of g_h requires a decay estimate of $\psi_{h,n}$ in the H^2 norm. Actually, collecting (6-18) and (6-19), we observe that

$$\begin{aligned}\|g_h\Phi_{\text{norm}}\|_{L^2(\Gamma_{t_0})}+h^{-1/2}\|((h\nabla-i\boldsymbol{F})g_h)\Phi_{\text{norm}}\|_{L^2(\Gamma_{t_0};\mathbb{R}^2)}\le C,\\ \|g_h\Phi_{\tan}\|_{L^2(\Gamma_{t_0})}+h^{-1/2}\|((h\nabla-i\boldsymbol{F})g_h)\Phi_{\tan}\|_{L^2(\Gamma_{t_0};\mathbb{R}^2)}\le C.\end{aligned}\tag{6-20}$$

<u>Step 2</u>: By the definition of g_h in (6-11), this function is compactly supported in $\Omega\cap\Gamma_{t_0}$. Hence, there exists a regular open set ω such that, for $h\in(0,h_0]$, $\operatorname{supp} g_h\subset\omega\subset\bar\omega\subset\Omega\cap\Gamma_{2t_0}$. Consequently g_h satisfies the Dirichlet boundary condition on $\partial\omega$. To prove that g_h is in the domain of the operator $\mathcal{P}_h$, it suffices to establish that

$$\partial_s\tilde\psi_{h,n}\in H^2(\Phi^{-1}(\omega)).\tag{6-21}$$

To that end, we consider the spectral equation satisfied by the eigenfunction $\psi_{h,n}$

$$-(h\nabla-i\boldsymbol{F})^2\psi_{h,n}=\lambda_n(h)\psi_{h,n}.\tag{6-22}$$

Using (A-5) with the potential $\widetilde{\boldsymbol{F}}$ in (4-3), (6-22) reads in the (s,t)-coordinates as

$$-\big(\mathfrak{a}^{-1}(h\partial_s-i\widetilde F_1)\mathfrak{a}^{-1}(h\partial_s-i\widetilde F_1)+h^2\mathfrak{a}^{-1}\partial_t\mathfrak{a}\partial_t\big)\tilde\psi_{h,n}=\lambda_n(h)\tilde\psi_{h,n},\tag{6-23}$$

that is,

$$h^2(\mathfrak{a}^{-2}\partial_s^2\tilde\psi_{h,n}+\partial_t^2\tilde\psi_{h,n})=f_1(s,t)\partial_s\tilde\psi_{h,n}+f_2(s,t)\partial_t\tilde\psi_{h,n}+f_3(s,t)\tilde\psi_{h,n},\tag{6-24}$$

where

$$f_1(s,t) = -h^2\mathfrak{a}^{-3}tk'(s) - 2i\mathfrak{a}^{-2}b_a(t)\Big(t - \frac{t^2}{2}k(s)\Big),$$
$$f_2(s,t) = h^2\mathfrak{a}^{-1}k(s),$$
$$f_3(s,t) = -ih\mathfrak{a}^{-3}tk'(s)b_a(t)\Big(t - \frac{t^2}{2}k(s)\Big) + h\mathfrak{a}^{-2}\frac{t^2}{2}k'(s) + \mathfrak{a}^{-2}b_a^2(t)\Big(t - \frac{t^2}{2}k(s)\Big)^2 - \lambda_n(h).$$

We differentiate with respect to s in (6-24), and get

$$\begin{aligned} h^2(\mathfrak{a}^{-2}\partial_s^2 + \partial_t^2)(\partial_s\tilde{\psi}_{h,n}) \\ = (f_1 - h^2\partial_s\mathfrak{a}^{-2})\partial_s^2\tilde{\psi}_{h,n} + f_2\,\partial_s\partial_t\tilde{\psi}_{h,n} + (\partial_s f_1 + f_3)\,\partial_s\tilde{\psi}_{h,n} + \partial_s f_2\,\partial_t\tilde{\psi}_{h,n} + \partial_s f_3\,\tilde{\psi}_{h,n}. \end{aligned} \tag{6-25}$$

Having $s \mapsto k(s)$ smooth, $\mathfrak{a} = 1 - tk(s)$ for $t \in (-2t_0, 2t_0)$, and $\psi_{n,h} \in \operatorname{Dom}\mathcal{P}_h$ ensures that the function in the right-hand side of (6-25) is in $L^2(\Phi^{-1}(\Omega\cap\Gamma_{2t_0}))$. Hence $\partial_s\tilde{\psi}_{h,n} \in H^1(\Omega\cap\Gamma_{2t_0})$ and satisfies

$$(\mathfrak{a}^{-2}\partial_s^2 + \partial_t^2)\partial_s\tilde{\psi}_{h,n} \in L^2(\Phi^{-1}(\Omega\cap\Gamma_{2t_0})). \tag{6-26}$$

Hence (6-21) follows from (6-26) using the interior elliptic estimates associated with the differential operator $L := (\mathfrak{a}^{-2}\partial_s^2 + \partial_t^2)$.

Step 3: We prove that

$$\mathcal{Q}_h(g_h) = \lambda_n(h)\|g_h\|^2_{L^2(\Omega)} + \widetilde{\mathcal{O}}(h^{5/2}), \tag{6-27}$$

where $\mathcal{Q}_h$ is the quadratic form introduced in (1-3).

With the notation introduced in (6-9), the estimates in (4-1) and (6-27) yield (5-1) for *any* $\theta \in \big(0, \frac{3}{8}\big)$.

We start by noticing that

$$\langle \mathcal{P}_h\varphi_h, G_h\rangle_{L^2(\Omega)} = \lambda_n(h)\langle\varphi_h, G_h\rangle_{L^2(\Omega)} + \langle(\mathcal{P}_h - \lambda_n(h))\varphi_h, G_h\rangle_{L^2(\Omega)}, \tag{6-28}$$

where φ_h is defined in (6-10) and

$$\widetilde{G}_h(s,t) = -(h^{1/2}\partial_s - i\zeta_a)g_h.$$

Recall that φ_h and G_h are compactly supported in $\Omega\cap\Gamma_{t_0}$ so that we can use the Frenet coordinates valid near the edge Γ. By (6-19) we have

$$\|(\mathcal{P}_h - \lambda_n(h))\varphi_h\|_{L^2(\Omega)} = \mathcal{O}(h^\infty) \tag{6-29}$$

and by (6-20)

$$\|G_h\|_{L^2(\Omega)} = \mathcal{O}(1). \tag{6-30}$$

By Hölder's inequality, we infer from (6-29) and (6-30)

$$\langle(\mathcal{P}_h - \lambda_n(h))\varphi_h, G_h\rangle_{L^2(\Omega)} = \mathcal{O}(h^\infty). \tag{6-31}$$

Furthermore, computing the integrals in the Frenet coordinates and integrating by parts, we find

$$\langle\varphi_h, G_h\rangle_{L^2(\Omega)} = \langle\mathfrak{a}(h^{1/2}\partial_s - i\zeta_a)\tilde{\varphi}_h + h^{1/2}(\partial_s\mathfrak{a})\tilde{\varphi}_h, \tilde{g}_h\rangle_{L^2(\mathbb{R}^2)} = \|g_h\|^2_{L^2(\Omega)} + \mathcal{O}(h^{9/8})\|g_h\|_{L^2(\Omega)}. \tag{6-32}$$

Here we get the $\mathcal{O}(h^{9/8})$ remainder by using that $\partial_s \mathfrak{a} = \mathcal{O}(ts)$, the Hölder inequality and Remark 6.1 on the decay estimates in (5-5) and (5-6) for $\psi_{h,n}$ as follows:

$$|\langle \mathfrak{a}(\partial_s \mathfrak{a})\tilde{\varphi}_h, \tilde{g}_h\rangle_{L^2(\mathbb{R}^2)}| \le C(A_4(\psi_{h,n})B_4(\psi_{h,n}))^{1/4}\|g_h\|_{L^2(\mathbb{R}^2)} = \mathcal{O}(h^{5/8})\|g_h\|_{L^2(\mathbb{R}^2)}.$$

By (4-1) and (6-12), we infer from (6-32)

$$\lambda_n(h)\langle \varphi_h, G_h\rangle_{L^2(\Omega)} = \lambda_n(h)\|g_h\|^2_{L^2(\Omega)} + \widetilde{\mathcal{O}}(h^{5/2}). \tag{6-33}$$

Therefore, inserting the estimates in (6-33) and (6-31) into (6-28), we find

$$\langle \mathcal{P}_h\varphi_h, G_h\rangle_{L^2(\Omega)} = \lambda_n(h)\|g_h\|^2_{L^2(\Omega)} + \widetilde{\mathcal{O}}(h^{5/2}). \tag{6-34}$$

Now, by Lemma A.2 (used with $\phi = 0$), we get

$$\operatorname{Re}\langle \mathcal{P}_h\varphi_h, G_h\rangle = \mathcal{Q}_h(g_h) - h^{1/2}\operatorname{Re}\langle R_h, g_h\rangle_{L^2(\Omega)}, \tag{6-35}$$

where the function R_h is defined via (A-3) as

$$\widetilde{R}_h(s,t) = (h\partial_s - i\widetilde{F}_1)\big((\partial_s\mathfrak{a}^{-1} - i\mathfrak{a}^{-1}\partial_s\widetilde{F}_1)(h\partial_s - i\widetilde{F}_1)\tilde{\varphi}_h - i\mathfrak{a}^{-1}(\partial_s\widetilde{F}_1)\tilde{\varphi}_h\big) + h^2\partial_t(\partial_s\mathfrak{a})\partial_t\tilde{\varphi}_h. \tag{6-36}$$

Our choice of gauge in Lemma A.1 ensures that $\widetilde{F}_2 = 0$ and $\widetilde{F}_1 = \mathcal{O}(t)$. By Remark 6.1 and (A-7), we have

$$\int_{\mathbb{R}}\int_{-t_0}^{t_0} |t|^N\big(|\tilde{\varphi}_h|^2 + \mathfrak{a}^{-1}h^{-1}|(h\partial_s - i\widetilde{F}_1)\tilde{\varphi}_h|^2 + h|\partial_t\tilde{\varphi}_h|^2\big)\mathfrak{a}\,ds\,dt = \mathcal{O}(h^{N/2}),$$
$$\int_{\mathbb{R}}\int_{-t_0}^{t_0} |s|^N\big(|\tilde{\varphi}_h|^2 + \mathfrak{a}^{-1}h^{-1}|(h\partial_s - i\widetilde{F}_1)\tilde{\varphi}_h|^2 + h|\partial_t\tilde{\varphi}_h|^2\big)\mathfrak{a}\,ds\,dt = \mathcal{O}(h^{N/8}).$$

Furthermore, by (6-19),

$$\int_{\mathbb{R}}\int_{-t_0}^{t_0} |t|^N(|\partial_s^2\tilde{\varphi}_h|^2 + |\partial_t^2\tilde{\varphi}_h|^2)\,ds\,dt = \mathcal{O}(h^{N/2-2}),$$
$$\int_{\mathbb{R}}\int_{-t_0}^{t_0} |s|^N(|\partial_s^2\tilde{\varphi}_h|^2 + |\partial_t^2\tilde{\varphi}_h|^2)\,ds\,dt = \mathcal{O}(h^{N/8-2}).$$

Now we can estimate $\widetilde{R}_h$ in (6-36), by expressing it as

$$\widetilde{R}_h = m_1(h\partial_s - i\widetilde{F}_1)^2\tilde{\varphi}_h + (m_2 + h\partial_s m_1)(h\partial_s - i\widetilde{F}_1)\tilde{\varphi}_h + h(\partial_s m_2)\tilde{\varphi}_h + h^2 m_3\partial_t^2\tilde{\varphi}_h + h^2(\partial_t m_3)\partial_t\tilde{\varphi}_h,$$

where

$$\begin{aligned} m_1 &= \partial_s\mathfrak{a}^{-1} - i\mathfrak{a}^{-1}\partial_s\widetilde{F}_1 = \mathcal{O}(ts), & \partial_s m_1 &= \mathcal{O}(t),\\ m_2 &= -i\mathfrak{a}^{-1}\partial_s\widetilde{F}_1 = \mathcal{O}(t^2 s), & \partial_s m_2 &= \mathcal{O}(t^3 s^2),\\ m_3 &= \partial_s\mathfrak{a} = \mathcal{O}(ts), & \partial_t m_3 &= \mathcal{O}(s). \end{aligned}$$

We get then that the norm of R_h satisfies

$$\|R_h\|_{L^2(\Omega)} = \mathcal{O}(h^{13/8}). \tag{6-37}$$

By Hölder's inequality, we infer from (6-37) and (6-12) the estimate

$$h^{1/2}|\operatorname{Re}\langle R_h, g_h\rangle_{L^2(\Omega)}| \le h^{1/2}\|R_h\|_{L^2(\Omega)}\|g_h\|_{L^2(\Omega)} = \widetilde{\mathcal{O}}(h^{5/2}).$$

Consequently, (6-34) and (6-35) yield (6-27).

Step 4: We refine the exponential decay of g_h. To that end, consider a fixed constant $0 < \alpha < \frac{1}{4}\alpha_2$, where α_2 is the constant in (6-15), and a real-valued Lipschitz function $\phi_{h,\alpha} \ge 0$, which will be either

$$\phi_{h,\alpha}(x) = \phi_{h,\alpha}^{\mathrm{norm}}(x) := \alpha h^{-1/2}\operatorname{dist}(x,\Gamma) \quad \text{or} \quad \phi_{h,\alpha}(x) = \phi_{h,\alpha}^{\tan}(x) := \alpha h^{-1/8}s(x).$$

We introduce the function $G_{h,\alpha}$ defined via (A-3) as

$$\widetilde{G}_{h,\alpha}(s,t) = -(h^{1/2}\partial_s - i\zeta_a)(e^{2\phi_{h,\alpha}}\tilde{g}_h(s,t)).$$

Since $\alpha < \frac{1}{4}\alpha_2$, we infer from (6-18) and (6-20)

$$\int_\Omega (\operatorname{dist}(x,\Gamma))^2|e^{\phi_{h,\alpha}}\varphi_h(x)|^2\,dx = \mathcal{O}(h),$$
$$\int_\Omega (s(x))^2|e^{\phi_{h,\alpha}}\varphi_h(x)|^2\,dx = \mathcal{O}(h^{1/4}),$$
$$\|G_{h,\alpha}\|_{L^2(\Omega)} = \mathcal{O}(1),$$

and also

$$\langle \mathcal{P}_h\varphi_h, G_{h,\alpha}\rangle_{L^2(\Omega)} = \lambda_n(h)\|e^{\phi_{h,\alpha}}g_h\|^2_{L^2(\Omega)} + \widetilde{\mathcal{O}}(h^{19/8}),$$

which results similarly to (6-34).

Now, we write by Lemma A.2,

$$\operatorname{Re}\langle \mathcal{P}_h\varphi_h, G_{h,\alpha}\rangle = \mathcal{Q}_h(e^{\phi_{h,\alpha}}g_h) - h^2\||\nabla\phi_{h,\alpha}|e^{\phi_{h,\alpha}}g_h\|^2_{L^2(\Omega)} - h^{1/2}\operatorname{Re}\langle R_h, e^{2\phi_{h,\alpha}}g_h\rangle_{L^2(\Omega)},$$

where R_h is introduced in (6-36). Since $\alpha < \frac{1}{4}\alpha_2$, we get from (6-18) and (6-19),

$$\|e^{\phi_{h,\alpha}}R_h\|_{L^2(\Omega)} = \mathcal{O}(h^{9/8}) \quad \text{and} \quad \langle R_h, e^{2\phi_{h,\alpha}}g_h\rangle_{L^2(\Omega)} = \mathcal{O}(h^{9/8})\|g_h\|_{L^2(\Omega)}.$$

Collecting the foregoing estimates, we get

$$\mathcal{Q}_h(e^{\phi_{h,\alpha}}g_h) = \lambda_n(h)\|e^{\phi_{h,\alpha}}g_h\|^2_{L^2(\Omega)} + \widetilde{\mathcal{O}}(h^{5/2}). \tag{6-38}$$

Now we can select $\alpha > 0$ small enough so that the following two estimates hold. The first estimate is

$$\int_\Omega \big(|g_h|^2 + h^{-1}|(h\nabla - i\boldsymbol{F})g_h|^2\big)\exp(\alpha h^{-1/2}\,d(x,\Gamma))\,dx \le C\|g_h\|^2_{L^2(\Omega)} + \widetilde{\mathcal{O}}(h^{3/2}), \tag{6-39}$$

and it follows after choosing $\phi_{h,\alpha} = \alpha h^{-1/2}\operatorname{dist}(x,\Gamma)$ and using (6-2). The second estimate follows by choosing $\phi_{h,\alpha} = \alpha h^{-1/8}s(x)$ and using (5-4); it reads as

$$\int_\Omega \big(|g_h|^2 + h^{-1}|(h\nabla - i\boldsymbol{F})g_h|^2\big)\exp(\alpha h^{-1/8}s(x))\,dx \le C\|g_h\|^2_{L^2(\Omega)} + \widetilde{\mathcal{O}}(h). \tag{6-40}$$

Step 5: Let $\theta \in \left(0, \frac{3}{8}\right)$. Collecting the estimates in (6-27), (6-39) and (6-40), we observe that the function g_h satisfies (5-1)$_\theta$, (5-3) and (5-4) with $r_h = \mathcal{O}(h^{3/4-\theta})$. We can then apply Proposition 5.1 and get (recall that $\|w_h\|_{L^2(\Omega)} \sim \|g_h\|_{L^2(\Omega)} \leq \sqrt{r_h}$ by (5-10))

$$\|(h^{3/8}\partial_\sigma - i\zeta_a)w_h\|_{L^2(\Omega)} \leq C_\theta h^{3/8-\theta/2}\left(\|g_h\|_{L^2(\Omega)} + \sqrt{r_h} + h^{3/8-3\theta/4}\right) = \mathcal{O}(h^{3/4-5\theta/4}).$$

Since this holds for any $\theta \in \left(0, \frac{3}{8}\right)$, we get that $\|(h^{3/8}\partial_\sigma - i\zeta_a)w_h\|_{L^2(\Omega)} = \widetilde{\mathcal{O}}(h^{3/4})$, thereby finishing the proof of Proposition 6.3. □

7. Lower bound

We fix a labeling $n \geq 1$ corresponding to the eigenvalue $\lambda_n(h)$ of the operator $\mathcal{P}_h$ introduced in (1-4). The purpose of this section is to obtain an accurate lower bound for $\lambda_n(h)$. This will be done by doing a spectral reduction via various auxiliary operators.

7A. ***Useful operators.*** We introduce operators, on the real line and in the plane, which will be useful to carry out a spectral reduction for the operator $\mathcal{P}_h$ and deduce the eigenvalue lower bounds that match with the established eigenvalue asymptotics in Theorem 1.2.

These new operators are defined via the spectral characteristics of the model operator introduced in Section 2B, namely, the spectral constants $\beta_a > 0$ and $\zeta_a < 0$ introduced in (1-10) and (1-12), and the positive normalized eigenfunction $\phi_a \in L^2(\mathbb{R})$ corresponding to β_a. We introduce the two operators

$$R_0^- : \psi \in L^2(\mathbb{R}^2) \mapsto \int_{\mathbb{R}} \phi_a(\tau)\psi(\,\cdot\,,\tau)\,d\tau \in L^2(\mathbb{R}), \tag{7-1}$$

$$R_0^+ : f \in L^2(\mathbb{R}) \mapsto f \otimes \phi_a \in L^2(\mathbb{R}^2), \tag{7-2}$$

where $(f \otimes \phi_a)(\sigma, \tau) := f(\sigma)\phi_a(\tau)$.

Note that $R_0^+ R_0^-$ is an orthogonal projector on $L^2(\mathbb{R}^2)$ whose image is $L^2(\mathbb{R}) \otimes \operatorname{span}(\phi_a)$. It is easy to check that the operator norms of $R_0^\pm$ are equal to 1; hence, for any $f \in L^2(\mathbb{R})$ and $\psi \in L^2(\mathbb{R}^2)$, we have

$$\|R_0^+ f\|_{L^2(\mathbb{R})} \leq \|f\|_{L^2(\mathbb{R})}, \quad \|R_0^- \psi\|_{L^2(\mathbb{R})} \leq \|\psi\|_{L^2(\mathbb{R}^2)}, \quad \|R_0^+ R_0^- \psi\|_{L^2(\mathbb{R}^2)} \leq \|\psi\|_{L^2(\mathbb{R}^2)}. \tag{7-3}$$

If we denote by π_a the projector in $L^2(\mathbb{R}_\tau)$ on the vector space generated by ϕ_a, we notice that

$$\Pi_0 := R_0^+ R_0^- = I \otimes \pi_a. \tag{7-4}$$

7B. ***Structure of bound states.*** Our aim is to determine a rough approximation of the bound state $\psi_{h,n}$ of $\mathcal{P}_h$ satisfying

$$\mathcal{P}_h \psi_{h,n} = \lambda_n(h)\psi_{h,n}, \tag{7-5}$$

this approximation being valid near the point of maximum curvature and reading as follows in the Frenet coordinates:

$$\tilde{\psi}_{h,n}(s,t) \approx h^{-5/16} e^{i\zeta_a s/h^{1/2}} \phi_a(h^{-1/2}t).$$

Associated with $\psi_{h,n}$, we introduced in (6-6) the function $u_{h,n}$ which can be seen as a function on $\mathbb{R}^2$ with L^2-norm satisfying (6-7). We recall that

$$u_{h,n}(\sigma,\tau)=h^{5/16}\chi(h^{\eta}\sigma)\chi(h^{\delta}\tau)\tilde{\psi}_{h,n}(h^{1/8}\sigma,h^{1/2}\tau),$$

where $\tilde{\psi}_{h,n}$ is the function assigned to $\psi_{h,n}$ by (A-3), $\chi\in C_c^{\infty}(\mathbb{R})$, $\operatorname{supp}\chi\subset[-1,1]$, $0\le\chi\le 1$ and $\chi=1$ on $\left[-\frac{1}{2},\frac{1}{2}\right]$.

We consider the function defined as

$$v_{h,n}(\sigma,\tau)=e^{-i\zeta_a\sigma/h^{3/8}}u_{h,n}(\sigma,\tau). \tag{7-6}$$

Approximating the function $v_{h,n}\sim\chi(h^{\eta}\sigma)\chi(h^{\delta}\tau)\phi_a(\tau)$ is the aim of the next proposition, which also yields an approximation of the bound state $\psi_{h,n}$ by the previous considerations.

Proposition 7.1. *Let $\mathcal{P}_h^{\mathrm{new}}$ be the operator in (4-6). The following hold*:

(1) $\|\mathcal{P}_h^{\mathrm{new}}v_{h,n}-(h^{-1}\lambda_n(h)-\beta_a)v_{h,n}\|_{L^2(\mathbb{R}^2)}=\mathcal{O}(h^{\infty})$.

(2) $\|v_{h,n}\|_{L^2(\mathbb{R}^2)}=1+\mathcal{O}(h^{1/2})$.

(3) $\|v_{h,n}-\Pi_0v_{h,n}\|_{L^2(\mathbb{R}^2)}=\mathcal{O}(h^{1/4})$.

(4) $\|\partial_\tau v_{h,n}-\partial_\tau\Pi_0v_{h,n}\|_{L^2(\mathbb{R}^2)}+\|\tau(v_{h,n}-\Pi_0v_{h,n})\|_{L^2(\mathbb{R}^2)}=\mathcal{O}(h^{1/4})$.

Proof. Proof of item (1). Let z_h be the function supported near Γ and defined in the Frenet coordinates by means of (A-3) as

$$\tilde{z}_h(s,t)=\chi(h^{-1/8+\eta}s)\chi(h^{-1/2+\delta}t). \tag{7-7}$$

We introduce the function involving the commutator of $\mathcal{P}_h$ and z_h acting on $\psi_{h,n}$,

$$f_h=[\mathcal{P}_h,z_h]\psi_{h,n}=(\mathcal{P}_hz_h-z_h\mathcal{P}_h)\psi_{h,n}. \tag{7-8}$$

By Remark 6.1, we may use the localization estimates in (5-7) and (5-8) with $g_h=\psi_{h,n}$ and $r_h=1$. Consequently,

$$\int_{\mathbb{R}^2}|\tilde{f}_h(s,t)|^2\,ds\,dt\le C\int_{\Omega}|f_h(x)|^2\,dx=\mathcal{O}(h^{\infty}),$$

where $\tilde{f}_h$ which is assigned to the function f_h in (7-8) is supported in the set

$$\left\{\left\{|s|\ge\tfrac{1}{2}h^{\eta-1/8}\right\}\cup\{|t|\ge\tfrac{1}{2}h^{\delta-1/2}\}\right\}\cap\{\{|s|\le h^{\eta-1/8}\}\cap\{|t|\le h^{\delta-1/2}\}\}.$$

We infer from (7-5), (4-2), (4-4) and (6-6),

$$\check{\mathcal{P}}_hu_{h,n}-\lambda_n(h)u_{h,n}=h^{5/16}\check{f}_h,$$

where

$$\check{f}_h(\sigma,\tau)=\tilde{f}_h(h^{1/8}\sigma,h^{1/2}\tau).$$

Consequently, after performing the change of variable ($\sigma=h^{-1/8}s$, $\tau=h^{-1/2}t$),

$$\|\check{\mathcal{P}}_hu_{h,n}-\lambda_n(h)u_{h,n}\|^2_{L^2(\mathbb{R}^2)}=\|\tilde{f}_h\|^2_{L^2(\mathbb{R}^2)}=\mathcal{O}(h^{\infty}). \tag{7-9}$$

By (4-6) and (7-6), we observe that

$$\check{\mathcal{P}}_h u_{h,n} = h e^{i\zeta_a \sigma/h^{3/8}} (\mathcal{P}_h^{\text{new}} + \beta_a) v_{h,n},$$

which after being inserted into (7-9), yields the estimate in item (1).

Remark 7.2. By (6-21), $\partial_\sigma v_{h,n} \in H^2(\mathbb{R}^2)$. Furthermore, by (6-19), the function f_h in (7-8) satisfies $\|\partial_\sigma \check{f}_h\|_{L^2(\mathbb{R}^2)} = \mathcal{O}(h^\infty)$. A slight adjustment of the proof of item (1) then yields

$$\|\partial_\sigma \mathcal{P}_h^{\text{new}} v_{h,n} - (h^{-1}\lambda_n(h) - \beta_a)\partial_\sigma v_{h,n}\|_{L^2(\mathbb{R}^2)} = \mathcal{O}(h^\infty).$$

Proof of item (2). By the normalization of $\psi_{h,n}$ and Remark 6.1, we have

$$1 = \int_\Omega |\psi_{h,n}|^2\,dx = \int_{\{|s(x)|<h^{-\eta+1/8}, |t(x)|<h^{-\delta+1/2}\}} |\psi_{h,n}|^2\,dx + \mathcal{O}(h^\infty),$$

$$\int_\Omega (1 - z_h^2)|\psi_h|^2\,dx = \mathcal{O}(h^\infty),$$

$$\int_\Omega \operatorname{dist}(x,\Gamma)|\psi_{h,n}|^2\,dx = \mathcal{O}(h^{1/2}).$$

We notice that the function z_h introduced above in (7-7) equals 1 in $\left\{|s(x)| < \frac{1}{2}h^{-\eta+1/8}, |t(x)| < \frac{1}{2}h^{-\delta+1/2}\right\}$.

Now we infer from (A-7)

$$\int_{\{|s|<h^{-\eta+1/8},\, |t|<h^{-\delta+1/8}\}} |\tilde{\psi}_{h,n}(s,t)|^2 |t|\,ds\,dt \le C \int_\Omega \operatorname{dist}(x,\Gamma)|\psi_{h,n}|^2\,dx = \mathcal{O}(h^{1/2})$$

and

$$\begin{aligned}\int_{\{|s|<h^{-\eta+1/8},\, |t|<h^{-\delta+1/8}\}} |\tilde{\psi}_{h,n}(s,t)|^2\,ds\,dt &= \int_{\{|s|<h^{-\eta+1/8},\, |t|<h^{-\delta+1/8}\}} |\tilde{\psi}_{h,n}(s,t)|^2(1 - tk(s))\,ds\,dt \\ &\quad + \int_{\{|s|<h^{-\eta+1/8},\, |t|<h^{-\delta+1/8}\}} |\tilde{\psi}_{h,n}(s,t)|^2\, tk(s)\,ds\,dt \\ &= 1 + \mathcal{O}(h^{1/2}).\end{aligned}$$

Similarly we get

$$\int_{\{|s|<\frac{1}{2}h^{-\eta+1/8},\, |t|<\frac{1}{2}h^{-\delta+1/8}\}} (1 - \tilde{z}_h^2)\,|\tilde{\psi}_{h,n}(s,t)|^2\,ds\,dt = \mathcal{O}(h^{1/2}).$$

Consequently, returning to (7-6), doing a change of variables and noticing that $\tilde{z}_h$ is supported in $\{|s| < h^{-\eta+1/8}, |t| < h^{-\delta+1/8}\}$, we get

$$\begin{aligned}\|v_{h,n}\|^2_{L^2(\mathbb{R}^2)} &= \int_{\{|s|<h^{-\eta+1/8},\, |t|<h^{-\delta+1/8}\}} |\tilde{\psi}_{h,n}|^2\,ds\,dt - \int_{\{|s|<h^{-\eta+1/8},\, |t|<h^{-\delta+1/8}\}} (1 - \tilde{z}_h^2)|\tilde{\psi}_h|^2\,ds\,dt \\ &= 1 + \mathcal{O}(h^{1/2}).\end{aligned}$$

Proof of items (3) *and* (4).

<u>Step 1</u>: We recall that the $\widetilde{\mathcal{O}}$ notation was introduced in (6-9). Note that Proposition 6.2 yields

$$\|h^{3/8}\partial_\sigma v_{h,n}\|_{L^2(\mathbb{R}^2)} = \widetilde{\mathcal{O}}(h^{3/8}). \tag{7-10}$$

By Remark 6.1, we can use (5-13) and (5-14) with $g_h = \psi_{h,n},\ r_h = 1$ (and $w_h = \check{u}_{h,n}$). In the same vein, we can use (5-5) and (5-6) too. Since $u_{h,n} = e^{i\zeta_a\sigma/h^{3/8}} v_{h,n}$, we get

$$\int_{\mathbb{R}^2} \big(|\partial_\tau v_{h,n}|^2 + |h^{3/8}\partial_\sigma v_{h,n} + i(b_a(\tau)\tau + \zeta_a)v_{h,n}|^2\big)\, d\tau\, d\sigma \le (\beta_a + \mathcal{O}(h^{1/2}))\|v_{h,n}\|^2_{L^2(\mathbb{R}^2)}. \tag{7-11}$$

By Cauchy's inequality and (7-10), we obtain, for any $\varepsilon > 0$,

$$\begin{aligned}\int_{\mathbb{R}^2} |h^{3/8}\partial_\sigma v_{h,n} + i(b_a(\tau)\tau+\zeta_a)v_{h,n}|^2\, d\sigma\, d\tau &\ge \int_{\mathbb{R}^2} \big((1-\varepsilon)|(b_a(\tau)\tau+\zeta_a)v_{h,n}|^2 - \varepsilon^{-1}|h^{3/8}\partial_\sigma v_{h,n}|^2\big)\, d\sigma\, d\tau \\ &\ge (1-\varepsilon)\int_{\mathbb{R}^2} |(b_a(\tau)\tau+\zeta_a)v_{h,n}|^2\, d\sigma\, d\tau - \widetilde{\mathcal{O}}(\varepsilon^{-1}h^{3/4}).\end{aligned}$$

We choose $\varepsilon = h^{3/8}$ and insert the resulting inequality into (7-11) to get

$$\int_{\mathbb{R}^2} \big(|\partial_\tau v_{h,n}|^2 + |(b_a(\tau)\tau + \zeta_a)v_{h,n}|^2\big)\, d\tau\, d\sigma \le \beta_a + \widetilde{\mathcal{O}}(h^{3/8}). \tag{7-12}$$

<u>Step 2</u>: In light of (7-4), let us introduce

$$r := \Pi_0 v_{h,n} \quad \text{and} \quad r_\perp := (I - \Pi_0)v_{h,n} = (I \otimes (I - \pi_a))v_{h,n}. \tag{7-13}$$

Using the last relation, and since the orthogonal projection π_a commutes with the operator $\mathfrak{h}_a[\zeta_a]$, we have the following two identities for almost every $\sigma \in \mathbb{R}$:

$$\int_{\mathbb{R}} |v_{h,n}(\sigma,\tau)|^2\, d\tau = \int_{\mathbb{R}} |r(\sigma,\tau)|^2\, d\tau + \int_{\mathbb{R}} |r_\perp(\sigma,\tau)|^2\, d\tau$$

and

$$\begin{aligned} q_{\zeta_a}(v_{h,n}(\sigma,\cdot)) &:= \int_{\mathbb{R}} \big(|\partial_\tau v_{h,n}(\sigma,\tau)|^2 + |(b_a(\tau)\tau + \zeta_a)v_{h,n}(\sigma,\tau)|^2\big)\, d\tau \\ &= q_{\zeta_a}(r(\sigma,\cdot)) + q_{\zeta_a}(r_\perp(\sigma,\cdot)) \\ &\ge \beta_a \int_{\mathbb{R}} |r(\sigma,\tau)|^2\, d\tau + \mu_2(\zeta_a)\int_{\mathbb{R}} |r_\perp(\sigma,\tau)|^2\, d\tau, \end{aligned} \tag{7-14}$$

by the min-max principle, where $\mu_2(\zeta_a)$ is the second eigenvalue of the operator $\mathfrak{h}_a[\zeta_a]$, satisfying $\mu_2(\zeta_a) > \beta_a$ (see Section 2A). Integrating with respect to σ, we get

$$\begin{aligned}\int_{\mathbb{R}^2} &\big(|\partial_\tau v_{h,n}(\sigma,\tau)|^2 + |(b_a(\tau)\tau + \zeta_a)v_{h,n}(\sigma,\tau)|^2\big)\, d\sigma\, d\tau \\ &\ge \beta_a \int_{\mathbb{R}^2} |r(\sigma,\tau)|^2\, d\sigma\, d\tau + \mu_2(\zeta_a)\int_{\mathbb{R}^2} |r_\perp(\sigma,\tau)|^2\, d\sigma\, d\tau. \end{aligned} \tag{7-15}$$

We deduce from (7-12) and the first item in Proposition 7.1

$$(\mu_2(\zeta_a) - \beta_a)\int_{\mathbb{R}^2} |r_\perp(\sigma,\tau)|^2\, d\sigma\, d\tau \le \widetilde{\mathcal{O}}(h^{3/8})\int_{\mathbb{R}^2} |r(\sigma,\tau)|^2\, d\sigma\, d\tau, \tag{7-16}$$

$$\int_{\mathbb{R}^2} |r(\sigma,\tau)|^2\, d\sigma\, d\tau = 1 + \widetilde{\mathcal{O}}(h^{3/8}), \tag{7-17}$$

$$\int_{\mathbb{R}^2} \big(|\partial_\tau r_\perp(\sigma,\tau)|^2 + |(b_a(\tau)\tau + \zeta_a)r_\perp(\sigma,\tau)|^2\big)\, d\sigma\, d\tau \le \widetilde{\mathcal{O}}(h^{3/8})\int_{\mathbb{R}^2} |r(\sigma,\tau)|^2\, d\sigma\, d\tau. \tag{7-18}$$

Step 3: Coming back to the definition of $r_\perp$ in (7-13), we still have to improve the error term in (7-16) to get the estimate of the third item in Proposition 7.1.

To that end, we will estimate the terms involving $\partial_\sigma v_{h,n}$ in (7-11). By (7-4) and dominated convergence, it is clear that Π_0 commutes with ∂_σ when acting on compactly supported functions of $H^1(\mathbb{R}^2)$:

$$\Pi_0 \partial_\sigma = \partial_\sigma \Pi_0. \tag{7-19}$$

By (2-11), ϕ_a is orthogonal to $(b_a(\tau)\tau + \zeta_a)\phi_a$ in $L^2(\mathbb{R})$, so

$$\pi_a (b_a(\tau)\tau + \zeta_a)\pi_a = 0,$$

which implies, by taking the tensor product,

$$\Pi_0 (b_a(\tau)\tau + \zeta_a)\Pi_0 = 0. \tag{7-20}$$

By (7-13), (7-19) and (7-20), we get

$$\langle r(\sigma,\tau), i(b_a(\tau)\tau + \zeta_a)\partial_\sigma r(\sigma,\tau)\rangle_{L^2(\mathbb{R}^2)} = 0.$$

Now, we inspect the term

$$\begin{aligned}
&\langle \partial_\sigma v_{h,n}, i(b_a(\tau)\tau + \zeta_a) r\rangle_{L^2(\mathbb{R}^2)} \\
&\qquad = -\langle v_{h,n}, i(b_a(\tau)\tau + \zeta_a)\partial_\sigma r\rangle_{L^2(\mathbb{R}^2)} \\
&\qquad = -\underbrace{\langle r, i(b_a(\tau)\tau + \zeta_a)\partial_\sigma r\rangle_{L^2(\mathbb{R}^2)}}_{=0} - \langle r_\perp, i(b_a(\tau)\tau + \zeta_a)\partial_\sigma r\rangle_{L^2(\mathbb{R}^2)} \\
&\qquad = -\langle r_\perp, i(b_a(\tau)\tau + \zeta_a)\partial_\sigma r\rangle_{L^2(\mathbb{R}^2)} = -\langle (b_a(\tau)\tau + \zeta_a) r_\perp, i\partial_\sigma r\rangle_{L^2(\mathbb{R}^2)}.
\end{aligned} \tag{7-21}$$

Since

$$\begin{aligned}
\|h^{3/8}\partial_\sigma r\|_{L^2(\mathbb{R}^2)} &= h^{3/8}\|\Pi_0 \partial_\sigma v_{h,n}\|_{L^2(\mathbb{R}^2)} && \text{(by (7-19))} \\
&\le h^{3/8}\|\partial_\sigma v_{h,n}\|_{L^2(\mathbb{R}^2)} && \text{(by (7-3))} \\
&= \widetilde{\mathcal{O}}(h^{3/8}) && \text{(by (7-10))},
\end{aligned}$$

we get by the Cauchy–Schwarz inequality, (7-21) and (7-18)

$$h^{3/8}|\langle \partial_\sigma v_{h,n}, i(b_a(\tau)\tau + \zeta_a) r\rangle_{L^2(\mathbb{R}^2)}| \le \|(b_a(\tau)\tau + \zeta_a) r_\perp\|_{L^2(\mathbb{R}^2)} \|h^{3/8}\partial_\sigma r\|_{L^2(\mathbb{R}^2)} = \widetilde{\mathcal{O}}(h^{9/16}). \tag{7-22}$$

Now, we can estimate the following inner product term by using (7-13) and (7-22):

$$\begin{aligned}
&\langle h^{3/8}\partial_\sigma v_{h,n}, i(b_a(\tau)\tau + \zeta_a) v_{h,n}\rangle_{L^2(\mathbb{R}^2)} \\
&\qquad = \langle h^{3/8}\partial_\sigma v_{h,n}, i(b_a(\tau)\tau + \zeta_a) r_\perp\rangle_{L^2(\mathbb{R}^2)} + \langle h^{3/8}\partial_\sigma v_{h,n}, i(b_a(\tau)\tau + \zeta_a) r\rangle_{L^2(\mathbb{R}^2)} \\
&\qquad = \langle h^{3/8}\partial_\sigma v_{h,n}, i(b_a(\tau)\tau + \zeta_a) r_\perp\rangle_{L^2(\mathbb{R}^2)} + \widetilde{\mathcal{O}}(h^{9/16}).
\end{aligned} \tag{7-23}$$

By the Cauchy–Schwarz inequality, (7-10), (7-18) and (7-23), we get

$$\begin{aligned}
|\langle h^{3/8}\partial_\sigma v_{h,n}, i(b_a(\tau)\tau + \zeta_a) v_{h,n}\rangle_{L^2(\mathbb{R}^2)}| &\le \|h^{3/8}\partial_\sigma v_{h,n}\| \, \|(b_a(\tau)\tau + \zeta_a) r_\perp\| + \widetilde{\mathcal{O}}(h^{9/16}) \\
&= \widetilde{\mathcal{O}}(h^{9/16}) = o(h^{1/2}).
\end{aligned} \tag{7-24}$$

Consequently,

$$\begin{aligned}
&\|h^{3/8}\partial_\sigma v_{h,n} + i(b_a(\tau)\tau+\zeta_a)v_{h,n}\|^2_{L^2(\mathbb{R}^2)}\\
&\qquad = \|h^{3/8}\partial_\sigma v_{h,n}\|^2_{L^2(\mathbb{R}^2)} + \|(b_a(\tau)\tau+\zeta_a)v_{h,n}\|^2_{L^2(\mathbb{R}^2)} + 2\,\mathrm{Re}\langle h^{3/8}\partial_\sigma v_{h,n}, i(b_a(\tau)\tau+\zeta_a)v_{h,n}\rangle_{L^2(\mathbb{R}^2)}\\
&\qquad \geq \|(b_a(\tau)\tau+\zeta_a)v_{h,n}\|^2_{L^2(\mathbb{R}^2)} + o(h^{1/2}).
\end{aligned}$$

Inserting the previous inequality into (7-11) we get the following improvement of (7-12):

$$\int_{\mathbb{R}^2} (|\partial_\tau v_{h,n}|^2 + |(b_a(\tau)\tau+\zeta_a)v_{h,n}|^2)\,d\tau\,d\sigma \leq \beta_a + \mathcal{O}(h^{1/2}). \tag{7-25}$$

Step 4: Now we are ready to finish the proof of items (3) and (4). By (7-15) and (7-14), we infer from (7-25) and (7-13),

$$(\mu_2(\zeta_a)-\beta_a)\int_{\mathbb{R}^2} |r_\perp(\sigma,\tau)|^2\,d\sigma\,d\tau \leq \mathcal{O}(h^{1/2})\int_{\mathbb{R}^2} |r(\sigma,\tau)|^2\,d\sigma\,d\tau,$$

$$\int_{\mathbb{R}^2} \big(|\partial_\tau r_\perp(\sigma,\tau)|^2 + |(b_a(\tau)\tau+\zeta_a)r_\perp(\sigma,\tau)|^2\big)\,d\sigma\,d\tau \leq \mathcal{O}(h^{1/2})\int_{\mathbb{R}^2} |r(\sigma,\tau)|^2\,d\sigma\,d\tau.$$

With (7-17) in hand, we get the estimates of items (3) and (4) of Proposition 7.1. □

7C. *Projection on a refined quasimode.* We wish to improve the approximation $v_{h,n} \sim \chi(h^\eta\sigma)\chi(h^\delta\tau)\phi_a(\tau)$ obtained in Proposition 7.1 by two ways which eventually are correlated: displaying the curvature effects in $v_{h,n}$ and getting better estimates of the errors. Along the proof of Proposition 7.1, curvature effects were neglected and absorbed in the error terms. Not neglecting the curvature, we get the approximation $v_{h,n} \sim \chi(h^\eta\sigma)\chi(h^\delta\tau)\phi_{a,h}(\tau)$, where $\phi_{a,h}(\tau)$ corrects $\phi_a(\tau)$ via curvature-dependent terms (see (7-31)). This is precisely stated in Proposition 7.3 after introducing the necessary preliminaries.

7C1. *Preliminaries.* In this subsection, we write $\kappa = k(0) = k_{\max}$ and $k_2 = k''(0)$. We consider the weighted L^2 space

$$X_{h,\delta} = L^2\big((-h^{-\delta}, h^{-\delta}); (1-h^{1/2}\kappa\tau)\,d\tau\big) \tag{7-26}$$

endowed with the Hilbertian norm

$$\|f\|_{X_{h,\delta}} = \left(\int_{-h^{-\delta}}^{h^{-\delta}} |f(\tau)|^2(1-h^{1/2}\kappa\tau)\,d\tau\right)^{1/2}.$$

This norm is equivalent to the usual norm of $L^2(-h^{-\delta}, h^{-\delta})$ provided h is sufficiently small.

With domain $H^2(-h^{-\delta}, h^{-\delta}) \cap H^1_0(-h^{-\delta}, h^{-\delta})$, consider the operator in (3-1) for $\xi = \zeta_a$:

$$\begin{aligned}
\mathcal{H}_{a,\kappa,h} = -\frac{d^2}{d\tau^2} + (b_a(\tau)\tau+\zeta_a)^2 + \kappa h^{1/2}(1-\kappa h^{1/2}\tau)^{-1}\partial_\tau + 2\kappa h^{1/2}\tau\Big(b_a(\tau)\tau+\zeta_a-\kappa h^{1/2}b_a(\tau)\frac{\tau^2}{2}\Big)^2\\
-\kappa h^{1/2}b_a(\tau)\tau^2(b_a(\tau)\tau+\zeta_a) + \kappa^2 h b_a(\tau)^2\frac{\tau^4}{4},
\end{aligned} \tag{7-27}$$

which is self-adjoint on the space $X_{h,\delta}$. This operator can be decomposed as follows:

$$\mathcal{H}_{a,\kappa,h} = \mathfrak{h}[\zeta_a] + \kappa h^{1/2}\mathfrak{h}^{(1)}[\zeta_a] + hL_h, \tag{7-28}$$

where $\mathfrak{h}[\zeta_a]$ is introduced in (2-1) and

$$\mathfrak{h}^{(1)}[\zeta_a] = \partial_\tau + 2\tau(b_a(\tau)\tau + \zeta_a)^2 - b_a(\tau)\tau^2(b_a(\tau)\tau + \zeta_a) \tag{7-29}$$

and

$$L_h = q_{1,h}(\tau)\partial_\tau + q_{2,h}(\tau), \quad \text{with } |q_{1,h}(\tau)| \le C_1|\tau|, \quad |q_{2,h}(\tau)| \le C_2(1+|\tau|^5), \tag{7-30}$$

where C_1, C_2 are positive constants independent of h, τ.

We introduce the following quasimode in the space $X_{h,\delta}$:

$$\phi_{a,h}(\tau) = \chi(h^\delta\tau)\big(\phi_a(\tau) + h^{1/2}\kappa\,\phi_a^{\mathrm{cor}}(\tau)\big), \tag{7-31}$$

where $\chi \in C_c^\infty(\mathbb{R}; [0,1])$, $\operatorname{supp}\chi \subset [-1,1]$, $\chi = 1$ on $\big[-\frac{1}{2}, \frac{1}{2}\big]$. The function ϕ_a is the positive ground state of $\mathfrak{h}[\zeta_a]$ with corresponding ground state energy β_a:

$$(\mathfrak{h}[\zeta_a] - \beta_a)\phi_a = 0.$$

We now explain the construction of ϕ_a^{cor}. By (7-28), starting from some ϕ_a^{cor} to be determined,

$$\begin{aligned}(\mathcal{H}_{a,\kappa,h} - \beta_a - h^{1/2}\kappa M_3(a))(\phi_a + h^{1/2}\kappa\,\phi_a^{\mathrm{cor}}) \\ = \kappa h^{1/2}\big((\mathfrak{h}[\zeta_a]-)\phi_a^{\mathrm{cor}} + (\mathfrak{h}^{(1)}[\zeta_a] - M_3(a))\phi_a\big) + h\mathcal{R}_{a,h},\end{aligned} \tag{7-32}$$

where

$$\mathcal{R}_{a,h} = L_h(\phi_a + h^{1/2}\kappa\,\phi_a^{\mathrm{cor}}) + \kappa^2(\mathfrak{h}^{(1)}[\zeta_a] - M_3(a))\phi_a^{\mathrm{cor}}.$$

Note that, by Remark 2.3, $\mathfrak{h}^{(1)}[\zeta_a]\phi_a - M_3(a)\phi_a$ is orthogonal to ϕ_a in $L^2(\mathbb{R})$. Hence we can choose

$$\phi_a^{\mathrm{cor}} = -\mathfrak{R}_a(\mathfrak{h}^{(1)}[\zeta_a]\phi_a - M_3(a)\phi_a), \tag{7-33}$$

so that the coefficient of $h^{1/2}$ in (7-32) vanishes. In this way, we infer from (7-32),

$$\big(\mathcal{H}_{a,\kappa,h} - \beta_a - h^{1/2}\kappa M_3(a)\big)(\phi_a + h^{1/2}\kappa\phi_a^{\mathrm{cor}}) = h\mathcal{R}_{a,h}.$$

Notice that $\phi_{a,h}$ is constructed so that it has compact support in $(-h^{-\delta}, h^{-\delta})$ and hence satisfies the Dirichlet conditions at $\tau = \pm h^{-\delta}$. Since, ϕ_a and ϕ_a^{cor} decay exponentially at infinity by Lemma 2.4, we deduce

$$\|\mathcal{H}_{a,\kappa,h}\phi_{a,h} - (\beta_a + h^{1/2}\kappa M_3(a))\phi_{a,h}\|_{X_{h,\delta}} = \mathcal{O}(h). \tag{7-34}$$

We denote by $\phi_{a,h}^{\mathrm{gs}}$ the normalized ground state of the Dirichlet realization of $\mathcal{H}_{a,\kappa,h}$ in the weighted space $X_{h,\delta}$ (i.e., in $L^2((-h^{-\delta}, h^{-\delta}); (1-h^{1/2}\kappa\tau)d\tau)$). By (3-8), the min-max principle and Proposition 3.2, we have

$$\lambda_1(\mathcal{H}_{a,\kappa,h}) = \beta_a + h^{1/2}\kappa M_3(a) + \mathcal{O}(h) \quad \text{and} \quad \lambda_2(\mathcal{H}_{a,\kappa,h}) \ge \mu_2(\zeta_a) + o(1), \tag{7-35}$$

so we infer from (7-34) and the Hölder inequality

$$\big\langle(\mathcal{H}_{a,\kappa,h}\phi_{a,h} - \lambda_1(\mathcal{H}_{a,\kappa,h}))(\phi_{a,h}^{\mathrm{gs}} - \phi_{a,h}), \phi_{a,h}^{\mathrm{gs}} - \phi_{a,h}\big\rangle_{X_{h,\delta}} = \mathcal{O}(h)\|\phi_{a,h}^{\mathrm{gs}} - \phi_{a,h}\|_{X_{h,\delta}}.$$

Thus, by the spectral theorem,

$$\|\phi_{a,h}^{\mathrm{gs}} - \phi_{a,h}\|_{X_{h,\delta}} + \|\tau(\phi_{a,h}^{\mathrm{gs}} - \phi_{a,h})\|_{X_{h,\delta}} + \|\partial_\tau(\phi_{a,h}^{\mathrm{gs}} - \phi_{a,h})\|_{X_{h,\delta}} = \mathcal{O}(h). \tag{7-36}$$

7C2. *New projections.* We fix $h_0 > 0$ so that $1 - h_0^{1/2-\delta}\kappa > \frac{1}{2}$. In the sequel, the parameter h varies in the interval $(0, h_0)$. Consider the space

$$X^2_{h,\delta} = L^2\big(\mathbb{R} \times (-h^{-\delta}, h^{\delta}); (1 - h^{1/2}\kappa\tau)\, d\sigma\, d\tau\big) \tag{7-37}$$

endowed with the weighted norm

$$\|v\|_{X^2_{h,\delta}} = \left(\int_{\mathbb{R}} \int_{-h^{-\delta}}^{h^{-\delta}} |v(\sigma, \tau)|^2 (1 - h^{1/2}\kappa\tau)\, d\sigma\, d\tau\right)^{1/2},$$

which is equivalent to the usual norm of $L^2(\mathbb{R} \times (-h^{-\delta}, h^{\delta}))$.

We introduce the two operators

$$R_h^- : v \in X^2_{h,\delta} \mapsto \int_{\mathbb{R}} \phi_{a,h}(\tau) v(\,\cdot\,, \tau)(1 - h^{1/2}\kappa\tau)\, d\tau \in L^2(\mathbb{R}), \tag{7-38}$$

$$R_h^+ : f \in L^2(\mathbb{R}) \mapsto f \otimes \phi_{a,h} \in X^2_{h,\delta}, \quad \text{where } f \otimes \phi_{a,h}(\sigma, \tau) = f(\sigma)\phi_{a,h}(\tau). \tag{7-39}$$

The image of $R_h^+ R_h^-$ is $L^2(\mathbb{R}) \otimes \operatorname{span}(\phi_{a,h})$. Furthermore, for all $v \in X^2_{h,\delta}$, the functions $R_h^+ R_h^- v$ and $v - R_h^+ R_h^- v$ are orthogonal in $X^2_{h,\delta}$, since the operator $R_h^+ R_h^-$ can be expressed as

$$\Pi_h := R_h^+ R_h^- = I \otimes \pi_{a,h}, \tag{7-40}$$

where $\pi_{a,h}$ is the orthogonal projection, in the weighted Hilbert space $X_{h,\delta}$, on the space $\operatorname{span} \phi_{a,h}$. With this projection in hand, we can approximate the truncated bound state $v_{h,n}$, introduced in (7-6), with better error terms, thereby improving Proposition 7.1.

Proposition 7.3. *The following holds*:

$$\|v_{h,n} - \Pi_h v_{h,n}\|_{X^2_{h,\delta}} + \|\partial_\tau(v_{h,n} - \Pi_h v_{h,n})\|_{X^2_{h,\delta}} + \|\tau(v_{h,n} - \Pi_h v_{h,n})\|_{X^2_{h,\delta}} = \widetilde{\mathcal{O}}(h^{5/16}),$$

where Π_h *is the projection in* (7-40).

Remark 7.4. By (7-31) and (7-32), we observe that

$$\|(\Pi_h - \Pi_0)v_{h,n}\|_{L^2(\mathbb{R}^2)} + \|(\partial_\tau \Pi_h - \partial_\tau \Pi_0)v_{h,n}\|_{L^2(\mathbb{R}^2)} + \|\tau(\Pi_h - \Pi_0)v_{h,n}\|_{L^2(\mathbb{R}^2)} = \mathcal{O}(h^{1/2}),$$

where Π_0 is the projection introduced in (7-4). Since the norm of $X^2_{h,\delta}$ is equivalent to the usual norm of L^2, Proposition 7.3 yields the following improvement of Proposition 7.1:

$$\|v_{h,n} - \Pi_0 v_{h,n}\|_{L^2(\mathbb{R}^2)} + \|\partial_\tau(v_{h,n} - \Pi_0 v_{h,n})\|_{L^2(\mathbb{R}^2)} + \|\tau(v_{h,n} - \Pi_0 v_{h,n})\|_{L^2(\mathbb{R}^2)} = \widetilde{\mathcal{O}}(h^{5/16}), \tag{7-41}$$

where Π_0 is the projection in (7-4). This remark will be useful in the next subsection.

Proof of Proposition 7.3. Step 1: We give here preliminary estimates that we will use in Step 3 below. Firstly, by Remark 6.1,

$$\int_{\mathbb{R}^2} \tau^4 |v_{h,n}(\sigma, \tau)|^2\, d\sigma\, d\tau = \mathcal{O}(1). \tag{7-42}$$

Secondly, we will prove that

$$\langle h^{3/8}\partial_\sigma v_{h,n}, (b_a(\tau)\tau+\zeta_a)v_{h,n}\rangle_{L^2(\mathbb{R}^2)} = \widetilde{\mathcal{O}}(h^{5/8}). \tag{7-43}$$

By (7-10) and Proposition 7.1,

$$\begin{aligned}|\langle h^{3/8}\partial_\sigma v_{h,n}, (b_a(\tau)\tau+\zeta_a)(v_{h,n}-\Pi_0 v_{h,n})\rangle_{L^2(\mathbb{R}^2)}| \\ \le \|h^{3/8}\partial_\sigma v_{h,n}\|_{L^2(\mathbb{R}^2)}\|(b_a(\tau)\tau+\zeta_a)(v_{h,n}-\Pi_0 v_{h,n})\|_{L^2(\mathbb{R}^2)} = \widetilde{\mathcal{O}}(h^{5/8}).\end{aligned}$$

Similarly, using (7-19) and Hölder's inequality, we write

$$\begin{aligned}|\langle (b_a(\tau)\tau+\zeta_a)h^{3/8}\partial_\sigma \Pi_0 v_{h,n}, v_{h,n}-\Pi_0 v_{h,n}\rangle_{L^2(\mathbb{R}^2)}| \\ \le \|h^{3/8}\Pi_0\partial_\sigma v_{h,n}\|_{L^2(\mathbb{R}^2)}\|(b_a(\tau)\tau+\zeta_a)(v_{h,n}-\Pi_0 v_{h,n})\|_{L^2(\mathbb{R}^2)} = \widetilde{\mathcal{O}}(h^{5/8}).\end{aligned}$$

Now, writing $v_{h,n} = \Pi_0 v_{h,n} + (v_{h,n}-\Pi_0 v_{h,n})$ and collecting the foregoing estimates, we get

$$\begin{aligned}&\langle h^{3/8}\partial_\sigma v_{h,n}, (b_a(\tau)\tau+\zeta_a)v_{h,n}\rangle_{L^2(\mathbb{R}^2)} \\ &\quad = \langle h^{3/8}\partial_\sigma v_{h,n}, (b_a(\tau)\tau+\zeta_a)\Pi_0 v_{h,n}\rangle_{L^2(\mathbb{R}^2)} + \widetilde{\mathcal{O}}(h^{5/8}) \\ &\quad = -\langle (b_a(\tau)\tau+\zeta_a)v_{h,n}, h^{3/8}\partial_\sigma \Pi_0 v_{h,n}\rangle_{L^2(\mathbb{R}^2)} + \widetilde{\mathcal{O}}(h^{5/8}) \quad \text{(by integration by parts).}\end{aligned}$$

Again, decomposing $v_{h,n}$ by the projection Π_0 and observing that (7-20) yields

$$\langle (b_a(\tau)\tau+\zeta_a)\Pi_0 v_{h,n}, h^{3/8}\partial_\sigma \Pi_0 v_{h,n}\rangle_{L^2(\mathbb{R}^2)} = 0,$$

we get

$$\begin{aligned}&\langle h^{3/8}\partial_\sigma v_{h,n}, (b_a(\tau)\tau+\zeta_a)v_{h,n}\rangle_{L^2(\mathbb{R}^2)} \\ &\quad = -\langle (b_a(\tau)\tau+\zeta_a)h^{3/8}\partial_\sigma \Pi_0 v_{h,n}, v_{h,n}-\Pi_0 v_{h,n}\rangle_{L^2(\mathbb{R}^2)} + \widetilde{\mathcal{O}}(h^{5/8}) = \widetilde{\mathcal{O}}(h^{5/8}),\end{aligned}$$

thereby obtaining (7-43).

<u>Step 2</u>: We introduce operators involving the ground state $\phi^{\mathrm{gs}}_{a,h}$ as follows. First we introduce the operators

$$\widetilde{R}_h^- : v \in X^2_{h,\delta} \mapsto \int_{\mathbb{R}} \phi^{\mathrm{gs}}_{a,h}(\tau) v(\,\cdot\,,\tau)(1-h^{1/2}\kappa\tau)\,d\tau \in L^2(\mathbb{R}), \tag{7-44}$$

$$\widetilde{R}_h^+ : f \in L^2(\mathbb{R}) \mapsto f\otimes\phi^{\mathrm{gs}}_{a,h} \in X^2_{h,\delta}, \quad \text{where } (f\otimes\phi^{\mathrm{gs}}_{a,h})(\sigma,\tau) = f(\sigma)\phi^{\mathrm{gs}}_{a,h}(\tau). \tag{7-45}$$

Denoting by $\tilde{\pi}_{a,h}$ the orthogonal projection, in $X_{h,\delta}$, on the space span $\phi^{\mathrm{gs}}_{a,h}$, we introduce

$$\widetilde{\Pi}_h := \widetilde{R}_h^+\widetilde{R}_h^- = I\otimes\tilde{\pi}_{a,h}. \tag{7-46}$$

By (7-36) and (7-40), we observe that, for all $g \in X_{h,\delta}$ and $f \in X^2_{h,,\delta}$, we have

$$\|(\widetilde{R}_h^- - R_h^-)g\|_{X_{h,\delta}} = \mathcal{O}(h)\|g\|_{X_{h,\delta}}, \quad \|(\widetilde{\Pi}_h - \Pi_h)f\|_{X^2_{h,\delta}} = \mathcal{O}(h)\|f\|_{X^2_{h,\delta}}.$$

So if we prove that

$$\|v_{h,n}-\widetilde{\Pi}_h v_{h,n}\|_{X^2_{h,\delta}} + \|\partial_\tau(v_{h,n}-\widetilde{\Pi}_h v_{h,n})\|_{X^2_{h,\delta}} + \|\tau(v_{h,n}-\widetilde{\Pi}_h v_{h,n})\|_{X^2_{h,\delta}} = \widetilde{\mathcal{O}}(h^{5/16}), \tag{7-47}$$

then we deduce the estimate in Proposition 7.3.

Step 3: Adapting the proof of Proposition 7.1, we prove now (7-47). By Remark 6.1, we can use (5-14) with $w_h = u_{h,n}$, $r_h = 1$, $m_h = \|u_{h,n}\|^2_{X^2_{h,\delta}} = 1 + \mathcal{O}(h^{1/2})$ and $\theta = \frac{1}{4}$. Thus

$$\int_{\mathbb{R}} \int_{-h^{-\delta}}^{h^{-\delta}} \Big(|\partial_\tau u_{h,n}|^2 + (1+2\kappa h^{1/2}\tau) \Big| \Big(h^{3/8}\partial_\sigma + i\Big(b_a\tau - \kappa h^{1/2} b_a \frac{\tau^2}{2}\Big)\Big) u_{h,n} \Big|^2 \Big)(1-\kappa h^{1/2}\tau)\, d\sigma\, d\tau \\ \leq (\beta_a + h^{1/2} M_3(a)\kappa + \mathcal{O}(h^{3/4}))\|u_{h,n}\|^2_{X^2_{h,\delta}}. \tag{7-48}$$

Since $u_{h,n} = e^{i\zeta_a\sigma/h^{3/8}} v_{h,n}$ (by (7-6)), we get

$$\int_{\mathbb{R}} \int_{-h^{-\delta}}^{h^{-\delta}} |\partial_\tau v_{h,n}|^2 (1-\kappa h^{1/2}\tau)\, d\sigma\, d\tau \\ + \int_{\mathbb{R}} \int_{-h^{-\delta}}^{h^{-\delta}} (1+2\kappa h^{1/2}\tau) \Big| \Big(h^{3/8}\partial_\sigma + i\Big(b_a\tau + \zeta_a - \kappa h^{1/2} b_a \frac{\tau^2}{2}\Big)\Big) v_{h,n} \Big|^2 (1-\kappa h^{1/2}\tau)\, d\sigma\, d\tau \\ \leq (\beta_a + h^{1/2} M_3(a)\kappa + \mathcal{O}(h^{3/4}))\|v_{h,n}\|^2_{X^2_{h,\delta}}. \tag{7-49}$$

Using (7-10), (7-43) and (7-42), we deduce the following estimate from (7-49):

$$\int_{\mathbb{R}} \int_{-h^{-\delta}}^{h^{-\delta}} \Big(|\partial_\tau v_{h,n}|^2 + (1+2\kappa h^{1/2}\tau) \Big| \Big(b_a\tau + \zeta_a - \kappa h^{1/2} b_a \frac{\tau^2}{2}\Big) v_{h,n} \Big|^2 \Big)(1-\kappa h^{1/2}\tau)\, d\sigma\, d\tau \\ \leq (\beta_a + h^{1/2} M_3(a)\kappa + \widetilde{\mathcal{O}}(h^{5/8}))\|v_{h,n}\|^2_{X^2_{h,\delta}}, \tag{7-50}$$

where we used also that $\|v_{h,n}\|^2_{X^2_{h,\delta}} = 1 + \mathcal{O}(h^{1/2})$, by (6-7) and (7-6).

Now we get (7-47) by decomposing $v_{h,n}$ in $X^2_{h,\delta}$ in the form

$$v_{h,n} = \tilde{r}_h + \tilde{r}_{h,\perp}, \quad \tilde{r}_h := \widetilde{\Pi}_h v_{h,n}, \quad \tilde{r}_{h,\perp} = (I - \widetilde{\Pi}_h) v_{h,n},$$

and by using the spectral asymptotics for the operator $\mathcal{H}_{h,a,\kappa}$, recalled in (7-35). □

7D. *Quasimodes for the effective operator.* Let us start with some heuristic considerations. The derivation of the eigenvalue upper bound of Theorem 4.1 suggested in the tangent variable the following one-dimensional effective operator (see (4-25)):

$$H_a^{\text{harm}} = -c_2(a)\partial_\sigma^2 - \frac{M_3(a)k''(0)}{2}\sigma^2, \tag{7-51}$$

where $c_2(a) > 0$ is introduced in (1-12).

Moreover, by Remark 4.3, it is natural to consider the quasimode

$$v_{h,n}^{\text{app}} = \big(\phi_a(\tau) + 2\mathfrak{R}_a((\zeta_a + b_a(\tau)\tau)\phi_a) i h^{3/8}\partial_\sigma + k_{\max} h^{1/2}\phi_a^{\text{cor}}(\tau)\big) f_n(\sigma),$$

where $\mathfrak{R}_a$ is the regularized resolvent introduced in (2-18), ϕ_a^{cor} is the function in (7-33), and f_n is the normalized n-th eigenfunction of the operator H_a^{harm}. Denoting by $\Pi_{h,n}^{\text{app}}$ the orthogonal projection, in $L^2(\mathbb{R}^2)$, on the space generated by $v_{h,n}^{\text{app}}$, we observe formally, by neglecting the terms with coefficients having order lower than $h^{3/4}$,

$$c_2(a)\Pi_{h,n}^{\text{app}}\mathcal{P}_h^{\text{new}} \approx h^{1/2}(M_3(a)k_{\max} + h^{1/4}H_a^{\text{harm}})\Pi_n^{\text{new}},$$

where Π_n^{new} is the projection, in $L^2(\mathbb{R}^2)$, on the space generated by the function $\varphi_a(\tau) f_n(\sigma)$, and

$$\varphi_a(\tau) := \phi_a(\tau) - 4(b_a(\tau)\tau + \zeta_a)\mathfrak{R}_a((b_a(\tau)\tau + \zeta_a)\phi_a(\tau)). \tag{7-52}$$

Guided by these heuristic observations, we will use the truncated bound state $v_{h,n}$ in (7-6) to construct quasimodes of the operator H_a^{harm} by projecting $v_{h,n}$ on the vector space generated by the function φ_a introduced in (7-52). To that end, we introduce the operator

$$R_0^{\text{new}} : v \in L^2(\mathbb{R}^2) \mapsto \int_{\mathbb{R}} \varphi_a(\tau) v(\,\cdot\,, \tau)\, d\tau \in L^2(\mathbb{R}). \tag{7-53}$$

We will prove the following proposition.

Proposition 7.5. *Let $n \in \mathbb{N}$ be fixed. The following hold*:

(1) $\|R_0^{\text{new}} v_{h,n} - (1 - 4I_2(a)) R_0^- v_{h,n}\|_{L^2(\mathbb{R})} = \mathcal{O}(h^{1/4})$, *where R_0^- is the operator in (7-1) and $I_2(a)$ is introduced in (2-17).*

(2) $\|R_0^{\text{new}} v_{h,n}\|_{L^2(\mathbb{R})} = 1 - 4I_2(a) + \mathcal{O}(h^{1/4})$.

(3) *For every $n \in \mathbb{N}$, there exists $h_n > 0$ such that, for all $h \in (0, h_n)$,*

$$\langle R_0^{\text{new}} v_{h,k}, R_0^{\text{new}} v_{h,k'} \rangle_{L^2(\mathbb{R})} = (1 - 4I_2(a))^2 \delta_{k,k'} + o(1) \quad (1 \le k, k' \le n), \tag{7-54}$$

and

$$M_n = \operatorname{span}(R_0^{\text{new}} v_{h,k},\ 1 \le k \le n) \quad \textit{satisfies} \quad \dim(M_n) = n. \tag{7-55}$$

(4) *We have as $h \to 0_+$*

$$\left\langle (H_a^{\text{harm}} - h^{-3/4}\Lambda_n(h)) R_0^{\text{new}} v_{h,n}, R_0^{\text{new}} v_{h,n} \right\rangle_{L^2(\mathbb{R})} = o(1) \|R_0^{\text{new}} v_{h,n}\|^2_{L^2(\mathbb{R})},$$

where

$$\Lambda_n(h) = h^{-1}\lambda_n(h) - \beta_a - M_3(a) k_{\max} h^{1/2},$$

and H_a^{harm} is the operator introduced in (7-51).

Proof. Proof of item (1). Consider $\Pi_0 = R_0^+ R_0^-$ the projection introduced in (7-4). By (2-17), $R_0^{\text{new}} R_0^+ = (1 - 4I_2(a))\mathrm{Id}$; hence, composing by R_0^- on the right gives

$$R_0^{\text{new}} \Pi_0 = (1 - 4I_2(a)) R_0^-.$$

Writing $v_{h,n} = \Pi_0 v_{h,n} + (v_{h,n} - \Pi_0 v_{h,n})$, we get

$$\begin{aligned} R_0^{\text{new}} v_{h,n} &= R_0^{\text{new}} \Pi_0 v_{h,n} + R_0^{\text{new}} (v_{h,n} - \Pi_0 v_{h,n}) \\ &= (1 - 4I_2(a)) R_0^- v_{h,n} + R_0^{\text{new}} (v_{h,n} - \Pi_0 v_{h,n}). \end{aligned}$$

Then we observe that

$$\|R_0^{\text{new}} (v_{h,n} - \Pi_0 v_{h,n})\|_{L^2(\mathbb{R})} \le \|\varphi_a\|_{L^2(\mathbb{R})} \|v_{h,n} - \Pi_0 v_{h,n}\|_{L^2(\mathbb{R}^2)} = \mathcal{O}(h^{1/4})$$

by Hölder's inequality and Proposition 7.1. This yields the conclusion of item (1).

Proof of item (2). By (2-20), $1-4I_2(a)>0$. By (7-1) and Proposition 7.1, we have

$$\|R_0^- v_{h,n}\|_{L^2(\mathbb{R})} = \|\Pi_0 v_{h,n}\|_{L^2(\mathbb{R}^2)} = 1+\mathcal{O}(h^{1/4}).$$

Now item (2) follows from item (1).

Proof of item (3). If $1\le k,k'\le n$ and $k\neq k'$, we have as $h\to 0_+$,

$$\langle v_{h,k}, v_{h,k'}\rangle_{L^2(\mathbb{R}^2)} = o(1)+\delta_{k,k'}.$$

By Proposition 7.1, we get further

$$\langle R_0^- v_{h,k}, R_0^- v_{h,k'}\rangle_{L^2(\mathbb{R})} = \langle \Pi_0 v_{h,k}, \Pi_0 v_{h,k'}\rangle_{L^2(\mathbb{R}^2)} = o(1)+\delta_{k,k'}.$$

Thus, by item (1),

$$\langle R_0^{\mathrm{new}} v_{h,k}, R_0^{\mathrm{new}} v_{h,k'}\rangle_{L^2(\mathbb{R})} = o(1)+\delta_{k,k'}.$$

With item (2) in hand, we get the conclusion of item (3).

Proof of item (4).

Step 1: We introduce the operator

$$\widetilde{R}_h^{\mathrm{new}} : v\in H^1(\mathbb{R}^2)\mapsto \int_{\mathbb{R}} \phi_{a,h}^{\mathrm{new}}(\tau, i\partial_\sigma) v(\,\cdot\,,\tau)\,d\tau \in L^2(\mathbb{R}), \tag{7-56}$$

where $\phi_{a,h}^{\mathrm{new}}(\tau, i\partial_\sigma)$ is the first-order differential operator

$$\phi_{a,h}^{\mathrm{new}}(\tau, i\partial_\sigma) := \phi_a(\tau) + 2h^{3/8}\mathfrak{R}_a((b_a(\tau)\tau+\zeta_a)\phi_a(\tau))i\partial_\sigma + \kappa h^{1/2}\phi_a^{\mathrm{cor}}(\tau), \tag{7-57}$$

$\kappa = k_{\max}$ and ϕ_a^{cor} is the function introduced in (7-33).

By Hölder's inequality, there exists a constant C_1 such that, for all $v\in H^1(\mathbb{R}^2)$,

$$\|\widetilde{R}_h^{\mathrm{new}} v\|_{L^2(\mathbb{R})} \le C_1(\|v\|_{L^2(\mathbb{R}^2)} + \|\partial_\sigma v\|_{L^2(\mathbb{R}^2)}). \tag{7-58}$$

Thus, by Proposition 7.1 and Remark 7.2,

$$\|\widetilde{R}_h^{\mathrm{new}}\mathcal{P}_h^{\mathrm{new}} v_{h,n} - (h^{-1}\lambda_n(h)-\beta_a)\widetilde{R}_h^{\mathrm{new}} v_{h,n}\|_{L^2(\mathbb{R})} = \mathcal{O}(h^\infty), \tag{7-59}$$

where $\mathcal{P}_h^{\mathrm{new}}$ is the operator in (4-6).

Step 2: We prove the estimate

$$\big\langle \big(c_2(a)\widetilde{R}_h^{\mathrm{new}}\mathcal{P}_h^{\mathrm{new}} - M_3(a)k_{\max}h^{1/2}R_0^{\mathrm{new}} - h^{3/4}H_a^{\mathrm{harm}}R_0^{\mathrm{new}}\big)v_{h,n}, R_0^{\mathrm{new}} v_{h,n}\big\rangle_{L^2(\mathbb{R})} = o(h^{3/4}). \tag{7-60}$$

We first observe that it results from (7-1), (7-10), (7-56), and (7-57),

$$\|\widetilde{R}_h^{\mathrm{new}} v_{h,n} - R_0^- v_{h,n}\|_{L^2(\mathbb{R})} = \mathcal{O}(h^{1/2}). \tag{7-61}$$

For the sake of simplicity, we write $\kappa = k(0) = k_{\max}$. We introduce the functions in $L^2(\mathbb{R})$:

$$f_1 = 2\mathfrak{R}_a((b_a(\tau)\tau+\zeta_a)\phi_a) \tag{7-62}$$

and (see (7-29) and (7-33))

$$f_2 = \phi_a^{\mathrm{cor}} = \mathfrak{R}_a\big(M_3(a)\phi_a - \phi_a' - 2\tau(b_a(\tau)\tau+\zeta_a)^2\phi_a + b_a(\tau)\tau^2(b_a(\tau)\tau+\zeta_a)\phi_a\big). \tag{7-63}$$

Recall the operators P_0, P_1, P_2, P_3, Q_h introduced in (4-10) and (4-11). Noticing the decomposition in (4-9), we write, for any function v with compact support in $\mathbb{R}^2$,

$$\begin{aligned}\widetilde{R}_h^{\mathrm{new}}\mathcal{P}_h^{\mathrm{new}}v = &\int_{\mathbb{R}} \phi_a(\tau)P_0 v(\sigma,\tau)\,d\tau + h^{3/8}\int_{\mathbb{R}}(if_1(\tau)\partial_\sigma P_0 + \phi_a(\tau)P_1)v(\sigma,\tau)\,d\tau\\ &+ h^{1/2}\int_{\mathbb{R}}(\phi_a(\tau)P_2 + \kappa f_2(\tau)P_0)v(\sigma,\tau)\,d\tau\\ &+ h^{3/4}\int_{\mathbb{R}}(\phi_a(\tau)P_3 + if_1(\tau)\partial_\sigma P_1)v(\sigma,\tau)\,d\tau + \mathrm{R}_{h,n}v,\end{aligned} \tag{7-64}$$

where

$$\begin{aligned}\mathrm{R}_{h,n}v = h^{7/8}\widetilde{R}_h^{\mathrm{new}}Q_h v + h^{7/8}&\int_{\mathbb{R}}(if_1(\tau)\partial_\sigma P_2 + \kappa f_2(\tau)P_0)v(\sigma,\tau)\,d\tau + h\kappa\int_{\mathbb{R}} f_2(\tau)P_2 v(\sigma,\tau)\,d\tau\\ &+ h^{5/4}\kappa\int_{\mathbb{R}} f_2(\tau)P_3 v(\sigma,\tau)\,d\tau + h^{9/8}\kappa\int_{\mathbb{R}} if_1(\tau)\partial_\sigma P_3 v(\sigma,\tau)\,d\tau.\end{aligned} \tag{7-65}$$

We now compute the first three terms on the right side of (7-64):

For the first term, since P_0 is self-adjoint in $L^2(\mathbb{R})$, we have

$$\int_{\mathbb{R}} \phi_a(\tau)P_0 v(\sigma,\tau)\,d\tau = \int_{\mathbb{R}} P_0\phi_a(\tau)v(\sigma,\tau)\,d\tau = 0.$$

For the second term, we have

$$\begin{aligned}\int_{\mathbb{R}} if_1(\tau)\partial_\sigma P_0 v(\sigma,\tau)\,d\tau &= \int_{\mathbb{R}} iP_0 f_1(\tau)\partial_\sigma v(\sigma,\tau)\,d\tau\\ &= \int_{\mathbb{R}} 2i\phi_a(\tau)(b_a(\tau)\tau+\zeta_a)\partial_\sigma v(\sigma,\tau)\,d\tau.\end{aligned}$$

Hence we find, by (4-10),

$$\int_{\mathbb{R}}\big(if_1(\tau)\partial_\sigma P_0 + \phi_a(\tau)P_1\big)v(\sigma,\tau)\,d\tau = 0.$$

For the third term, noticing that

$$P_0 f_2 = M_3(a)\phi_a - \phi_a' - 2\tau(b_a(\tau)\tau+\zeta_a)^2\phi_a + b_a(\tau)\tau^2(b_a(\tau)\tau+\zeta_a)\phi_a$$

and

$$\int_{\mathbb{R}} \phi_a(\tau)P_2 v(\sigma,\tau)\,d\tau = \kappa\int_{\mathbb{R}}\big(-\phi_a'(\tau) + 2\tau(b_a(\tau)\tau+\zeta_a)^2\phi_a(\tau) - b_a(\tau)\tau^2(b_a(\tau)\tau+\zeta_a)\phi_a(\tau)\big)v\,d\tau,$$

we get

$$\begin{aligned}(W_2 v)(\sigma) &:= \int_{\mathbb{R}}\big(\phi_a(\tau)P_2 + \kappa f_2(\tau)P_0\big)v(\sigma,\tau)\,d\tau\\ &= \int_{\mathbb{R}}\big(\phi_a(\tau)P_2 + \kappa(P_0 f_2(\tau))\big)v(\sigma,\tau)\,d\tau\\ &= \kappa\int_{\mathbb{R}}\big(M_3(a)\phi_a(\tau) - 2\phi_a'(\tau)\big)v(\sigma,\tau)\,d\tau.\end{aligned}$$

By the forgoing computations, (7-64) becomes

$$\widetilde{R}_h^{\text{new}}\mathcal{P}_h^{\text{new}}v = h^{1/2}W_2v + h^{3/4}W_3v + \mathrm{R}_{h,n}v, \tag{7-66}$$

with

$$(W_3v)(\sigma) := \int_{\mathbb{R}} \big(\phi_a(\tau)P_3 + if_1(\tau)\partial_\sigma P_1\big)v(\sigma,\tau)\,d\tau. \tag{7-67}$$

We estimate $W_2v_{h,n}$ by writing $v_{h,n} = \Pi_0 v_{h,n} + (v_{h,n} - \Pi_0 v_{h,n})$, with Π_0 the projection introduced in (7-4), and by using (7-41). Eventually, since $P_0\Pi_0 = 0$ and $\langle \phi_a, \phi_a'\rangle_{L^2(\mathbb{R})} = 0$, we get by Remark 2.3,

$$\|W_2v_{h,n} - M_3(a)\kappa R_0^- v_{h,n}\|_{L^2(\mathbb{R})} = o(h^{1/4}). \tag{7-68}$$

We still have to estimate the terms involving W_3 and $\mathrm{R}_{h,n}$ in (7-66) when $v = v_{h,n}$. By choosing η small enough, the error term

$$\mathrm{r}_n(\sigma,h) := \mathrm{R}_{h,n}v_{h,n}, \tag{7-69}$$

with $\mathrm{R}_{h,n}$ introduced in (7-65), satisfies

$$\langle \mathrm{r}_n(\,\cdot\,,h), R_0^{\text{new}}v_{h,n}\rangle_{L^2(\mathbb{R})} = o(h^{3/4}). \tag{7-70}$$

The technical proof of (7-70) is given in Appendix B. So we are left (see (7-67)) with estimating

$$W_3v_{h,n} = w_1 + w_2, \tag{7-71}$$

where

$$w_1(\sigma) := \int_{\mathbb{R}} \phi_a(\tau)P_3v_{h,n}(\sigma,\tau)\,d\tau,$$
$$w_2(\sigma) := \int_{\mathbb{R}} if_1(\tau)\partial_\sigma P_1 v_{h,n}(\sigma,\tau)\,d\tau.$$

In light of the definition of P_3 in (4-10) and R_0^- in (7-1), we write

$$w_1(\sigma) = -\partial_\sigma^2 R_0^- v_{h,n}(\sigma) + \frac{k''(0)\sigma^2}{2}w(\sigma),$$

where

$$w(\sigma) = \int_{\mathbb{R}} \big(\partial_\tau + 2\tau(b_a(\tau)\tau+\zeta_a)^2 - b_a(\tau)\tau(b_a(\tau)\tau+\zeta_a)\big)\phi_a(\tau)\,v_{h,n}(\sigma,\tau)\,d\tau.$$

Using Proposition 7.1 and that $v_{h,n}$ is supported in $\{|\sigma| \le h^{-\eta}\}$, we get

$$\|\sigma^2(w - M_3(a)R_0^- v_{h,n})\|_{L^2(\mathbb{R})} = \mathcal{O}(h^{1/4-2\eta}).$$

Hence

$$\left\| w_1 - \left(-\partial_\sigma^2 + \frac{k''(0)M_3(a)}{2}\sigma^2\right)R_0^- v_{h,n}\right\|_{L^2(\mathbb{R})} = \mathcal{O}(h^{1/4-2\eta}). \tag{7-72}$$

Furthermore, by (4-10) and (7-62), the term w_2 can be expressed as

$$\begin{aligned} w_2(\sigma) &= 2\partial_\sigma^2 \int_{\mathbb{R}} f_1(\tau)(\zeta_a + b_a(\tau)\tau)v_{h,n}(\sigma,\tau)\,d\tau \\ &= 4\partial_\sigma^2 \int_{\mathbb{R}} (b_a(\tau)\tau+\zeta_a)\mathfrak{R}_a\big((b_a(\tau)\tau+\zeta_a)\phi_a(\tau)\big)v_{h,n}(\sigma,\tau)\,d\tau. \end{aligned} \tag{7-73}$$

Collecting (7-72) and (7-73), along with the definition of R_0^{new} in (7-53), we infer from (7-71)

$$\left\| W_3 v_{h,n} - \left(-\partial_\sigma^2 R_0^{\text{new}} + \frac{k''(0) M_3(a)}{2} \sigma^2 R_0^- \right) v_{h,n} \right\|_{L^2(\mathbb{R})} = \mathcal{O}(h^{1/4-2\eta}). \tag{7-74}$$

By Hölder's inequality, we infer from (7-68) and (7-74)

$$\begin{aligned} &h^{1/2} \langle (W_2 - M_3(a)\kappa R_0^-) v_{h,n}, R_0^{\text{new}} v_{h,n} \rangle_{L^2(\mathbb{R})} \\ &\quad + h^{3/4} \left\langle W_3 v_{h,n} - \left(-\partial_\sigma^2 R_0^{\text{new}} + \frac{k''(0) M_3(a)}{2} \sigma^2 R_0^- \right) v_{h,n}, R_0^{\text{new}} v_{h,n} \right\rangle_{L^2(\mathbb{R})} = o(h^{3/4}) \| R_0^{\text{new}} v_{h,n} \|_{L^2(\mathbb{R})}. \end{aligned}$$

By (7-66) and (7-70), we get from the above estimate

$$\left\langle \left(\widetilde{R}_h^{\text{new}} \mathcal{P}_h^{\text{new}} - h^{1/2} M_3(a) \kappa R_0^- - h^{3/4} \widetilde{H} \right) v_{h,n}, R_0^{\text{new}} v_{h,n} \right\rangle_{L^2(\mathbb{R})} = o(h^{3/4}) \| R_0^{\text{new}} v_{h,n} \|_{L^2(\mathbb{R})},$$

where

$$\widetilde{H} := -\partial_\sigma^2 R_0^{\text{new}} + \frac{k''(0) M_3(a)}{2} \sigma^2 R_0^-.$$

Finally, by item (1) and Proposition 2.5, we get (7-60).

Step 3: Using Steps 1 and 2, we are now able to finish the proof of item (4). By (1-12) and (2-20), $c_2(a) = 1 - 4I_2(a)$; hence (7-61) and item (1) yield that

$$\| c_2(a) \widetilde{R}_h^{\text{new}} v_{h,n} - R_0^{\text{new}} v_{h,n} \|_{L^2(\mathbb{R})} = \mathcal{O}(h^{1/4}). \tag{7-75}$$

Collecting (7-59), (7-60) and (7-75), we get

$$\left\langle h^{3/4} H_a^{\text{harm}} R_0^{\text{new}} v_{h,n} - \Lambda_n(h) R_0^{\text{new}} v_{h,n}, R_0^{\text{new}} v_{h,n} \right\rangle_{L^2(\mathbb{R})} = \mathcal{O}(|\Lambda_n(h)| h^{1/4}) + o(h^{3/4}),$$

where, by (6-4) and Theorem 4.1,

$$|\Lambda_n(h)| = |h^{-1} \lambda_n(h) - \beta_a - M_3(a) k_{\max} h^{1/2}| = o(h^{1/2}).$$

Thus, we obtain

$$\left\langle h^{3/4} H_a^{\text{harm}} R_0^{\text{new}} v_{h,n} - \Lambda_n(h) R_0^{\text{new}} v_{h,n}, R_0^{\text{new}} v_{h,n} \right\rangle_{L^2(\mathbb{R})} = o(h^{3/4}).$$

Dividing by $h^{3/4}$ and using item (2), we get item (4). □

With Proposition 7.5 in hand, we can now finish the proof of Theorem 1.2.

Proof of Theorem 1.2. The upper bound of $\lambda_n(h)$ follows from Theorem 4.1. For the lower bound of $\lambda_n(h)$, consider $u = \sum_{k=1}^n a_k R_0^{\text{new}} v_{h,k}$ such that $\|u\|_{L^2(\mathbb{R})} = 1$, where R_0^{new} is introduced in (7-53). From Proposition 7.5 we have

$$((1 - 4I_2(a))^2 + o(1)) \sum_{k=1}^n |a_k|^2 = 1$$

and

$$\langle (H_a^{\text{harm}} - h^{-3/4} \Lambda_n(h)) u, u \rangle_{L^2(\mathbb{R})} \le o(1) \sum_{k=1}^n |a_k|^2.$$

Consequently,

$$\max_{u\in M_n,\,\|u\|=1}\langle (H_a^{\mathrm{harm}}-h^{-3/4}\Lambda_n(h))u,u\rangle_{L^2(\mathbb{R})}=o(1),$$

where M_n is the space defined in (7-55). By the min-max principle

$$\sqrt{\frac{M_3(a)k''(0)c_2(a)}{2}}(2n-1)\le h^{-3/4}\Lambda_n(h)+o(1),$$

thereby leading to

$$\lambda_n(h)\ge \beta_a h+M_3(a)k_{\max}h^{3/2}+\sqrt{\frac{M_3(a)k''(0)c_2(a)}{2}}(2n-1)h^{7/4}+o(h^{7/4}). \qquad \square$$

Appendix A: Frenet coordinates near the magnetic edge

We introduce the Frenet coordinates near Γ. We refer the reader to [Fournais and Helffer 2010, Appendix F] and [Assaad et al. 2019] for a similar setup.

Let $s\mapsto M(s)\in\Gamma$ be the arc length parametrization of Γ such that

- $\nu(s)$ is the unit normal of Γ at the point $M(s)$ pointing towards P_1,
- $T(s)$ is the unit tangent vector of Γ at the point $M(s)$, such that $(T(s),\nu(s))$ is a direct frame, i.e., $\det(T(s),\nu(s))=1$.

We define the curvature k of Γ as $T'(s)=k(s)\nu(s)$. *Working under Assumption 1.1*, we assume without loss of generality that $s_0=0$, where s_0 is the unique maximum of the curvature at Γ ($k(0)=k_{\max}$).

For $t_0>0$, we define the transformation $\Phi=\Phi_{t_0}$ as

$$\Phi:\mathbb{R}\times(-t_0,t_0)\to\Gamma_{t_0}:=\{x\in\mathbb{R}^2:\operatorname{dist}(x,\Gamma)<t_0\},\qquad (s,t)\mapsto M(s)+t\nu(s). \tag{A-1}$$

We pick t_0 sufficiently small so that Φ is a diffeomorphism, whose Jacobian is

$$\mathfrak{a}(s,t):=J_\Phi(s,t)=1-t\,k(s). \tag{A-2}$$

We consider the following correspondence between functions u in $H^1_{\mathrm{loc}}(\Gamma_{t_0})$ and those $\tilde u$ in $H^1_{\mathrm{loc}}(\mathbb{R}\times(-t_0,t_0))$:

$$\tilde u(s,t)=u(\Phi(s,t)), \tag{A-3}$$

and vice versa.

Moreover, we assign to the potential $\boldsymbol{F}$ in (1-1) a vector field $\widetilde{\boldsymbol{F}}\in H^1_{\mathrm{loc}}(\mathbb{R}\times(-t_0,t_0))$ as

$$\boldsymbol{F}(x)=(F_1(x),F_2(x))\mapsto\widetilde{\boldsymbol{F}}(s,t)=(\widetilde F_1(s,t),\widetilde F_2(s,t)),$$

where

$$\widetilde F_1(s,t)=\mathfrak{a}(s,t)\boldsymbol{F}(\Phi(s,t))\cdot T(s)\quad\text{and}\quad \widetilde F_2(s,t)=\boldsymbol{F}(\Phi(s,t))\cdot\nu(s). \tag{A-4}$$

Consequently,

$$(h\nabla-i\boldsymbol{F})^2=\mathfrak{a}^{-1}(h\partial_s-i\widetilde F_1)\mathfrak{a}^{-1}(h\partial_s-i\widetilde F_1)+\mathfrak{a}^{-1}(h\partial_t-i\widetilde F_2)\mathfrak{a}(h\partial_t-i\widetilde F_2). \tag{A-5}$$

Note that

$$\operatorname{curl}\widetilde{\boldsymbol{F}}(s,t)=(1-tk(s))\operatorname{curl}\boldsymbol{F}(\Phi(s,t))=(1-tk(s))(\mathbb{1}_{\{t>0\}}+a\mathbb{1}_{\{t<0\}}), \tag{A-6}$$

where $\operatorname{curl}\widetilde{\boldsymbol{F}}=\partial_s\widetilde F_2-\partial_t\widetilde F_1$ and $\operatorname{curl}\boldsymbol{F}=\partial_{x_1}F_2-\partial_{x_2}F_1$ is as in (1-2).

Furthermore, we present the change of variable formulas (for functions compactly supported in Γ_{t_0}):

$$\begin{aligned}\int_{\Gamma_{t_0}} |u|^2\,dx &= \int_{\mathbb{R}}\int_{-t_0}^{t_0} |\tilde{u}|^2\,\mathfrak{a}\,dt\,ds,\\ \int_{\Gamma_{t_0}} |(h\nabla - i\boldsymbol{F})u|^2\,dx &= \int_{\mathbb{R}}\int_{-t_0}^{t_0} \big(\mathfrak{a}^{-2}|(h\partial_s - i\widetilde{F}_1)\tilde{u}|^2 + |(h\partial_t - i\widetilde{F}_2)\tilde{u}|^2\big)\mathfrak{a}\,dt\,ds.\end{aligned} \tag{A-7}$$

Now, we make a global change of gauge ω as follows:

Lemma A.1. *There exists a function $\omega \in H^2(\Phi^{-1}(\Gamma_{t_0}\cap\Omega))$ such that*

$$\widetilde{\boldsymbol{F}} - \nabla_{s,t}\omega = \begin{pmatrix} -b_a(t)\big(t - \frac{1}{2}t^2 k(s)\big) \\ 0 \end{pmatrix} \quad \text{in } \Phi^{-1}(\Gamma_{t_0}\cap\Omega),$$

where $t \mapsto b_a(t)$ is defined by $b_a(t) = \mathbb{1}_{\{t>0\}} + a\,\mathbb{1}_{\{t<0\}}$.

Proof. For $(s,t)\in\Phi^{-1}(\Gamma_{t_0}\cap\Omega)$, let $\omega(s,t) = \int_0^t \widetilde{F}_2(s,t')\,dt' + \int_0^s \widetilde{F}_1(s',0)\,ds'$. This choice of ω and (A-6) establish the lemma. □

The gauge of Lemma A.1 is adequate when working with functions localized near the edge Γ. With this choice of gauge, we have the following identity which is useful to analyze the decay of functions localized near Γ.

Lemma A.2. *Assume that $\varphi \in H^2(\Omega)$ with compact support in $\Omega\cap\Gamma_{t_0}$. Let g and G be the functions defined (by means of (A-3)) as*

$$\tilde{g}(s,t) = (h^{1/2}\partial_s - i\zeta_a)\tilde{\varphi}(s,t) \quad \text{and} \quad \widetilde{G}(s,t) = -(h^{1/2}\partial_s - i\zeta_a)(e^{2\tilde{\phi}}\tilde{g}),$$

where ζ_a is the constant in Section 2B and ϕ is a Lipschitz real-valued function on Ω. If $g \in H^2(\Omega)$, then

$$\operatorname{Re}\langle \mathcal{P}_h\varphi, G\rangle_{L^2(\Omega)} = \mathcal{Q}_h(e^{\phi}g) - h^2\||\nabla g|e^{\phi}\varphi\|^2_{L^2(\Omega)} - h^{1/2}\operatorname{Re}(\mathsf{T}_h).$$

Here $\mathcal{Q}_h$ is the quadratic form introduced in (4-11) and

$$\mathsf{T}_h = \big\langle (h\partial_s - i\widetilde{F}_1)\big((\partial_s\mathfrak{a}^{-1} - i\mathfrak{a}^{-1}\partial_s\widetilde{F}_1)(h\partial_s - i\widetilde{F}_1)\tilde{\varphi} - i\mathfrak{a}^{-1}(\partial_s\widetilde{F}_1)\tilde{\varphi}\big) + h^2\partial_t(\partial_s\mathfrak{a})\partial_t\tilde{\varphi}, e^{2\tilde{\phi}}\tilde{g}\big\rangle_{L^2(\mathbb{R})}.$$

Proof. We assume that $\widetilde{F}_2 = 0$ and get from (A-5) and (A-2)

$$\langle \mathcal{P}_h\varphi, G\rangle_{L^2(\Omega)} = \langle (h\partial_s - i\widetilde{F}_1)\mathfrak{a}^{-1}(h\partial_s - i\widetilde{F}_1)\varphi + h^2\partial_t\mathfrak{a}\partial_t\varphi, (h^{1/2}\partial_s - i\zeta_a)(e^{2\phi}g)\rangle_{L^2(\mathbb{R}^2)}, \tag{A-8}$$

where we dropped the tildes from the notation for the sake of simplicity. Notice that

$$\begin{aligned}(h^{1/2}\partial_s - i\zeta_a)\partial_t\mathfrak{a}\partial_t\varphi &= \partial_t((h^{1/2}\partial_s - i\zeta_a)\mathfrak{a}\partial_t\varphi)\\ &= \partial_t(\mathfrak{a}\partial_t(h^{1/2}\partial_s - i\zeta_a)\varphi) + h^{1/2}\partial_t(\partial_s\mathfrak{a})\partial_t\varphi\\ &= \partial_t\mathfrak{a}\partial_t g + h^{1/2}\partial_t(\partial_s\mathfrak{a})\partial_t\varphi,\end{aligned}$$

and

$$
\begin{aligned}
&(h^{1/2}\partial_s - i\zeta_a)(h\partial_s - i\widetilde{F}_1)\mathfrak{a}^{-1}(h\partial_s - i\widetilde{F}_1)\varphi \\
&\quad = (h\partial_s - i\widetilde{F}_1)\big((h^{1/2}\partial_s - i\zeta_a) - ih^{1/2}(\partial_s\widetilde{F}_1)\big)\mathfrak{a}^{-1}(h\partial_s - i\widetilde{F}_1)\varphi \\
&\quad = (h\partial_s - i\widetilde{F}_1)\big(\mathfrak{a}^{-1}(h\partial_s - i\widetilde{F}_1)(h^{1/2}\partial_s - i\zeta_a)\varphi - ih^{1/2}(\partial_s\widetilde{F}_1)\mathfrak{a}^{-1}(h\partial_s - i\widetilde{F}_1)\varphi\big) \\
&\qquad\qquad + h^{1/2}(h\partial_s - i\widetilde{F}_1)\big((\partial_s\mathfrak{a}^{-1})(h\partial_s - i\widetilde{F}_1)\varphi - i\mathfrak{a}^{-1}(\partial_s\widetilde{F}_1)\varphi\big) \\
&\quad = (h\partial_s - i\widetilde{F}_1)\mathfrak{a}^{-1}(h\partial_s - i\widetilde{F}_1)g + h^{1/2}(h\partial_s - i\widetilde{F}_1)\big((\partial_s\mathfrak{a}^{-1} - i\mathfrak{a}^{-1}\partial_s\widetilde{F}_1)(h\partial_s - i\widetilde{F}_1)\varphi - i\mathfrak{a}^{-1}(\partial_s\widetilde{F}_1)\varphi\big).
\end{aligned}
$$

By integration by parts, we infer from (A-8)

$$
\langle \mathcal{P}_h\varphi, G\rangle_{L^2(\Omega)} = \langle \mathcal{P}_h g, e^{2\phi}g\rangle_{L^2(\Omega)} - h^{1/2}\mathsf{T}_h. \tag{A-9}
$$

Finally, by integration by parts, we get

$$
\mathrm{Re}\langle \mathcal{P}_h g, e^{2\phi}g\rangle_{L^2(\Omega)} = \mathcal{Q}_h(e^{\phi}g) - h^2\||\nabla\phi|e^{\phi}g\|^2_{L^2(\Omega)}. \qquad \square
$$

Appendix B: Control of a remainder term

The aim of this appendix is to prove the estimate in (7-70). We fix a positive integer $n \geq 1$ and two positive constants $\eta \in \big(0, \frac{1}{8}\big)$ and $\delta \in \big(0, \frac{1}{12}\big)$.

For all $h > 0$, let $v_{h,n}$ be the function introduced in (7-6) which is supported in $\{|\sigma| < h^{-\eta},\ |\tau| < h^{-\delta}\}$. Moreover, by (7-6) and Propositions 6.2 and 6.3, we observe that,

$$
\text{for all } \theta \in \big(0, \tfrac{3}{8}\big), \text{ there exists } C_\theta > 0 \text{ such that } \|\partial_\sigma^j v_{h,n}\|_{L^2(\mathbb{R}^2)} \leq C_\theta h^{-j\theta} \quad (0 \leq j \leq 2). \tag{B-1}
$$

Consider two functions $f \in L^2(\mathbb{R})$ and $p \in L^1_{\mathrm{loc}}(\mathbb{R}^2)$ so that,

$$
\text{for all } \alpha \geq 1, \quad \tau^\alpha f(\tau) \in L^2(\mathbb{R}),
$$

and there exist $k \geq 1$ and C such that

$$
|p(\sigma,\tau)| \leq C(|\sigma|^k + |\tau|^k + 1) \quad (\sigma, \tau \in \mathbb{R}).
$$

For $j \in \{0, 1, 2\}$, we introduce the function

$$
w_j(\sigma) = \int_{\mathbb{R}} f(\tau)p(\sigma,\tau)\partial_\sigma^j v_{h,n}(\sigma,\tau)\,d\tau, \tag{B-2}
$$

whose support is included in $\{|\sigma| < h^{-\eta}\}$, by the considerations on the support of $v_{h,n}$.

Lemma B.1. *Given $\eta \in \big(0, \frac{1}{8}\big)$, there exist two positive constants $h_0, C > 0$ such that*

$$
\|w_j\|_{L^2(\mathbb{R})} \leq C\, h^{-(k+j/2)\eta}
$$

for all $h \in (0, h_0)$ and $j \in \{0, 1, 2\}$.

Proof. By Hölder's inequality

$$
|w_j(\sigma)|^2 \leq \left(\int_{\mathbb{R}} |f(\tau)|^2|p(\sigma,\tau)|^2\,d\tau\right)\left(\int_{\mathbb{R}} |\partial_\sigma^j v_{h,n}(\sigma,\tau)|^2\,d\tau\right). \tag{B-3}
$$

For σ in the support of w_j, we have

$$\int_{\mathbb{R}} |f(\tau)|^2 |p(\sigma,\tau)|^2 \, d\tau \le C \int_{\mathbb{R}} |f(\tau)|^2 (1+|\tau|^k+|\sigma|^k)^2 \, d\tau \le \widetilde{C}_k (1+h^{-2k\eta}).$$

Inserting this into (B-3) then integrating with respect to σ, we get

$$\int_{\mathbb{R}} |w_j(\sigma)|^2 \, d\sigma \le \widetilde{C}_k (1+h^{-2k\eta}) \int_{\mathbb{R}^2} |\partial_\sigma^j v_{h,n}(\sigma,\tau)|^2 \, d\sigma \, d\tau.$$

Finally, we use (B-1) with $\theta = \eta$. □

We will encounter functions of the form

$$\mathfrak{w}_j(\sigma) = \int_{\mathbb{R}} g(\tau) q(\sigma) \partial_\tau^j v_{h,n}(\sigma,\tau) \, d\tau \quad (j \in \{1,2\},\ \sigma \in \mathbb{R}), \tag{B-4}$$

where $g \in H^j(\mathbb{R})$ and $q \in H^1_{\mathrm{loc}}(\mathbb{R})$ satisfy,

$$\text{for all } \alpha \ge 1, \quad \tau^\alpha g^{(i)}(\tau) \in L^2(\mathbb{R}) \ \ (1 \le i \le j),$$

and

$$\text{there exists } k \ge 1 \text{ such that there exists } C_k > 0 \text{ such that } \ |q(\sigma)| \le C_k (1+|\sigma|^k) \ \ (\sigma \in \mathbb{R}).$$

Lemma B.2. *Given $\eta \in \left(0, \frac{1}{8}\right)$, there exist two positive constants h_0 and C such that*

$$\|\mathfrak{w}_j\|_{L^2(\mathbb{R})} \le C h^{-(k+1)\eta}$$

for all $h \in (0, h_0]$ and $j \in \{1,2\}$.

Proof. Using integration by parts and that $v_{h,n}$ is with compact support, we get

$$\mathfrak{w}_j(\sigma) = (-1)^j \int_{\mathbb{R}} g^{(j)}(\tau) q(\sigma) v_{h,n}(\sigma,\tau) \, d\tau.$$

This function has the form of functions in Lemma B.1, with $f(\tau) = g^{(j)}(\tau)$ and $p(\sigma,\tau) = q(\sigma)$. □

The inner product of the remainder, $\mathrm{r}_n(\sigma,h)$ in (7-69), and the function, $R_0^{\mathrm{new}} v_{h,n}$ in (7-53), can be expressed as the inner product of a linear combination of functions having the forms in Lemmas B.1 and B.2. The polynomials we encounter are of degree 6 at most. More precisely,

$$\langle \mathrm{r}_n(\,\cdot\,,h), R_0^{\mathrm{new}} v_{h,n} \rangle_{L^2(\mathbb{R})} = h^{7/8} A_1 + h^{7/8} A_2 + h A_3 + h^{9/8} A_4 + h^{5/4} A_5,$$

where

$$A_1 = \langle a_{1,1}, b_1 \rangle_{L^2(\mathbb{R})} + h^{3/8} \langle a_{1,2}, b_2 \rangle_{L^2(\mathbb{R})} + h^{1/2} \langle a_{1,3}, b_1 \rangle_{L^2(\mathbb{R})},$$
$$A_2 = \langle a_{2,1}, b_2 \rangle_{L^2(\mathbb{R})} + \langle a_{2,2}, b_1 \rangle_{L^2(\mathbb{R})},$$
$$A_3 = \langle a_3, b_1 \rangle_{L^2(\mathbb{R})}, \quad A_4 = \langle a_4, b_2 \rangle_{L^2(\mathbb{R})}, \quad A_5 = \langle a_5, b_1 \rangle_{L^2(\mathbb{R})},$$

and

$$a_{1,1} = \int_{\mathbb{R}} g_1(\tau) Q_h v_{h,n} \, d\tau, \quad a_{1,2} = \int_{\mathbb{R}} g_2(\tau) Q_h v_{h,n} \, d\tau, \quad a_{1,3} = \int_{\mathbb{R}} g_3(\tau) Q_h v_{h,n} \, d\tau,$$
$$a_{2,1} = \int_{\mathbb{R}} f_1(\tau) P_2 v_{h,n} d\tau, \quad a_{2,2} = \kappa \int_{\mathbb{R}} f_2(\tau) P_0 v_{h,n} \, d\tau,$$

$$a_3 = \kappa \int_{\mathbb{R}} f_2(\tau) P_2 v_{h,n}\, d\tau, \quad a_4 = \kappa \int_{\mathbb{R}} f_1(\tau) P_3 v_{h,n}\, d\tau, \quad a_5 = \kappa \int_{\mathbb{R}} f_2(\tau) P_3 v_{h,n}\, d\tau,$$
$$b_1 = \int_{\mathbb{R}} g(\tau) v_{h,n}\, d\tau, \quad b_2 = i \int_{\mathbb{R}} g(\tau) \partial_\sigma v_{h,n}\, d\tau.$$

Here, Q_h is the operator introduced in (4-11), P_0, P_1, P_2, P_3 are the operators introduced in (4-10), f_1, f_2 are the functions introduced in (7-62)-(7-63), the functions g_1, g_2, g_3 and g are defined as follows (see (7-57) and (7-53))

$$g_1 = \phi_a, \quad g_2 = f_1 = 2\mathfrak{R}_a((b_a(\tau)\tau + \zeta_a)\phi_a),$$
$$g_3 = \kappa f_2 = \kappa \mathfrak{R}_a\big(M_3(a)\phi_a - \phi_a' - 2\tau(b_a(\tau)\tau + \zeta_a)^2\phi_a + b_a(\tau)\tau^2(b_a(\tau)\tau + \zeta_a)\phi_a\big),$$
$$g = \phi_a - 4(b_a(\tau)\tau + \zeta_a)\mathfrak{R}_a((b_a(\tau)\tau + \zeta_a)\phi_a).$$

So, we get

$$\langle \mathrm{r}_n(\,\cdot\,, h), R_0^{\mathrm{new}} v_{h,n}\rangle_{L^2(\mathbb{R})} = \mathcal{O}(h^{7/8-8\eta}).$$

By choosing $\eta < \frac{1}{64}$, we get (7-70).

Acknowledgments

Kachmar is partially supported by the Center for Advanced Mathematics Sciences (CAMS, American University of Beirut).

References

[Assaad 2021] W. Assaad, "The breakdown of superconductivity in the presence of magnetic steps", *Commun. Contemp. Math.* **23**:2 (2021), art. id. 2050005. MR Zbl

[Assaad and Kachmar 2022] W. Assaad and A. Kachmar, "Lowest energy band function for magnetic steps", *J. Spectr. Theory* **12**:2 (2022), 813–833. MR Zbl

[Assaad et al. 2019] W. Assaad, A. Kachmar, and M. Persson-Sundqvist, "The distribution of superconductivity near a magnetic barrier", *Comm. Math. Phys.* **366**:1 (2019), 269–332. MR Zbl

[Bernoff and Sternberg 1998] A. Bernoff and P. Sternberg, "Onset of superconductivity in decreasing fields for general domains", *J. Math. Phys.* **39**:3 (1998), 1272–1284. MR Zbl

[Bolley and Helffer 1993] C. Bolley and B. Helffer, "An application of semi-classical analysis to the asymptotic study of the supercooling field of a superconducting material", *Ann. Inst. H. Poincaré Phys. Théor.* **58**:2 (1993), 189–233. MR Zbl

[Bonnaillie-Noël 2003] V. Bonnaillie, *Analyse mathématique de la supraconductivité dans un domaine à coins: méthodes semi-classiques et numériques*, Ph.D. thesis, Université Paris Sud-Paris XI, 2003, available at https://theses.hal.science/tel-00005430.

[Bonnaillie-Noël 2005] V. Bonnaillie, "On the fundamental state energy for a Schrödinger operator with magnetic field in domains with corners", *Asymptot. Anal.* **41**:3-4 (2005), 215–258. MR Zbl

[Bonnaillie-Noël and Dauge 2006] V. Bonnaillie-Noël and M. Dauge, "Asymptotics for the low-lying eigenstates of the Schrödinger operator with magnetic field near corners", *Ann. Henri Poincaré* **7**:5 (2006), 899–931. MR Zbl

[Bonnaillie-Noël and Fournais 2007] V. Bonnaillie-Noël and S. Fournais, "Superconductivity in domains with corners", *Rev. Math. Phys.* **19**:6 (2007), 607–637. MR Zbl

[Bonnaillie-Noël et al. 2007] V. Bonnaillie-Noël, M. Dauge, D. Martin, and G. Vial, "Computations of the first eigenpairs for the Schrödinger operator with magnetic field", *Comput. Methods Appl. Mech. Engrg.* **196**:37-40 (2007), 3841–3858. MR Zbl

[Bonnaillie-Noël et al. 2022] V. Bonnaillie-Noël, F. Hérau, and N. Raymond, "Purely magnetic tunneling effect in two dimensions", *Invent. Math.* **227**:2 (2022), 745–793. MR Zbl

[Dauge and Helffer 1993] M. Dauge and B. Helffer, "Eigenvalues variation, I: Neumann problem for Sturm–Liouville operators", *J. Differential Equations* **104**:2 (1993), 243–262. MR Zbl

[Fournais and Persson-Sundqvist 2015] S. Fournais and M. Persson Sundqvist, "Lack of diamagnetism and the Little–Parks effect", *Comm. Math. Phys.* **337**:1 (2015), 191–224. MR Zbl

[Fournais and Helffer 2006] S. Fournais and B. Helffer, "Accurate eigenvalue asymptotics for the magnetic Neumann Laplacian", *Ann. Inst. Fourier* (*Grenoble*) **56**:1 (2006), 1–67. MR Zbl

[Fournais and Helffer 2010] S. Fournais and B. Helffer, *Spectral methods in surface superconductivity*, Progr. Nonlinear Differential Eq. Appl. **77**, Birkhäuser, Boston, 2010. MR Zbl

[Fournais et al. 2022] S. Fournais, B. Helffer, and A. Kachmar, "Tunneling effect induced by a curved magnetic edge", pp. 315–350 in *The physics and mathematics of Elliott Lieb*: *the* 90*th anniversary*, *I*, edited by R. L. Frank et al., Eur. Math. Soc., Berlin, 2022. MR Zbl

[Helffer and Kachmar 2017] B. Helffer and A. Kachmar, "Eigenvalues for the Robin Laplacian in domains with variable curvature", *Trans. Amer. Math. Soc.* **369**:5 (2017), 3253–3287. MR Zbl

[Helffer and Morame 2001] B. Helffer and A. Morame, "Magnetic bottles in connection with superconductivity", *J. Funct. Anal.* **185**:2 (2001), 604–680. MR Zbl

[Helffer and Morame 2004] B. Helffer and A. Morame, "Magnetic bottles for the Neumann problem: curvature effects in the case of dimension 3 (general case)", *Ann. Sci. École Norm. Sup.* (4) **37**:1 (2004), 105–170. MR Zbl

[Helffer and Pan 2003] B. Helffer and X.-B. Pan, "Upper critical field and location of surface nucleation of superconductivity", *Ann. Inst. H. Poincaré C Anal. Non Linéaire* **20**:1 (2003), 145–181. MR Zbl

[Hislop and Soccorsi 2015] P. D. Hislop and E. Soccorsi, "Edge states induced by Iwatsuka Hamiltonians with positive magnetic fields", *J. Math. Anal. Appl.* **422**:1 (2015), 594–624. MR Zbl

[Hislop et al. 2016] P. D. Hislop, N. Popoff, N. Raymond, and M. P. Sundqvist, "Band functions in the presence of magnetic steps", *Math. Models Methods Appl. Sci.* **26**:1 (2016), 161–184. MR Zbl

[Iwatsuka 1985] A. Iwatsuka, "Examples of absolutely continuous Schrödinger operators in magnetic fields", *Publ. Res. Inst. Math. Sci.* **21**:2 (1985), 385–401. MR Zbl

[Lu and Pan 1999a] K. Lu and X.-B. Pan, "Eigenvalue problems of Ginzburg–Landau operator in bounded domains", *J. Math. Phys.* **40**:6 (1999), 2647–2670. MR Zbl

[Lu and Pan 1999b] K. Lu and X.-B. Pan, "Estimates of the upper critical field for the Ginzburg–Landau equations of superconductivity", *Phys. D* **127**:1-2 (1999), 73–104. MR Zbl

[Lu and Pan 2000] K. Lu and X.-B. Pan, "Gauge invariant eigenvalue problems in $\mathbb{R}^2$ and in $\mathbb{R}^2_+$", *Trans. Amer. Math. Soc.* **352**:3 (2000), 1247–1276. MR Zbl

[Peeters and Matulis 1993] F. M. Peeters and A. Matulis, "Quantum structures created by nonhomogeneous magnetic fields", *Phys. Rev. B* **48**:20 (1993), art. id. 15166.

[Reijniers and Peeters 2000] J. Reijniers and F. M. Peeters, "Snake orbits and related magnetic edge states", *J. Phys. Condens. Matter* **12**:47 (2000), art. id. 9771.

[Saint-James and de Gennes 1963] D. Saint-James and P. G. Gennes, "Onset of superconductivity in decreasing fields", *Phys. Lett.* **7**:5 (1963), 306–308.

[Tilley and Tilley 1990] D. R. Tilley and J. Tilley, *Superfluidity and superconductivity*, Routledge, New York, 1990.

Received 6 Sep 2021. Revised 25 Mar 2022. Accepted 11 Jul 2022.

WAFAA ASSAAD: wafaa.assaad@liu.edu.lb
Faculty of Arts and Sciences, Lebanese International University, Beirut, Lebanon

BERNARD HELFFER: bernard.helffer@univ-nantes.fr
Laboratoire de Mathématiques Jean Leray, Université de Nantes, Nantes, France

AYMAN KACHMAR: akachmar@cuhk.edu.cn
School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Shenzhen, China

ANALYSIS AND PDE
Vol. 17 (2024), No. 2, pp. 587–616
DOI: 10.2140/apde.2024.17.587

NECESSARY DENSITY CONDITIONS FOR SAMPLING AND INTERPOLATION IN SPECTRAL SUBSPACES OF ELLIPTIC DIFFERENTIAL OPERATORS

KARLHEINZ GRÖCHENIG AND ANDREAS KLOTZ

We prove necessary density conditions for sampling in spectral subspaces of a second-order uniformly elliptic differential operator on $\mathbb{R}^d$ with slowly oscillating symbol. For constant-coefficient operators, these are precisely Landau's necessary density conditions for bandlimited functions, but for more general elliptic differential operators it has been unknown whether such a critical density even exists. Our results prove the existence of a suitable critical sampling density and compute it in terms of the geometry defined by the elliptic operator. In dimension $d = 1$, functions in a spectral subspace can be interpreted as functions with variable bandwidth, and we obtain a new critical density for variable bandwidth. The methods are a combination of the spectral theory and the regularity theory of elliptic partial differential operators, some elements of limit operators, certain compactifications of $\mathbb{R}^d$, and the theory of reproducing kernel Hilbert spaces.

1. Introduction

The classical Paley–Wiener space is the subspace $\mathrm{PW}_\Omega = \{f \in L^2(\mathbb{R}) : \operatorname{supp} \hat{f} \subseteq [-\Omega, \Omega]\}$ of $L^2(\mathbb{R})$. Using Fourier inversion, one sees that the point evaluation $f \mapsto f(x)$ is bounded on PW_Ω. The fundamental questions about PW_Ω are originally motivated by problems in signal processing and information theory: when is $f \in \mathrm{PW}_\Omega$ completely and stably determined by its samples $\{f(s) : s \in S\}$ on a set $S \subseteq \mathbb{R}$? On which sets $S \subseteq \mathbb{R}$ can every sequence $(a_s)_{s\in S} \in \ell^2(S)$ be interpolated by a function f in PW_Ω, so that $f(s) = a_s$ for all $s \in S$? These questions were answered by Beurling [1989] and Landau [1967].

Theorem A. (i) *Assume that S is uniformly separated and*

$$A\|f\|_2^2 \le \sum_{s\in S} |f(s)|^2 \le B\|f\|_2^2 \quad \textit{for all } f \in \mathrm{PW}_\Omega. \tag{1-1}$$

Then

$$D^-(S) = \liminf_{r\to\infty} \inf_{x\in\mathbb{R}} \frac{\#(S \cap [x-r, x+r])}{2r} \ge \frac{\Omega}{\pi}. \tag{1-2}$$

(ii) *If for all $a \in \ell^2(S)$ there exists $f \in \mathrm{PW}_\Omega$ such that $f(s) = a_s$, $s \in S$, then*

$$D^+(S) = \limsup_{r\to\infty} \sup_{x\in\mathbb{R}} \frac{\#(S \cap [x-r, x+r])}{2r} \le \frac{\Omega}{\pi}. \tag{1-3}$$

This research was funded in whole or in part by the Austrian Science Fund (FWF) 10.55776/P31887.
MSC2020: primary 35J99, 46E22, 47B32, 54D35, 94A20; secondary 42C40.
Keywords: spectral subspace, Paley–Wiener space, bandwidth, Beurling density, sampling, interpolation, elliptic operator, regularity theory, slow oscillation, Higson compactification.

In the established terminology, a set that satisfies a sampling inequality of the form (1-1) is called a sampling set for the underlying space PW_Ω, or a set of stable sampling. A set on which arbitrary ℓ^2-data can be interpolated is called a set of interpolation. The expressions $D^-(S)$ and $D^+(S)$ are called the lower and the upper Beurling density.

The number Ω/π in (1-2) and (1-3) is an important invariant of the space PW_Ω and has an interpretation in information theory. Since, roughly speaking, the densities $D^\pm(S)$ measure the average number of samples in S per unit length, the *necessary density conditions* of Theorem A say that at least Ω/π samples per unit length are required to recover a function in PW_Ω from $f|_S$, whereas at most Ω/π values per unit length are permitted to solve the interpolation problem in PW_Ω. Thus the density Ω/π represents a critical value below which (stable) sampling is impossible, and above which interpolation is impossible. Indeed, these questions about sampling and interpolation were at the origin of Shannon's information theory [1948], and the uniform sampling theorem with $S=\alpha\mathbb{Z}$ is still considered the basis of analog-digital conversion in modern signal processing. The ratio $D^\pm(S)/\Omega$ is a measure for the redundancy, thus for the performance quality, of the sampling set S. The theory of Beurling, Kahane, and Landau provides a rigorous mathematical formulation for the existence of a critical density for arbitrary sets S (in place of $\alpha\mathbb{Z}$). Although we will not touch this question here, we mention that the conditions of Theorem A yield almost a characterization of sets of sampling and of interpolation: *in dimension $d=1$, if S is uniformly separated and $D^-(S)>1$, then S is a sampling set for* PW_Ω, *and if $D^+(S)<1$, then S is a set of interpolation for* PW_Ω. See [Kahane 1962; Beurling 1989; Seip 2004] for an exposition of the sampling theory in the classical Paley–Wiener space.

The connection with partial differential operators comes from the observation that PW_Ω is a spectral subspace of the differential operator $H=-d^2/dx^2$. Using the Fourier transform $\mathscr{F}$, this differential operator is unitarily equivalent to the multiplication operator $\mathscr{F}(-d^2f/dx^2)(\xi)=\xi^2\hat{f}(\xi)$. In this representation of $-d^2/dx^2$ the spectral projection on the interval $[0,\Omega]$ is given by $\chi_{[0,\Omega]}(H)f=\mathscr{F}^{-1}(\chi_{[0,\Omega]}(\xi^2)\hat{f})$. This implies

$$\mathrm{PW}_\Omega=\chi_{[0,\Omega^2]}(H)L^2(\mathbb{R}).$$

This observation is the starting point for many generalizations of Paley–Wiener spaces and sampling theorems. In this work we study the question of necessary density conditions for sampling and interpolation in the spectral subspaces of a self-adjoint uniformly elliptic differential operator

$$H_a=-\sum_{j,k=1}^{d}\partial_j a_{jk}(x)\partial_k$$

acting on $L^2(\mathbb{R}^d)$ with a smooth positive definite (matrix) symbol $a=(a_{jk}(x))_{j,k=1,\dots,d}$. The Paley–Wiener space associated to H_a is the spectral subspace

$$\mathrm{PW}_\Omega(H_a)=\chi_{[0,\Omega]}(H_a)L^2(\mathbb{R}^d),$$

where, as usual, $\chi_{[0,\Omega]}(H_a)$ is the orthogonal projection corresponding to the spectrum $[0,\Omega]$.

If the symbol $a(x)=a$ is constant, then H_a is similar to the Laplace operator, and the corresponding spectral subspace can be described with Fourier techniques. For this case necessary density conditions for sampling and interpolation are already contained in [Landau 1967]. Optimal sufficient conditions

for sampling in $\mathbb{R}^d$ in terms of a covering density were obtained in [Beurling 1966]. However, if H_a is a uniformly elliptic differential operator with variable coefficients, then the standard techniques break down, and it was an open question whether a critical density exists for sampling and interpolation in the spectral subspaces of H_a, and how to compute this critical density.

We will answer this question for a class uniformly elliptic operators. We say a smooth symbol with all derivatives bounded, $a \in C_b^\infty(\mathbb{R}^d, \mathbb{C}^{d\times d})$, is slowly oscillating if $\lim_{|x|\to\infty} |\partial_k a(x)| = 0$ for $k = 1, \dots, d$.

Theorem B. *If a is slowly oscillating, then there exists a critical density for sampling and interpolation for* $\mathrm{PW}_\Omega(H_a)$.

Adapting the measure to the geometry associated to the differential operator H_a, the critical density can be determined explicitly. This is our main result.

Theorem C. *Assume $H_a = -\sum_{j,k=1}^d \partial_j a_{jk} \partial_k$ is a self-adjoint uniformly elliptic operator with slowly oscillating symbol $a \in C_b^\infty(\mathbb{R}^d, \mathbb{C}^{d\times d})$. Let $d\nu(x) = (\det a(x))^{-1/2}\, dx$ be the associated measure.*

(i) *If $S \subseteq \mathbb{R}^d$ is a set of stable sampling for* $\mathrm{PW}_\Omega(H_a)$ *then*

$$D_\nu^-(S) = \liminf_{r\to\infty} \inf_{x\in\mathbb{R}^d} \frac{\#(S \cap B_r(x))}{\nu(B_r(x))} \geq \frac{|B_1|}{(2\pi)^d} \Omega^{d/2}.$$

(ii) *If $S \subseteq \mathbb{R}^d$ is a set of interpolation for* $\mathrm{PW}_\Omega(H_a)$, *then*

$$D_\nu^+(S) = \limsup_{r\to\infty} \sup_{x\in\mathbb{R}^d} \frac{\#(S \cap B_r(x))}{\nu(B_r(x))} \leq \frac{|B_1|}{(2\pi)^d} \Omega^{d/2}.$$

Except for the modified definition of the density, the formulation of the theorem is identical to Landau's theorem [1967]. By contrast, the method of proof is vastly different as Fourier methods are not available for the proof of Theorem C. In addition we draw the new insight that the appropriate notion of density must be linked to the geometry defined by a. For compact manifolds the link between density and geometry was already observed in [Ortega-Cerdà and Pridhnani 2012].

For the special case of a symbol that is asymptotically constant at infinity we can use the standard Beurling densities in $\mathbb{R}^d$ from (1-2) and (1-3) (with intervals replaced by Euclidean balls $B_r(x)$) and obtain the following consequence.

Corollary D. *Assume that $a \in C_b^\infty(\mathbb{R}^d, \mathbb{C}^{d\times d})$ is asymptotically constant, i.e.,* $\lim_{x\to\infty} a(x) = b$. *Let* $\Sigma_\Omega^b = \{\xi \in \mathbb{R}^d : b\xi \cdot \xi \leq \Omega\}$.

(i) *If $S \subseteq \mathbb{R}^d$ is a set of sampling for the Paley–Wiener space* $\mathrm{PW}_\Omega(H_a)$, *then*

$$D^-(S) \geq \frac{|\Sigma_\Omega^b|}{(2\pi)^d} = (\det b)^{-1/2} \frac{|B_1|}{(2\pi)^d} \Omega^{d/2}.$$

(ii) *If $S \subseteq \mathbb{R}^d$ is a set of interpolation for the Paley–Wiener space* $\mathrm{PW}_\Omega(H_a)$, *then*

$$D^+(S) \leq \frac{|\Sigma_\Omega^b|}{(2\pi)^d}.$$

We note that the same critical density holds for the Paley–Wiener space of the constant-coefficient differential operator H_b. Since H_a may be considered a perturbation of H_b and since the Beurling density $D^{\pm}(S)$ is an asymptotic quantity, it is to be expected that the necessary density for $\mathrm{PW}_\Omega(H_a)$ coincides with the necessary density for $\mathrm{PW}_\Omega(H_b)$.

Let us put these statements into context.

Sampling in spectral subspaces. Several researchers have created an extensive qualitative theory of sampling in spectral subspaces of a general unbounded, positive, self-adjoint operator H on a Hilbert space $\mathscr{H}$. In this case the abstract Paley–Wiener space is defined as $\mathrm{PW}_{[0,\Omega]}(H) = \chi_{[0,\Omega]}(H)\mathscr{H}$. Usually $\mathscr{H} = L^2(X, \mu)$ and $\mathrm{PW}_{[0,\Omega]}(H)$ is a reproducing kernel Hilbert space. In this situation many authors have proved the existence of sampling sets [Coulhon et al. 2012; Feichtinger et al. 2016; Filbir and Mhaskar 2011; Feichtinger and Pesenson 2004; Pesenson 2000; 2001; Pesenson and Zayed 2009]. In particular the set-up of [Coulhon et al. 2012; Pesenson 1999; Pesenson and Zayed 2009] covers the case of H being a self-adjoint uniformly elliptic differential operator on $L^2(\mathbb{R}^d)$. The construction of sampling sets in these abstract Paley–Wiener spaces requires some smoothness properties of functions in $\mathrm{PW}_\Omega(H)$ and a Bernstein-type inequality (see (2-1) below). The result then is that a "sufficiently dense" subset in X is a sampling set and a "sufficiently sparse" subset of X is a set of interpolation. What remained unknown is the existence of a critical density against which one could compare the quality of the construction. Theorems B, C, and Corollary D address this gap for uniformly elliptic differential operators. Once a critical sampling density is established, one may aim for sampling sets near the critical density. The question of optimal sampling sets in spectral subspaces is wide open; in fact, it has become meaningful only after the critical density is known explicitly. This problem is already difficult for multivariate bandlimited functions $\mathrm{PW}_K = \{f \in L^2(\mathbb{R}^d) : \operatorname{supp} \hat{f} \subseteq K\}$ for compact spectrum $K \subseteq \mathbb{R}^d$ and was solved only recently in [Matei and Meyer 2010; Olevskiĭ and Ulanovskii 2008]. A possible general approach is via the construction of Fekete sets and weak limits, as was carried out in [Gröchenig et al. 2019] for Fock spaces with a general weight.

Insight for partial differential operators. Although the spectral subspaces of a partial differential operator are natural objects, they seem to have received little attention. To the best of our knowledge, nothing is known about the nature of the corresponding reproducing kernel and the behavior of functions in the spectral subspaces PW_Ω. Our investigation reveals several properties of the reproducing kernel, such as the behavior of its diagonal and some form of off-diagonal decay. These are key properties for the proofs of Theorems B and C, and we hope that these also hold some interest for partial differential operators.

Variable bandwidth. Our original motivation comes from a new concept of variable bandwidth. In [Gröchenig and Klotz 2017] we argued that the spectral subspaces of the Sturm–Liouville operator $-(d/dx)a(d/dx)$ on $L^2(\mathbb{R})$ for some function $a > 0$ can be taken as spaces with variable bandwidth. We proved that $a(x)^{-1/2}$ is a measure for the bandwidth near x (the largest active frequency at position x). The function a thus parametrizes the local bandwidth. For $a = \text{const.}$, the spectral subspace is just the classical Paley–Wiener space PW_Ω. For the special case of an eventually constant parametrizing function a, i.e., a is constant outside an interval $[-R, R]$, we computed the critical density for sampling in

$\mathrm{PW}_\Omega(-(d/dx)a(d/dx))$. The proof required intricate details of the scattering theory of one-dimensional Schrödinger operators. Theorem C, formulated for dimension $d = 1$, yields a significant extension of the density theorem for the sampling of functions of variable bandwidth.

Corollary E. *Assume that $a \in C_b^\infty(\mathbb{R})$ is bounded, $a > 0$, and $\lim_{x\to\pm\infty} a'(x) = 0$. Let $\mathrm{PW}_\Omega(H_a)$ be the Paley–Wiener space associated to H_a.*

If S is a sampling set for $\mathrm{PW}_\Omega(H_a)$, then

$$D_\nu^-(S) \geq \frac{\Omega^{1/2}}{\pi}.$$

Similarly, if S is a set of interpolation for $\mathrm{PW}_\Omega(H_a)$, then

$$D_\nu^+(S) \leq \frac{\Omega^{1/2}}{\pi}.$$

Methods. The proofs of Theorems B and C combine ideas and techniques from several areas of analysis.

Critical density in reproducing kernel Hilbert spaces. Originally, density theorems in the style of Landau — and there are dozens in analysis — were proved from scratch. In our approach we apply the results on sampling and interpolation in general reproducing kernel Hilbert spaces from [Führ et al. 2017]. The main insight was that it suffices to verify some geometric conditions on the measure space, such as a doubling condition of the underlying measure, and of the reproducing kernel, such as some form of off-diagonal decay. Once these conditions are satisfied, one obtains the existence of a critical density and can calculate it in terms of the averaged trace of the reproducing kernel. Since the geometric conditions are trivially satisfied for $\mathbb{R}^d$, our main technical difficulty is to understand the reproducing kernel of the spectral subspaces of a self-adjoint uniformly elliptic differential operator.

Regularity theory and heat kernel estimates. To study this reproducing kernel, we use the fundamental results of the regularity theory of elliptic differential operators. With these tools we investigate the smoothness of the reproducing kernel and compare various Sobolev norms on $\mathrm{PW}_\Omega(H_a)$. See Lemma 2.1 and Proposition 2.2. For an important technical detail (Proposition 2.2) we will need Gaussian estimates for the heat kernel, which we expect to play a key role in extensions of our theory.

Limit operators and slowly varying symbols. To connect asymptotic properties of the symbol a of a partial differential operator H_a to the spectral theory of H_a, we use the notion of limit operators. Although we do not use any elaborate results from this theory (see [Georgescu 2011; Rabinovich et al. 2004a; Špakula and Willett 2017]), limit operators are central to our arguments.

Higson compactification of $\mathbb{R}^d$. An important structure underlying the proof of Theorem C is a compactification of $\mathbb{R}^d$, the so-called Higson compactification. This is the compactification arising as the maximal ideal space of the C^*-algebra of slowly oscillating functions on $\mathbb{R}^d$. By Gelfand theory every slowly oscillating function can be identified with a continuous function on the Higson compactification $h\mathbb{R}^d$; see, e.g., [Rabinovich et al. 2004a; Roe 2003; Shteinberg 2000]. On a technical level we will show that for slowly oscillating symbols the mapping $x \to T_{-x}k_x$ of centered reproducing kernels can be extended continuously to the compactification $h\mathbb{R}^d$ (Proposition 6.3).

The underlying philosophy is summarized in the following diagram; we write $T_x f(z) = f(z-x)$ for the translation operator and k_x for the reproducing kernel of $\mathrm{PW}_\Omega(H_a)$:

$$\{T_{-x}a : x \in \mathbb{R}^d\} \text{ compact} \quad \Longrightarrow \quad \{T_{-x}H_aT_x : x \in \mathbb{R}^d\} \text{ compact} \quad \Longrightarrow \quad \{T_{-x}k_x : x \in \mathbb{R}^d\} \text{ compact.}$$

Thus, if $T_{x_n}a \to b$ in a suitable topology, then $T_{-x_n}H_aT_{x_n} \to H_b$ and the sequence of centered reproducing kernels $T_{-x_n}k_{x_n}$ converges to the reproducing kernel of $\mathrm{PW}_\Omega(H_b)$. In the considered examples the limit operator H_b is simpler than the original operator H_a, and this facilitates information about the reproducing kernel of $\mathrm{PW}_\Omega(H_a)$.

The paper is organized as follows. Section 2 prepares the background material on regularity theory, symbol classes for partial differential operators, and reproducing kernel Hilbert spaces. We prove the basic properties of the Paley–Wiener space $\mathrm{PW}_\Omega(H_a)$. Section 3 gives the precise formulation of the general density theorem for $\mathrm{PW}_\Omega(H_a)$. Its proof is given in Sections 4 and 5. In Section 6 we calculate the critical density for sampling in $\mathrm{PW}_\Omega(H_a)$ for the class of slowly varying symbols (Theorem C and Corollary D). We conclude with an outlook and collect additional material in Appendices A and B.

2. Preliminaries

2A. ***Notation.*** For a function f on $\mathbb{R}^d$ and $x, z \in \mathbb{R}^d$ we define the translation operator $T_x f(z) = f(z-x)$. The open Euclidean ball of radius r at x is $B_r(x)$, and $B_r = B_r(0)$.

We use standard multi-index notation; thus the differential operator D^α is $\partial^{|\alpha|}/(\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d})$ and the multivariate binomial symbol is $\binom{\alpha}{\gamma} = \prod_{j=1}^d \binom{\alpha_j}{\gamma_j}$ for multi-indices $\alpha, \gamma \in \mathbb{N}_0^d$.

We will denote the space of uniformly continuous and bounded functions on $\mathbb{R}^d$ with values in a Banach space X by $C_b^u(\mathbb{R}^d, X)$. The indices c, ∞, and 0 refer to the subspaces of compactly supported, smooth, and vanishing-at-infinity functions in $C(\mathbb{R}^d)$. Thus $C_b^\infty(\mathbb{R}^d, X)$ consists of all smooth X-valued functions with bounded derivatives of all orders. The space $C^\infty(\mathbb{R}^d, X)$ has the Fréchet space topology induced by the seminorms $|f|_{R,\alpha} = \sup_{x \in B_R(0)} \|D^\alpha f(x)\|_X$. If $X = \mathbb{C}$, we write $C_b^\infty(\mathbb{R}^d)$, etc.

The Fourier transform of $f \in L^1(\mathbb{R}^d)$ is

$$\mathscr{F} f(\omega) = \hat{f}(\omega) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(x) e^{-ix\cdot\omega}\, dx,$$

and $\mathscr{F}$ extends to a unitary operator on $L^2(\mathbb{R}^d)$ as usual. For every $s \geq 0$ the Sobolev space W_2^s is defined by

$$W_2^s = \left\{ f \in L^2(\mathbb{R}^d) : \|f\|_{W_2^s} = \left[(2\pi)^{-d/2} \int_{\mathbb{R}^d} |\hat{f}(\omega)|^2 (1+|\omega|^2)^s\, d\omega \right]^{1/2} < \infty \right\}.$$

If $s \in \mathbb{N}$, then $\|f\|_{W_2^s} \asymp \sum_{|\alpha| \leq s} \|D^\alpha f\|_2$. By the Sobolev embedding theorem, $W_2^s \hookrightarrow C_0(\mathbb{R}^d)$ for $s > d/2$.

Recall that a reproducing kernel Hilbert space $\mathscr{H}$ is a Hilbert space of functions defined on a set X such that $f(x) = \langle f, k_x \rangle_{\mathscr{H}}$ for all $f \in \mathscr{H}$ and $x \in X$. We write $k(x, y) = \overline{k_x(y)}$ for the reproducing kernel of $\mathscr{H}$. See, e.g., [Aronszajn 1950]. In particular, the Sobolev space W_2^s is a reproducing kernel Hilbert space with reproducing kernel $T_x\kappa,\ x \in \mathbb{R}^d$, where $\hat{\kappa}(\omega) = \hat{\kappa}_s(\omega) = (1+|\omega|^2)^{-s}$, by direct computation or by [Wendland 2005].

2B. *The generalized Paley–Wiener space and its basic properties.* Pesenson's idea [1998; 2001] (see also [Pesenson and Zayed 2009]) was to define an abstract Paley–Wiener space as a spectral subspace associated to an arbitrary positive, self-adjoint operator $H \geq 0$ with domain $\mathscr{D}(H)$ on a Hilbert space $\mathscr{H}$ and a spectral interval $[0, \Omega]$. Let $\chi_{[0,\Omega]}(H)$ be the spectral projection of H. Then the generalized Paley–Wiener space is defined as

$$\mathrm{PW}_\Omega(H) = \chi_{[0,\Omega]}(H)\mathscr{H}.$$

Equivalently, for a positive, self-adjoint operator, one can define the Paley–Wiener space $\mathrm{PW}_\Omega(H)$ by a Bernstein-type inequality: $f \in \mathrm{PW}_\Omega(H)$, if and only if $f \in \mathscr{D}(H^k)$ for all $k \in \mathbb{N}$, and

$$\|H^k f\|_2 \leq \Omega^k \|f\|_2 \quad \text{for all } k \in \mathbb{N}. \tag{2-1}$$

This is an easy consequence of the spectral theorem; see [Gröchenig and Klotz 2010; Pesenson 2001; Pesenson and Zayed 2009].

If $H = -d^2/dx^2$ on $L^2(\mathbb{R})$, then

$$\mathrm{PW}_\Omega(H) = \{f \in L^2(\mathbb{R}) : \operatorname{supp} \hat{f} \subseteq [-\sqrt{\Omega}, \sqrt{\Omega}]\}$$

is precisely the classical Paley–Wiener space, or in engineering language the space of band-limited functions with bandwidth $2\sqrt{\Omega}$.

Convention. In this work we consider positive, formally self-adjoint differential expressions $H = H_a$ of the form

$$H_a f = -\sum_{j,k=1}^{d} \partial_j a_{jk} \partial_k f, \qquad f \in W_2^2. \tag{2-2}$$

Here the *matrix symbol* $a \in C_b^\infty(\mathbb{R}^d, \mathbb{C}^{d\times d})$ is positive definite; i.e., $a_{jk} = \bar{a}_{kj} \in C_b^\infty(\mathbb{R}^d)$ and there exists $\theta > 0$ such that $a(x)\xi \cdot \xi \geq \theta|\xi|^2$ for all $\xi, x \in \mathbb{R}^d$. Then H_a is a positive, uniformly elliptic self-adjoint operator on $\mathbb{R}^d$ with domain $\mathscr{D}(H_a) = W_2^2$. In particular $C_c^\infty(\mathbb{R}^d)$ is a core for H_a; i.e., H_a is the operator closure of $H_a|_{C_c^\infty(\mathbb{R}^d)}$. The regularity theory of elliptic differential operators asserts that for every $k \in \mathbb{N}_0$ there is a $c_k \in \mathbb{R}$ such that

$$H_a^k + c_k : W_2^{2k} \to L^2(\mathbb{R}^d)$$

is a Hilbert space isomorphism. See [Zimmer 1990, Theorem 6.3.12] or the standard references [Agmon 1965; Shubin 1992]. For further use we record the fact that a uniformly elliptic operator is one-to-one on its domain and thus

$$0 \text{ is not an eigenvalue of } H_a. \tag{2-3}$$

To see this, we use the ellipticity and $f \in W_2^2$. Then the identity

$$\langle H_a f, f\rangle = \int \sum_{j,k} a_{jk} \partial_k f(x) \overline{\partial_j f(x)}\, dx = 0$$

implies that $\partial_j f \equiv 0$; thus $f = 0$.

Remark. We regard the mapping $a \mapsto H_a$ as a mapping from functions to operators (a symbolic calculus) and refer to a as the (matrix) symbol of the operator. This terminology differs slightly from

the usage in PDE, where the (principal) symbol of the differential operator $\sum_{|\alpha|\le m} a_\alpha D^\alpha$ is the function $p(x,\xi)=\sum_{|\alpha|=m} a_\alpha \xi^\alpha$ on $\mathbb{R}^{2d}$. For the second-order differential operator H_a in (2-2) the principal symbol is $p(x,\xi)=a(x)\xi\cdot\xi$. Since H_a is self-adjoint, the coefficients a_α are all real for $|\alpha|=2$.

First we verify that $\mathrm{PW}_\Omega(H_a)$ embeds in every Sobolev space.

Lemma 2.1. *The Paley–Wiener space* $\mathrm{PW}_\Omega(H_a)$ *is continuously embedded in all Sobolev spaces* W_2^s, $s\ge 0$, *and in* $C_0^\infty(\mathbb{R}^d)$. *As a consequence, on* $\mathrm{PW}_\Omega(H_a)$, *the* L^2*-norm and the Sobolev norms are equivalent.*

Proof. Let $f\in \mathrm{PW}_\Omega(H_a)$ and $k\in\mathbb{N}$. By elliptic regularity and Bernstein's inequality (2-1), $\|f\|_{W_2^{2k}} \asymp \|(H^k+c_k)f\|_2 \le (\Omega^k+|c_k|)\|f\|_2$. Consequently, $f\in\bigcap_{k\in\mathbb{N}} W_2^{2k}=\bigcap_{s\ge 0} W_2^s \subseteq C_0^\infty(\mathbb{R}^d)$ via the Sobolev embedding. □

Embeddings of Paley–Wiener spaces different from Lemma 2.1 can be found in [Feichtinger and Pesenson 2004].

Next we show that $\mathrm{PW}_\Omega(H_a)$ is a reproducing kernel Hilbert space in $L^2(\mathbb{R}^d)$.

Proposition 2.2. *There exists a reproducing kernel* $k_x\in \mathrm{PW}_\Omega(H_a)$ *such that* $\chi_{[0,\Omega]}(H_a)f(x)=\langle f,k_x\rangle$ *for all* $f\in L^2(\mathbb{R}^d)$ *and all* $x\in\mathbb{R}^d$. *In addition, there are positive constants* c, C *such that*

$$0<c\le \|k_x\|_2\le C \quad \textit{for all } x\in\mathbb{R}^d. \tag{2-4}$$

Proof. Let $f\in \mathrm{PW}_\Omega(H_a)$ and $s>d/2$. By Lemma 2.1, $f\in W_2^s$ and $\|f\|_2\asymp\|f\|_{W_2^s}$. Since W_2^s is a reproducing kernel Hilbert space, we obtain

$$|f(x)|=|\langle f,T_x\kappa\rangle_{W_2^s}|\le \|T_x\kappa\|_{W_2^s}\|f\|_{W_2^s}\le C\|f\|_2.$$

Thus $\mathrm{PW}_\Omega(H_a)$ is a reproducing kernel Hilbert space with kernel $k_x\in \mathrm{PW}_\Omega(H_a)$.

For the lower bound in (2-4) we do not have a proof based exclusively on regularity theory. Instead we refer to [Coulhon et al. 2012, Lemma 3.19], where the lower bound for the reproducing kernel was derived by means of heat kernel estimates. As some details and notation differ, we reproduce the proof in Appendix B. □

Proposition 2.3. *The mapping* $x\mapsto k_x$ *is continuous from* $\mathbb{R}^d$ *to* W_2^s, $s\ge 0$.

Proof. Since $k_x\in PW_\Omega(H_a)$ and $\|k_x\|_2$ is bounded by (2-4), Lemma 2.1 and the Sobolev embedding theorem imply that $C_1=\sup_{x,y\in\mathbb{R}}|\nabla k_x(y)|$ is finite; therefore

$$\begin{aligned}\|k_x-k_y\|_{W_2^s}^2 &\le C\|k_x-k_y\|_2^2 = C\big(k_x(x)-k_x(y)-k_y(x)+k_y(y)\big)\\ &\le 2C\sup_{z,w}|\nabla k_z(w)|\,|x-y|\le C'|x-y|.\end{aligned}$$

Consequently $x\to k_x$ is continuous. □

2C. *Sampling and interpolation in* $\mathrm{PW}_\Omega(H_a)$ *and the Beurling densities.* Let μ be a Borel measure on $\mathbb{R}^d$ that is equivalent to Lebesgue measure in the sense that $d\mu=h\,dx$ for a measurable function h with $0<c\le h(x)\le C$ for all $x\in\mathbb{R}^d$.

The lower Beurling density of S with respect to μ is defined as

$$D_\mu^-(S) = \liminf_{r\to\infty} \inf_{x\in\mathbb{R}^d} \frac{\#(S\cap B_r(x))}{\mu(B_r(x))}, \tag{2-5}$$

and the upper Beurling density of S is

$$D_\mu^+(S) = \limsup_{r\to\infty} \sup_{x\in\mathbb{R}^d} \frac{\#(S\cap B_r(x))}{\mu(B_r(x))}.$$

If $d\mu = dx$ we omit the subscript and write $D^\pm(S)$.

For sampling in reproducing kernel Hilbert spaces the relevant measure is $d\mu(x) = k(x,x)\,dx$. We call the Beurling density with respect to this measure the dimension-free density and write $D_0^\pm(S)$ for $D_\mu^\pm(S)$.

We say that the reproducing kernel k of a reproducing kernel Hilbert space $\mathscr{H} \subseteq L^2(\mathbb{R}^d, dx)$ satisfies the *weak localization* property (WL) if for every $\varepsilon > 0$ there is a constant $r = r(\varepsilon)$ such that

$$\sup_{x\in\mathbb{R}^d} \int_{\mathbb{R}^d\setminus B_r(x)} |k(x,y)|^2\,dy < \varepsilon^2. \tag{WL}$$

The discrete analog of the weak localization is the so-called *homogeneous approximation property* (HAP) of the reproducing kernel: Assume that S is such that $\{k_s : s \in S\}$ is a *Bessel sequence* for $\mathscr{H}$; i.e., S satisfies the upper sampling inequality $\sum_{s\in S}|f(s)|^2 \le C\|f\|_2^2$ for all $f \in \mathscr{H}$. Then for every $\varepsilon > 0$ there is a constant $r = r(\varepsilon)$ such that

$$\sup_{x\in\mathbb{R}^d} \sum_{s\in S\setminus B_r(x)} |k(x,s)|^2 < \varepsilon^2. \tag{HAP}$$

Under the assumptions of weak localization (WL) and (2-4), an upper sampling inequality implies that for some (and hence all) $\rho > 0$

$$\max_{x\in\mathbb{R}^d} \#(S\cap B_\rho(x)) < \infty.$$

We call such a set S *relatively separated*. See also [Führ et al. 2017, Lemma 3.7].

The two localization properties (WL) and (HAP) are the key properties of the reproducing kernel required for an abstract density theorem to hold. For reproducing kernel Hilbert spaces embedded in $L^2(\mathbb{R}^d)$ this can be stated as follows [Führ et al. 2017, Corollary 4.1].

Theorem 2.4. *Let $\mathscr{H} \subseteq L^2(\mathbb{R}^d, dx)$ be a reproducing kernel Hilbert space with kernel k. Assume that k satisfies the boundedness property (2-4) on the diagonal, the weak localization (WL) and the homogeneous approximation property (HAP).*

(i) *If S is a sampling set for $\mathscr{H}$, then $D_0^-(S) \ge 1$.*

(ii) *If S is an interpolating set for $\mathscr{H}$, then $D_0^+(S) \le 1$.*

This result holds under a set of natural assumptions on metric measure spaces and conditions on the reproducing kernel. We will not dwell on the geometric conditions, e.g., doubling measure, as these are clearly satisfied for $\mathbb{R}^d$ with μ equivalent to Lebesgue measure. We want to verify Theorem 2.4 for $\mathscr{H} = \mathrm{PW}_\Omega(H_a)$ for a suitable class of symbols a. The boundedness of the diagonal of the kernel was

already established in Proposition 2.2, (2-4). To prove Theorems B and C we therefore need to verify the properties (WL) and (HAP) for the reproducing kernel Hilbert space $\mathrm{PW}_\Omega(H_a)$.

Observe that (WL) is equivalent to the condition

$$\sup_{x\in\mathbb{R}^d}\int_{\mathbb{R}^d\setminus B_r(0)} |T_{-x}k_x(y)|^2\,dy < \varepsilon^2 \tag{2-6}$$

for the *centered* reproducing kernels. We will show the stronger statement that the set $\{T_{-x}k_x : x\in\mathbb{R}^d\}$ is relatively compact in $L^2(\mathbb{R}^d)$. The Riesz–Kolmogorov compactness theorem then implies (2-6) and thus (WL).

The proof of (HAP) requires some additional local regularity of k_x. We will use prominently elliptic regularity theory to show that $\{T_{-x}k_x : x\in\mathbb{R}^d\}$ is relatively compact in all Sobolev spaces W_2^s. For the proof of (HAP) it is fundamental that the point evaluation on $\mathrm{PW}_\Omega(H_a)$ can be expressed two-fold as

$$f(x) = \langle f, k_x\rangle_{L^2} = \langle f, T_x\kappa\rangle_{W_2^s} \quad \text{for all } f\in \mathrm{PW}_\Omega(H_a). \tag{2-7}$$

2D. ***Classes of symbols, limit operators.*** First we define the relevant symbol classes. Let

$$\tau_x(H_a) = T_{-x}H_aT_x = H_{T_{-x}a} \tag{2-8}$$

be the conjugation of H_a by the translation T_x. If $a\in C_b^\infty(\mathbb{R}^d,\mathbb{C}^{d\times d})$, observe that $\tau_x(H_a)$ is again a self-adjoint, uniformly elliptic operator with domain W_2^2 and core $C_c^\infty(\mathbb{R}^d)$. In this section we describe symbol classes that ensure that $\{\tau_x(H_a)f : x\in\mathbb{R}^d\}$ is relatively compact in $L^2(\mathbb{R}^d)$ for all $f\in C_c^\infty(\mathbb{R}^d)$. Equivalently, every sequence $\tau_{x_k}(H_a)f$ has a norm-convergent subsequence. If (x_k) is bounded, this follows from the continuity of $x\mapsto T_xf$. To treat unbounded sequences we need some terminology.

Since in Section 6 we will deal with a nonmetrizable compactification of $\mathbb{R}^d$, we formulate most results for nets $(x_\lambda)_{\lambda\in\Lambda}$ instead of sequences. (Here Λ is a directed set with a partial order $\succeq$ and we write $\lim_\lambda x_\lambda$ for the limit of a net when it exists.)

Definition 2.5. Assume $a\in C_b^\infty(\mathbb{R}^d,\mathbb{C}^{d\times d})$. If the net $(x_\lambda)_{\lambda\in\Lambda}\subset\mathbb{R}^d$ diverges to infinity and there is an operator $H\in\mathscr{B}(W_2^2, L^2(\mathbb{R}^d))$ such that $\lim_\lambda \tau_{x_\lambda}(H_a)f = Hf$ for all $f\in C_c^\infty(\mathbb{R}^d)$, then H is called a *limit operator* of H_a.

Remark 2.6. (i) Existence and uniqueness of the limit operator follow from the Banach–Steinhaus theorem.

(ii) We do not even scratch the surface of the method of limit operators: see, amongst many others, [Rabinovich et al. 2004a; 2004b; Špakula and Willett 2017], and in the C^*-algebra setting [Davies and Georgescu 2013; Georgescu 2011; 2018].

(iii) Limit operators are related to compactifications of $\mathbb{R}^d$. An example can be found in Section 6B.

2D1. *Compact orbits.* Identity (2-8) suggests that compactness properties of $\{\tau_x(H_a) : x\in\mathbb{R}^d\}$ are related to compactness properties of $\{T_{-x}a : x\in\mathbb{R}^d\}$, so we investigate these first.

Lemma 2.7. (i) *If $f\in C_b^\infty(\mathbb{R}^d)$, then $\{T_xf : x\in\mathbb{R}^d\}$ is relatively compact in the Fréchet space $C^\infty(\mathbb{R}^d)$ with respect to its topology of uniform convergence of all derivatives on compact sets.*

(ii) *In particular, if* $\lim_\lambda T_{x_\lambda} f = g$ *pointwise, then* $\lim_\lambda T_{x_\lambda} f = g$ *in* $C^\infty(\mathbb{R}^d)$. *The limit function* g *is again in* $C_b^\infty(\mathbb{R}^d)$.

Proof. (i) The space $C^\infty(\mathbb{R}^d)$ has the Heine–Borel property [Rudin 1973, 1.46], so it suffices to verify that $\{T_x f : x \in \mathbb{R}^d\}$ is bounded in $C^\infty(\mathbb{R}^d)$, which means that

$$\|D^\alpha T_x f\|_{L^\infty(B_r(0))} < C_{\alpha,r} \quad \text{for all } x \in \mathbb{R}^d \text{ and all } r > 0,\ \alpha \in \mathbb{N}_0^d.$$

But this is trivial for $f \in C_b^\infty(\mathbb{R}^d)$, since all derivatives are globally bounded.

(ii) We apply the following observation: A net converges to a limit g if and only if every subnet has a subnet that converges to g. By (i) every subnet of $(T_{x_\lambda} f)_{k\in\mathbb{N}}$ has a subnet $(T_{z_\lambda} f)_{k\in\mathbb{N}}$ that converges in $C^\infty(\mathbb{R}^d)$ (to the limit function g). We conclude that $(T_{x_\lambda} f)_{k\in\mathbb{N}}$ converges to g in $C^\infty(\mathbb{R}^d)$. As all functions and their derivatives of all orders are bounded and continuous, this is true for the limit as well. □

Proposition 2.8. *Let* $a \in C_b^\infty(\mathbb{R}^d, \mathbb{C}^{d\times d})$, $k, m \in \mathbb{N}_0$, *and assume* $\lim_\lambda T_{-x_\lambda} a = b$ *pointwise. Then, for every* $f \in W_2^{2m+2k}$

$$\lim_\lambda \|(\tau_{x_\lambda}(H_a^k) - H_b^k) f\|_{W_2^{2m}} = 0.$$

Proof. We treat the case $k = 1$ first and assume for the moment that $f \in C_c^\infty(\mathbb{R}^d)$. Set $a^{(\lambda)} = T_{-x_\lambda} a$. We can express H_a in the form $H_a = \sum_{|\beta|\le 2} a_\beta D^\beta$, with coefficients $a_\beta \in C_b^\infty(\mathbb{R}^d)$, and estimate, for every multindex α with $|\alpha| \le 2m$,

$$|D^\alpha(H_{a^{(\lambda)}} - H_b) f| = \Big| D^\alpha \sum_{|\beta|\le 2} (a_\beta^{(\lambda)} - b_\beta) D^\beta f \Big| = \Big| \sum_{|\beta|\le 2} \sum_{|\gamma|\le|\alpha|} \binom{\alpha}{\gamma} D^\gamma (a_\beta^{(\lambda)} - b_\beta) D^{\alpha-\gamma+\beta} f \Big|.$$

By Lemma 2.7 we have $\lim_\lambda D^\gamma a^{(\lambda)} = D^\gamma b$ uniformly on compact sets, so the convergence is actually uniform on supp f, and thus

$$\lim_\lambda \|D^\alpha(H_{a^{(\lambda)}} - H_b) f\|_\infty = 0.$$

Consequently

$$\begin{aligned}\|(H_{a^{(\lambda)}} - H_b) f\|_{W_2^{2m}} &\le C \max_{|\alpha|\le 2m} \|D^\alpha(H_{a^{(\lambda)}} - H_b) f\|_2 \\ &\le C|\operatorname{supp} f|^{1/2} \max_{|\alpha|\le 2m} \|D^\alpha(H_{a^{(\lambda)}} - H_b) f\|_\infty \to 0.\end{aligned}$$

As $C_c^\infty(\mathbb{R}^d)$ is dense in W_2^{2m+2}, and the operators $H_{a^{(\lambda)}}$ are uniformly bounded from W_2^{2m+2} to W_2^{2m}, a standard density argument (see, e.g., [Teschl 2009, Lemma 1.14]) implies $\|(H_{a^{(\lambda)}} - H_b) f\|_{W_2^{2m}} \to 0$ for all $f \in W_2^{2m+2}$.

For $k > 1$ observe that

$$H_a^k f - H_b^k f = H_a^{k-1}(H_a f - H_b f) + (H_a^{k-1} - H_b^{k-1}) H_b f.$$

As $\lim_\lambda \|(H_{a^{(\lambda)}} - H_b) f\|_{W_2^{2m}} = 0$ for $f \in W_2^{2m+2}$, the result follows by induction on k. □

Remark. The statement of the proposition and its proof are valid under the following more general conditions: $a_\lambda, b \in C_b^\infty(\mathbb{R}^d)$, $a_\lambda \xrightarrow{C^\infty} b$, and (H_{a_λ}) is uniformly bounded from W_2^2 to $L^2(\mathbb{R}^d)$.

Though not needed in the sequel, we state an interesting corollary that shows how compactness properties of the orbit $\{T_x a : x \in \mathbb{R}^d\}$ are transferred to compactness properties of $\{\tau_x(H_a) : x \in \mathbb{R}^d\}$.

Corollary 2.9. *If $a \in C_b^\infty(\mathbb{R}^d, \mathbb{C}^{d\times d})$ and $f \in C_c^\infty(\mathbb{R}^d)$ the set $\{\tau_x(H_a)f : x \in \mathbb{R}^d\}$ is relatively compact in every Sobolev space W_2^s, $s > 0$.*

Proof. The set $\{T_x a : x \in \mathbb{R}^d\}$ is relatively compact in $C^\infty(\mathbb{R}^d)$ by Lemma 2.7, and Proposition 2.8 says that the mapping $a \mapsto H_a f$ is continuous from $\overline{\{T_x a : x \in \mathbb{R}^d\}}^{C^\infty(\mathbb{R}^d)}$ to W_2^s. □

2D2. *Slowly oscillating symbols.* In the next step we single out a subclass of operators for which the spectral theory is sufficiently simple. In our approach it is essential that the limit operators do not have the endpoint 0 and Ω of the spectrum as eigenvalues. The limits of translates of *slowly oscillating* symbols are constant, if they exist (Lemma 2.13 below), so the limit operators are similar to the Laplacian. This will be used in Section 6 to compute the critical density.

Definition 2.10. An X-valued function $f \in C_b^u(\mathbb{R}^d, X)$ is slowly oscillating[1] if for all compact subsets $M \subset \mathbb{R}^d$

$$\lim_{|x|\to\infty} \sup_{m\in M} \|f(x) - f(x+m)\|_X = 0.$$

In fact, it suffices to use the closed unit ball $\bar{B}_1$ instead of an arbitrary compact set M.

We denote the space of all slowly oscillating functions on $\mathbb{R}^d$ by $C_h(\mathbb{R}^d, X)$ and define $C_h^\infty(\mathbb{R}^d, X) = C_h(\mathbb{R}^d, X) \cap C_b^\infty(\mathbb{R}^d, X)$.

The space $C_h(\mathbb{R}^d)$ with the $\|\cdot\|_\infty$-norm and pointwise multiplication is a commutative C^*-subalgebra of $C_b^u(\mathbb{R}^d)$.

We will need the following characterization of $C_h^\infty(\mathbb{R}^d, X)$. The statement is folklore, but we do not know a formal reference. For completeness we sketch the simple proof.

Lemma 2.11. *A function f is in $C_h^\infty(\mathbb{R}^d, X)$ if and only if $f \in C_b^\infty(\mathbb{R}^d, X)$ and $\lim_{|x|\to\infty} \partial_k f(x) = 0$ for all $1 \le k \le d$.*

Proof. Assume that $f \in C_b^\infty(\mathbb{R}^d, X)$ and $\lim_{|x|\to\infty} \partial_k f(x) = 0$ for all $1 \le k \le d$ and choose $M = [-h, h]^d$. Writing $m \in M$ as $m = \sum_{k=1}^d h_k e_k$ with $|h_k| \le h$, the difference in Definition 2.10 is

$$f\left(x + \sum_{k=1}^d h_k e_k\right) - f(x) = \sum_{k=0}^{d-1} \int_{x+\sum_{l\le k} h_l e_l}^{x+\sum_{l\le k+1} h_l e_l} \partial_{k+1} f.$$

This implies that $\sup_{m\in M} \|f(x+m) - f(x)\|_X \to 0$ for $|x| \to \infty$.

Conversely, assume that $f \in C_h^\infty(\mathbb{R}^d, X)$. Fix $\varepsilon > 0$ and let $\eta \in C_c^\infty(\mathbb{R}^d)$ with $\operatorname{supp} \eta \subset B_1$, $\eta \ge 0$, $\int_{\mathbb{R}^d} \eta(x)\,dx = 1$. Then $\eta_\tau(x) = \tau^{-d}\eta(\tau^{-1}x)$, $\tau > 0$, is an approximate unit. This implies that for $f \in C_u^b(\mathbb{R}^d, X)$ bounded and uniformly continuous

$$\lim_{\tau\to 0+} \sup_{x\in\mathbb{R}^d} \|f(x) - f * \eta_\tau(x)\|_X = 0. \tag{2-9}$$

To estimate the partial derivative of $f \in C_h^\infty(\mathbb{R}^d, X)$ we introduce the approximate unit:

$$\|\partial_k f(x)\|_X \le \|\partial_k f(x) - \eta_\tau * \partial_k f(x)\|_X + \|\eta_\tau * \partial_k f(x)\|_X = \mathrm{I}_\tau + \mathrm{II}_\tau.$$

[1] In the literature f is also called "of vanishing oscillation at infinity" or a Higson function.

Since all derivatives of $f \in C_h^\infty(\mathbb{R}^d, X)$ are bounded and uniformly continuous, by (2-9) there exists τ_0 such that $\mathrm{I}_\tau < \varepsilon/2$ for $t < \tau_0$. Fix $\tau < \tau_0$ and observe that $\eta_\tau * \partial_k f = \partial_k \eta_\tau * f$ and that $\int_{\mathbb{R}^d} \partial_k \eta_\tau(y)\, dy = 0$. So

$$\begin{aligned}\mathrm{II}_\tau = \|\eta_\tau * \partial_k f(x)\|_X &= \left\| \int_{\mathbb{R}^d} (f(x) - f(y)) \partial_k \eta_\tau(x-y)\, dy \right\|_X \\ &\le \sup_{y \in B_\tau(x)} \|f(x) - f(y)\|_X \int_{\mathbb{R}^d} |\partial_k \eta_\tau(x-y)|\, dy \\ &\le C_\tau \sup_{y \in B_\tau(0)} \|f(x) - f(x+y)\|_X.\end{aligned}$$

As $f \in C_h^\infty(\mathbb{R}^d, X)$, there is $R > 0$ such that $\mathrm{II}_\tau \le \varepsilon/(2C_\tau)$ for $|x| > R$, and thus $\|\eta_\tau * \partial_k f(x)\|_X < \varepsilon$ for $|x| > R$. □

Example 2.12. A typical example of a genuinely slowly oscillating function is $a(x) = \sin |x|^{1/2}(1 - \varphi(x))$ for some $\varphi \in C_c^\infty(\mathbb{R}^d)$ with $\varphi(x) = 1$ near 0. (The cut-off of the singularity at 0 serves to make all derivatives of a bounded, but, of course, it is immaterial for the asymptotic behavior.)

Our interest in $C_h(\mathbb{R}^d, X)$ stems from the following fact (see [Rabinovich et al. 2004a, Proposition 2.4.1]):

Lemma 2.13. *Assume that $f \in C_h(\mathbb{R}^d, X)$ and $(x_\lambda)_{\lambda \in \Lambda} \subset \mathbb{R}^d$ diverges to infinity, $|x_\lambda| \to \infty$. If $\lim_\lambda T_{-x_\lambda} f(x) = g(x)$ exists for all $x \in \mathbb{R}^d$, then g is constant.*

Proof. Let $x, x' \in \mathbb{R}^d$. Definition 2.10 with $M = \{x, x'\}$ shows that for all $\varepsilon > 0$ there exists an index $\lambda_\varepsilon = \lambda_\varepsilon(x, x')$ such that $\|f(x + x_\lambda) - f(x_\lambda)\|_X < \varepsilon/2$ and $\|f(x' + x_\lambda) - f(x_\lambda)\|_X < \varepsilon/2$ for all $\lambda \succeq \lambda_\varepsilon$. So $\|f(x + x_\lambda) - f(x' + x_\lambda)\|_X < \varepsilon$ for all $\lambda \succeq \lambda_\varepsilon$. If $g = \lim_\lambda T_{-x_\lambda} f$ exists, it follows that $\|g(x) - g(x')\|_X \le \varepsilon$. As $\varepsilon > 0$ was arbitrary, g must be constant. □

3. Statement of the density theorem

We state our main theorems. A first version describes a general setup for symbols in the class $C_b^\infty(\mathbb{R}^d, \mathbb{C}^{d\times d})$ under additional assumptions on the spectra of the limit operators. We then formulate a corollary for slowly oscillating symbols, where the assumptions on the limit operators are automatically satisfied. We discuss possible applications of the general version in Section 7.

Theorem 3.1. *Assume that $H_a = -\sum_{j,k=1}^d \partial_j a_{jk} \partial_k$ is uniformly elliptic with symbol $a \in C_b^\infty(\mathbb{R}^d, \mathbb{C}^{d\times d})$. Let $\mathrm{PW}_\Omega(H_a)$ be the Paley–Wiener space as defined in Section 2B. Assume that Ω is not an eigenvalue of any limit operator H_b.*

- *If S is a set of stable sampling for $\mathrm{PW}_\Omega(H_a)$, then*

$$D_0^-(S) \ge 1.$$

- *If S is a set of interpolation for $\mathrm{PW}_\Omega(H_a)$, then*

$$D_0^+(S) \le 1.$$

The following consequence is a more explicit version of Theorem B of the Introduction, where we have used the equivalence of Lemma 2.11 to avoid the formal definition of $C_h^\infty(\mathbb{R}^d)$.

Corollary 3.2. *Assume that $H_a = -\sum_{j,k=1}^d \partial_j a_{jk} \partial_k$ is uniformly elliptic with symbol $a \in C_h^\infty(\mathbb{R}^d, \mathbb{C}^{d\times d})$.*

- *If S is a set of stable sampling for $\mathrm{PW}_\Omega(H_a)$, then $D_0^-(S) \geq 1$.*
- *If S is a set of interpolation for $\mathrm{PW}_\Omega(H_a)$, then $D_0^+(S) \leq 1$.*

Proof of Corollary 3.2. If $a \in C_h^\infty(\mathbb{R}^d, \mathbb{C}^{d\times d})$, then by Lemma 2.7 every net $(x_\lambda)_{\lambda\in\Lambda} \subset \mathbb{R}^d$ that diverges to infinity has a subnet $(x_\mu)_{\mu\in M}$ such that $\lim_\mu T_{-x_\mu} a = b$ in the topology of $C^\infty(\mathbb{R}^d)$ for a symbol b. This symbol b is constant by Lemma 2.13 and positive definite; so H_b is similar to the Laplacian and has no point spectrum. □

4. Proof of weak localization of the kernel

To prove Theorem 3.1 we invoke Theorem 2.4 and verify its main hypotheses (WL) and (HAP) on the reproducing kernel.

Let $Q_h = [-h/2, h/2]^d$ be the cube of side-length h, and let $\varphi_x^h(y) = h^{-d}\chi_{Q_h}((y-x)/h) = T_x\varphi_0^h(y)$ be the usual approximate unit.

Lemma 4.1. *We have $\lim_{h\to 0}\|\chi_{[0,\Omega]}(H_a)\varphi_x^h - k_x\|_2 = 0$ uniformly in $x \in \mathbb{R}^d$.*

Proof. Let $f \in \mathrm{PW}_\Omega(H_a)$. Then

$$\begin{aligned}|\langle f, \chi_{[0,\Omega]}(H_a)\varphi_x^h - k_x\rangle| &= |\langle f, \varphi_x^h\rangle - f(x)| = h^{-d}\left|\int_{Q_h(x)} (f(y)-f(x))\,dy\right| \\ &\leq h^{-d}\int_{Q_h(x)} |f(y)-f(x)|\,dy \leq \sup_{z\in\mathbb{R}^d}|\nabla f(z)|\,h^{-d}\int_{Q_h(x)}|y-x|\,dy \\ &\leq C\|\nabla f\|_\infty h.\end{aligned}$$

Since $f \in W_2^s(\mathbb{R}^d)$ for all $s \geq 0$, we apply first the Sobolev embedding (with $s > d/2+1$) and then Lemma 2.1 and obtain

$$\|\nabla f\|_\infty \leq C_1\|f\|_{W_2^s} \leq C\|f\|_2,$$

since $f \in \mathrm{PW}_\Omega(H_a)$. Consequently,

$$|\langle f, \chi_{[0,\Omega]}(H_a)\varphi_x^h - k_x\rangle| \leq Ch\|f\|_2.$$

Taking the supremum over $f \in \mathrm{PW}_\Omega(H_a)$, we obtain

$$\|\chi_{[0,\Omega]}(H_a)\varphi_x^h - k_x\|_2 = \sup_{f\in\mathrm{PW}_\Omega(H_a),\|f\|_2=1}\langle f, \chi_{[0,\Omega]}(H_a)\varphi_x^h - k_x\rangle \leq Ch.$$

As this estimate is independent of x, we have shown that $\chi_{[0,\Omega]}(H_a)\varphi_x^h \to k_x$ in $L^2(\mathbb{R}^d)$ uniformly in x. □

The following result relates the reproducing kernel of a limit operator of H_a to the original kernel. It expresses a form of continuous dependence of the reproducing kernel of the matrix symbol of H_a. We will denote the point spectrum of an operator H by $\sigma_p(H)$.

Theorem 4.2. *Let H_a with symbol $a \in C_b^\infty(\mathbb{R}^d, \mathbb{C}^{d\times d})$, and let $(x_\lambda)_{\lambda\in\Lambda} \subset \mathbb{R}^d$ be an unbounded net such that $\lim_\lambda T_{-x_\lambda} a = b$ pointwise. Assume that $\Omega \notin \sigma_p(H_b)$. Let $\tilde{k}$ be the reproducing kernel of $\mathrm{PW}_\Omega(H_b)$. Then*

$$\lim_\lambda T_{-x_\lambda} k_{x_\lambda} = \tilde{k}_0,$$

with convergence in W_2^s for every $s \geq 0$.

Before the proof we remind the reader of the following standard facts of spectral theory; see, e.g., [Teschl 2009, Chapter 6.6]. Although in the literature these results are formulated for sequences of operators, the statements and proofs are equally valid for nets.[2]

Let H_λ, $\lambda \in \Lambda$, and H_b be self-adjoint operators with a common core $\mathscr{D}$. If $H_\lambda f \to H_b f$ for all $f \in \mathscr{D}$, then, for every $F \in C_b(\mathbb{R})$,

$$F(H_\lambda) f \to F(H_b) f \quad \text{for all } f \in L^2(\mathbb{R}^d). \tag{4-1}$$

Furthermore, if $\chi_{\{\alpha\}}(H_b) = \chi_{\{\beta\}}(H_b) = 0$, i.e., $\alpha, \beta \notin \sigma_p(H_b)$, then

$$\chi_{[\alpha,\beta]}(H_\lambda) f \to \chi_{[\alpha,\beta]}(H_b) f \quad \text{for all } f \in L^2(\mathbb{R}^d). \tag{4-2}$$

Proof of Theorem 4.2. We split the difference $T_{-x_\lambda} k_{x_\lambda} - \tilde{k}_0$ into three terms and then estimate their W_2^s-norms separately:

$$\begin{aligned}
&\|T_{-x_\lambda} k_{x_\lambda} - \tilde{k}_0\|_{W_2^s} \\
&\quad \leq \|T_{-x_\lambda} k_{x_\lambda} - T_{-x_\lambda}\chi_{[0,\Omega]}(H_a)\varphi^h_{x_\lambda}\|_{W_2^s} + \|T_{-x_\lambda}\chi_{[0,\Omega]}(H_a)\varphi^h_{x_\lambda} - \chi_{[0,\Omega]}(H_b)\varphi^h_0\|_{W_2^s} + \|\chi_{[0,\Omega]}(H_b)\varphi^h_0 - \tilde{k}_0\|_{W_2^s} \\
&\quad = (I) + (II) + (III).
\end{aligned}$$

Choose $\varepsilon > 0$.

<u>Step 1</u>: Expression (I) can be estimated by

$$\|T_{-x_\lambda} k_{x_\lambda} - T_{-x_\lambda}\chi_{[0,\Omega]}(H_a)\varphi^h_{x_\lambda}\|_{W_2^s} = \|k_{x_\lambda} - \chi_{[0,\Omega]}(H_a)\varphi^h_{x_\lambda}\|_{W_2^s} \leq C_s \|k_{x_\lambda} - \chi_{[0,\Omega]}(H_a)\varphi^h_{x_\lambda}\|_2.$$

The first equality holds by the translation invariance of the Sobolev norm; the second inequality is a consequence of Lemma 2.1. By Lemma 4.1 there exists $h_\varepsilon > 0$ such that, for every $0 < h < h_\varepsilon$,

$$\|k_x - \chi_{[0,\Omega]}(H_a)\varphi^h_x\|_2 < \frac{\varepsilon}{3C_s}$$

for *all* $x \in \mathbb{R}^d$. So for $h < h_\varepsilon$, we obtain $(I) < \varepsilon/3$. Similarly, we achieve $(III) < \varepsilon/3$ for every $h < h'_\varepsilon$.

<u>Step 2</u>: To bound the decisive term (II), we bring in limit operators and elliptic regularity theory. Set $a_\lambda = T_{-x_\lambda} a$. First note that

$$T_{-x_\lambda}\chi_{[0,\Omega]}(H_a)\varphi^h_{x_\lambda} = T_{-x_\lambda}\chi_{[0,\Omega]}(H_a)T_{x_\lambda}\varphi^h_0 = \chi_{[0,\Omega]}(\tau_{x_\lambda} H_a)\varphi^h_0 = \chi_{[0,\Omega]}(H_{a_\lambda})\varphi^h_0.$$

We have to verify that

$$\lim_\lambda \|\chi_{[0,\Omega]}(H_{a_\lambda})\varphi^h_0 - \chi_{[0,\Omega]}(H_b)\varphi^h_0\|_{W_2^s} = 0. \tag{4-3}$$

[2]The cited results use the strong operator topology. As this topology is metrizable on bounded sets, the convergence of nets is equivalent to the convergence of sequences.

For L^2-convergence ($s=0$) we argue as follows. By Lemma 2.7 the translates $T_{-x_\lambda}a$ converge to the matrix b uniformly on compact sets. Proposition 2.8 implies that $H_{T_{-x_\lambda}a}f \to H_b f$ for $f \in W_2^s$, $s \geq 0$. To apply (4-2), we note that $C_c^\infty(\mathbb{R}^d)$ is a common core for all H_{a_λ} and for H_b and that $0 \notin \sigma_p(H_b)$ by (2-3) and $\Omega \notin \sigma_p(H_b)$ by assumption. Therefore (4-3) follows from (4-2).

For the convergence of (4-3) in general Sobolev spaces W_2^s it suffices to treat the case $s=2k$ for every integer k. Recall that by the results on elliptic regularity in Section 2B the operator $(H_a^k + c_k)$ defines an isomorphism from $W_2^{2k}(\mathbb{R}^d)$ to $L^2(\mathbb{R}^d)$, and since $\tau_{x_\lambda}(H_a^k + c_k) = H_{a_\lambda}^k + c_k$ we obtain

$$\|H_{a_\lambda}^k + c_k\|_{W_2^{2k}\to L^2} = \|H_a^k + c_k\|_{W_2^{2k}\to L^2} < \infty.$$

The Sobolev norm can be estimated by the L^2-norm

$$\|f\|_{W_2^{2k}} = \|T_x f\|_{W_2^{2k}} \leq C_s\|(H_a^k + c_k)T_x f\|_2 = C_s\|T_{-x}(H_a^k + c_k)T_x f\|_2$$

independently of $x \in \mathbb{R}^d$. Thus (II) can be estimated by the L^2-norm, namely

$$\begin{aligned}\|\chi_{[0,\Omega]}(H_{a_\lambda})\varphi_0^h - \chi_{[0,\Omega]}(H_b)\varphi_0^h\|_{W_2^s} &\leq C_s\|(H_{a_\lambda}^k + c_k)\chi_{[0,\Omega]}(H_{a_\lambda})\varphi_0^h - (H_{a_\lambda}^k + c_k)\chi_{[0,\Omega]}(H_b)\varphi_0^h\|_2\\ &\leq C_s\|(H_{a_\lambda}^k + c_k)\chi_{[0,\Omega]}(H_{a_\lambda})\varphi_0^h - (H_b^k + c_k)\chi_{[0,\Omega]}(H_b)\varphi_0^h\|_2\\ &\qquad + C_s\|(H_b^k + c_k)\chi_{[0,\Omega]}(H_b)\varphi_0^h - (H_{a_\lambda}^k + c_k)\chi_{[0,\Omega]}(H_b)\varphi_0^h\|_2\\ &= A_\lambda + B_\lambda.\end{aligned}$$

By Proposition 2.8 we have $(H_{a_\lambda}^k + c_k)f \to (H_b^k + c_k)f$ in L^2-norm for all $f \in W_2^{2k}$. In particular, this holds for $f = \chi_{[0,\Omega]}(H_b)\varphi_0^h$; thus $\lim_\lambda B_\lambda = 0$.

For the first term we use spectral theory again. Define $F \in C_c(\mathbb{R})$ such that its restriction to $[0,\Omega]$ satisfies

$$F(t) = t^k + c_k \quad \text{for } t \in [0,\Omega].$$

Then $F(t)\chi_{[0,\Omega]}(t) = (t^k + c_k)\chi_{[0,\Omega]}(t)$, and $\lim_\lambda F(\tau_{x_\lambda}(H_a))f = F(H_b)f$ for all $f \in L^2(\mathbb{R}^d)$ by (4-1). Since the product of bounded operators is continuous in the strong operator topology, it follows that

$$\begin{aligned}\lim_\lambda (H_{a_\lambda}^k + c_k)\chi_{[0,\Omega]}(H_{a_\lambda})\varphi_0^h &= \lim_\lambda F(H_{a_\lambda})\big(\lim_\lambda \chi_{[0,\Omega]}(H_{a_\lambda})\varphi_0^h\big)\\ &= F(H_b)\chi_{[0,\Omega]}(H_b)\varphi_0^h = (H_b^k + c_k)\chi_{[0,\Omega]}(H_b)\varphi_0^h,\end{aligned}$$

and so $\lim_\lambda A_\lambda = 0$.

We can finish the proof as follows. We have already chosen $h < \min\{h_\varepsilon, h'_\varepsilon\}$ so that the terms (I) and (III) are $< \varepsilon/3$ for all $\lambda \in \Lambda$. For this fixed $h > 0$ we can find an index λ_0 such that

$$(II) \leq C_s\|(H_{a_\lambda}^k + c_k)\chi_{[0,\Omega]}(H_{a_\lambda})\varphi_0^h - (H_{a_\lambda}^k + c_k)\chi_{[0,\Omega]}(H_b)\varphi_0^h\|_2 < \frac{\varepsilon}{3}$$

for all $\lambda \succeq \lambda_0$. Altogether we obtain $\|T_{-x_\lambda}k_{x_\lambda} - \tilde{k}_0\|_2 \leq (I) + (II) + (III) < \varepsilon$. □

Theorem 4.3. *Assume that H_a is uniformly elliptic with symbol $a \in C_b^\infty(\mathbb{R}^d, \mathbb{C}^{d\times d})$ and that no limit operator has the eigenvalue Ω. Then the set $\{T_{-x}k_x : x \in \mathbb{R}^d\}$ is relatively compact in W_2^s for every $s \geq 0$.*

Proof. This follows directly from Theorem 4.2. Let $(x_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}^d$ be an arbitrary sequence. By Lemma 2.7 the sequence $T_{-x_n}a$ has a C^∞-convergent subsequence $T_{-x_{n_l}}a$. If $(x_{n_l})_{l\in\mathbb{N}}$ is bounded, we can assume without loss of generality that $x_{n_l} \to x \in \mathbb{R}^d$, and $T_{-x_{n_l}}k_{x_{n_l}} \to T_{-x}k_x$ in W_2^s by the continuity of the translations and Proposition 2.3. If $(x_{n_l})_{l\in\mathbb{N}}$ is unbounded, we can assume $|x_{n_l}| \to \infty$. This case is settled by Theorem 4.2 and yields the convergence of $T_{-x_{n_l}}k_{x_{n_l}}$. □

A combination of the above arguments yields the weak localization (WL) .

Theorem 4.4. *Assume that H_a is uniformly elliptic with symbol $a \in C_b^\infty(\mathbb{R}^d, \mathbb{C}^{d\times d})$ and that no limit operator has the eigenvalue Ω. Let k be the reproducing kernel of* $\mathrm{PW}_\Omega(H_a)$. *Then k satisfies the weak localization property* (WL), *i.e.*,

$$\lim_{R\to\infty} \int_{|y-x|>R} |k(x,y)|^2\, dy = 0.$$

Proof. By Theorem 4.3 (for $s = 0$) the set $\{T_{-x}k_x : x \in \mathbb{R}^d\}$ is relatively compact in $L^2(\mathbb{R}^d)$. The Riesz–Kolmogorov theorem implies that for all $\varepsilon > 0$ there is $R > 0$ such that for all $x \in \mathbb{R}^d$

$$\int_{\mathbb{R}^d\setminus B_R(0)} |T_{-x}k_x(y)|^2\, dy < \varepsilon^2.$$

By a change of variable this expression reads as

$$\int_{|y-x|>R} |k(x,y)|^2\, dy < \varepsilon^2,$$

and this is (WL). □

5. Proof of the homogeneous approximation property (HAP)

Next we prove the homogeneous approximation property. Recall that $T_x\kappa$ is the reproducing kernel for W_2^s with $\hat{\kappa}(\omega) = (1+|\omega|^2)^{-s}$.

Lemma 5.1. *If S is a relatively separated set in $\mathbb{R}^d$, then $\{T_x\kappa : x \in S\}$ is a Bessel sequence for W_2^s, $s > d/2$.*

Proof. By standard facts of frame theory (see, e.g., [Heil 2011, Theorem 7.6]) the Bessel property is equivalent to the boundedness of the Gramian $G = (\langle T_x\kappa, T_y\kappa\rangle_{W_2^s})_{x,y\in S}$ on $\ell^2(S)$. To deduce the boundedness of G we first show that G possesses exponential off-diagonal decay and then apply Schur's test. The off-diagonal decay follows from a (well-known) calculation. Let $\mathscr{J}_r$ denote the Bessel function of the first kind and $\mathscr{K}_r$ the modified Bessel function of the second kind. Then by [Wendland 2005, Theorem 6.13] or [Grafakos 2004, Appendix B]

$$\begin{aligned}
\langle T_x\kappa, T_y\kappa\rangle_{W_2^s} &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \widehat{T_x\kappa}(\omega)\overline{\widehat{T_y\kappa}}(\omega)(1+|\omega|^2)^s\, d\omega \\
&= (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{-i(x-y)\omega}(1+|\omega|^2)^{-s}\, d\omega \\
&= C|x-y|^{-(d-2)/2} \int_0^\infty (1+r^2)^{-s}\, \mathscr{J}_{(d-2)/2}(r|x-y|) r^{d/2}\, dr \\
&= C'|x-y|^{s-d/2} \mathscr{K}_{s-d/2}(|x-y|).
\end{aligned}$$

Using the asymptotic decay $\mathscr{K}_r(x) \sim \sqrt{\pi/(2x)}e^{-x}$ for $x \to \infty$, see, e.g., [DLMF 2020, equation 10.25.3], the off-diagonal decay of G is

$$|\langle T_x\kappa, T_y\kappa\rangle_{W_2^s}| \le C''|x-y|^{s-d/2-1/2}e^{-|x-y|} \quad (|x-y| \to \infty). \tag{5-1}$$

The off-diagonal decay of the Gramian implies the boundedness of the Gramian as follows. By (5-1) there exists $N_0 \in \mathbb{N}$ such that $|G_{xy}| \le Ce^{-c|x-y|}$ if $|x-y| > N_0$. Obviously, $|G_{xy}| \le \|\kappa\|^2_{W_2^s}$ is bounded for all x, y.

For $x \in S$ and $k \in \mathbb{N}_0$ set $A_k(x) = \{y \in S : k < |y-x| \le k+1\}$. Since $S \subset \mathbb{R}^d$ is relatively separated, there exists $r > 0$ such that

$$\max \#(S \cap B_r(x)) < \infty.$$

A covering argument (of a large ball $B_R(z)$ by balls $B_r(x)$) implies that $\#(S \cap B_R(z)) = \mathscr{O}(R^d)$ for arbitrary $R > 0$. Consequently we also obtain $\#A_k(x) \le Ck^d$ independent of x. Then

$$\begin{aligned}\sum_{y\in S}|G_{xy}| &= \sum_{k=0}^{\infty}\sum_{y\in A_k(x)}|G_{xy}| = \sum_{k=0}^{N_0}\sum_{y\in A_k(x)}|G_{xy}| + \sum_{k>N_0}\sum_{y\in A_k(x)}|G_{xy}| \\ &\le C_0\,\#(B_{N_0+1}(x)\cap S) + C\sum_{k>N_0}e^{-ck}\,\#A_k(x) \\ &\le C_1(N_0+1)^d + C_2\sum_{k>N_0}e^{-ck}k^d.\end{aligned}$$

This expression is bounded independently of x. Now Schur's test implies that the Gramian is bounded on $\ell^2(S)$. □

Theorem 5.2 (HAP). *Assume that H_a is uniformly elliptic with symbol $a \in C_b^\infty(\mathbb{R}^d, \mathbb{C}^{d\times d})$ and that $\Omega \notin \sigma_p(H_b)$ for every limit operator H_b. Let $\{k_x : x \in S\}$ be a Bessel sequence in $\mathrm{PW}_\Omega(H_a)$. Then for every $\varepsilon > 0$ there exists an $R > 0$ such that for all $y \in \mathbb{R}^d$*

$$\sum_{x\in S\setminus B_R(y)}|k(y,x)|^2 < \varepsilon^2.$$

Proof. If $\{k_x : x \in S\}$ is a Bessel sequence of reproducing kernels, then S is relatively separated in $\mathbb{R}^d$ (see [Führ et al. 2017, Lemma 3.7]). Lemma 5.1 implies that $\{T_x\kappa : x \in S\}$ is also a Bessel sequence in W_2^s for $s > d/2$.

Choose $\varepsilon > 0$. Since $\{T_{-x}k_x : x \in \mathbb{R}^d\}$ is relatively compact in W_2^s for $s \ge 0$ by Theorem 4.3, the Riesz–Kolmogorov theorem for translation-invariant Banach spaces [Feichtinger 1984] asserts that there exists a $R = R_\varepsilon > 0$ and a function $\psi \in C_c^\infty(\mathbb{R}^d)$ satisfying $\psi|_{B_{R/2}(0)} = 1$, $\operatorname{supp}\psi \subseteq B_R(0)$ such that

$$\|T_{-x}k_x(1-\psi)\|_{W_2^s} \le \varepsilon \quad \text{for all } x \in \mathbb{R}^d.$$

We now use the fundamental observation (2-7) that the point evaluation in $\mathrm{PW}_\Omega(H_a)$ can be expressed in two ways. For $f = k_x$ we have

$$k(y,x) = \langle k_y, k_x\rangle_{L^2} = \langle k_y, T_x\kappa\rangle_{W_2^s}. \tag{5-2}$$

Since $\{T_x\kappa : x \in S\}$ is a Bessel sequence in W_2^s with bound B, the set $\{T_{x-y}\kappa : x \in S\}$ is a Bessel sequence with the same bound. Observe that for $|u| > R$ we obtain

$$\langle \psi T_{-y}k_y, T_u\kappa\rangle_{W_2^s} = T_{-y}k_y(u)\psi(u) = 0.$$

This implies

$$\begin{aligned}\sum_{x\in S\setminus B_R(y)} |\langle k_y, T_x\kappa\rangle_{W_2^s}|^2 &= \sum_{x\in S\setminus B_R(y)} |\langle T_{-y}k_y, T_{x-y}\kappa\rangle_{W_2^s}|^2 \\ &= \sum_{x\in S\setminus B_R(y)} |\langle (1-\psi)T_{-y}k_y, T_{x-y}\kappa\rangle_{W_2^s}|^2 \\ &\le B\|(1-\psi)T_{-y}k_y\|_{W_2^s}^2 \le B\varepsilon^2,\end{aligned}$$

and this is the homogeneous approximation property, (HAP). □

Proof of Theorem 3.1. After the verification of the properties (WL) and (HAP) of the reproducing kernel, the version for the dimension-free density of Theorem 2.4 is applicable and yields Theorem 3.1. The statement asserts the existence of a critical density for sampling and interpolation with the dimension-free Beurling density $D_0^{\pm}(S)$. □

6. Geometric Beurling densities

In this section we derive results for the geometric densities (2-5). According to Theorem 2.4 this step requires the computation of the averaged trace $|\mu(B_r(x))|^{-1}\int_{B_r(x)} k(x,x)\,d\mu(x)$ of the reproducing kernel. This version of the density theorems is of interest because it relates the critical density in $\mathrm{PW}_\Omega(H_a)$ to the geometry defined by the differential operator H_a. The explicit computation of the averaged trace becomes possible by introducing a suitable compactification of $\mathbb{R}^d$ and then extending the centered kernels $T_{-x}k_x$ to this compactification.

6A. *The basic computation*: *constant coefficients.* For reference we mention the case when $H_b = -\sum_{j,k}\partial_j b_{jk}\partial_k = -\sum_{j,k} b_{jk}\partial_j\partial_k$ is a differential operator with constant coefficients b_{jk}. Define

$$\Sigma_\Omega^b = \{\xi \in \mathbb{R}^d : b\xi\cdot\xi \le \Omega\} = b^{-1/2}\overline{B_{\Omega^{1/2}}(0)},$$

with volume

$$|\Sigma_\Omega^b| = \det(b^{-1/2})|B_{\Omega^{1/2}}(0)| = \det(b^{-1/2})\Omega^{d/2}|B_1|. \tag{6-1}$$

Since $\widehat{H_b f}(\xi) = \sum_{j,k} b_{jk}\xi_j\xi_k\hat{f}(\xi) = (b\xi\cdot\xi)\hat{f}(\xi)$, the spectral subspace is

$$\mathrm{PW}_\Omega(H_b) = \chi_{[0,\Omega]}(H_b)L^2(\mathbb{R}^d) = \{f \in L^2(\mathbb{R}^d) : \operatorname{supp}\hat{f} \subseteq \Sigma_\Omega^b\}.$$

The kernel of $\mathrm{PW}_\Omega(H_b)$ is

$$\tilde{k}(x,y) = (2\pi)^{-d/2}(\mathscr{F}^{-1}\chi_{\Sigma_\Omega^b})(x-y), \tag{6-2}$$

whence

$$\tilde{k}(x,x) = (2\pi)^{-d/2}(\mathscr{F}^{-1}\chi_{\Sigma_\Omega^b})(0) = \frac{|\Sigma_\Omega^b|}{(2\pi)^d} = \frac{|B_1|}{(2\pi)^d}\det(b^{-1/2})\Omega^{d/2}.$$

By Landau's theorem [1967] a sampling set S for $\mathrm{PW}_\Omega(H_b)$ has Beurling density at least $D^-(S) \ge |\Sigma_\Omega^b|/(2\pi)^d$.

6B. *The Higson compactification.* We recall how a compactification arises in Gelfand theory. Let $C_\gamma(\mathbb{R}^d)$ be a unital C^* algebra of functions on $\mathbb{R}^d$ satisfying the embeddings $C_0(\mathbb{R}^d) \subset C_\gamma(\mathbb{R}^d) \subset C_b(\mathbb{R}^d)$. The *maximal ideal space* M_γ of $C_\gamma(\mathbb{R}^d)$ is the space of all multiplicative homomorphisms $\varphi : C_\gamma(\mathbb{R}^d) \to \mathbb{C}$. Equipped with the weak-star topology M_γ is a compact Hausdorff space. The point evaluations $\delta_x(f) = f(x)$ constitute an embedding γ of $\mathbb{R}^d$ into M_γ via $\gamma(x) = \delta_x$, and $\gamma(\mathbb{R}^d)$ is dense in M_γ and homeomorphic to $\mathbb{R}^d$. Thus, the pair (γ, M_γ) is a compactification of $\mathbb{R}^d$, which we will call $\gamma\mathbb{R}^d$. The corona of $\gamma\mathbb{R}^d$ is $\partial_\gamma\mathbb{R}^d = \gamma\mathbb{R}^d \setminus \gamma(\mathbb{R}^d)$. By abuse of notation we will identify a point $x \in \mathbb{R}^d$ with its point evaluation δ_x. Then $C_\gamma(\mathbb{R}^d)$ is isometrically isomorphic to $C(\gamma\mathbb{R}^d)$. We denote the image of $f \in C_\gamma(\mathbb{R}^d)$ in $C(\gamma\mathbb{R}^d)$ by $\bar{f}$. See, e.g., [Engelking 1977] for the basics of compactifications, and [Gamelin 1969] for compactifications of function algebras.

As noted in Section 2D2 the space $C_h(\mathbb{R}^d)$ of slowly oscillating functions with supremum norm is a commutative unital C^*-algebra. Thus there is a compactification $h\mathbb{R}^d$ of $\mathbb{R}^d$, the *Higson compactification*, such that $C_h(\mathbb{R}^d)$ is isometrically isomorphic to $C(h\mathbb{R}^d)$. It is known that $h\mathbb{R}^d$ is nonmetrizable, and even more, points of the corona $h\mathbb{R}^d \setminus \mathbb{R}^d$ can only be reached by nets; see, e.g., [Rabinovich et al. 2004a, 2.4.10]. Therefore we need to work with nets instead of sequences.

The relevance of the Higson compactification and the algebra of slowly oscillating functions in our context is given by the fact that translations act trivially on the corona $\partial_h\mathbb{R}^d$.

Lemma 6.1. *If $x_\lambda \to \eta \in \partial_h\mathbb{R}^d$, then $x + x_\lambda \to \eta$ for all $x \in \mathbb{R}^d$.*

Proof. By definition, $x_\lambda \to \eta \in \partial_h\mathbb{R}^d$ if $f(x_\lambda) \to \bar{f}(\eta) = \eta(f)$ for every $f \in C_h(\mathbb{R}^d)$. From the definition of $C_h(\mathbb{R}^d)$ we obtain

$$\lim_\lambda |f(x_\lambda) - f(x + x_\lambda)| = 0$$

for every $x \in \mathbb{R}^d$, so $f(x_\lambda + x) \to \bar{f}(\eta)$ for every $f \in C_h(\mathbb{R}^d)$ as well. □

One can show that $h\mathbb{R}^d$ is the *maximal* compactification of $\mathbb{R}^d$ with this property: every $C_\gamma(\mathbb{R}^d)$ as above with translations acting trivially on $\partial_\gamma\mathbb{R}^d$ is a subalgebra of $C_h(\mathbb{R}^d)$.

We need the following fact [Roe 2003].

Proposition 6.2. *Let $C_\gamma(\mathbb{R}^d)$ be a C^*-algebra of functions on $\mathbb{R}^d$ as above with corresponding compactification $\gamma\mathbb{R}^d$ of $\mathbb{R}^d$. If $f \in C_\gamma(\mathbb{R}^d)$ satisfies*

$$\bar{f}|_{\partial_\gamma\mathbb{R}^d} \equiv 0$$

then $f \in C_0(\mathbb{R}^d)$.

Proof. Let $(x_\lambda)_{\lambda\in\Lambda} \subset \mathbb{R}^d$ be an unbounded net, $\lim_\lambda |x_\lambda| = \infty$. As $\gamma\mathbb{R}^d$ is compact, every subnet of (x_λ) has a convergent subnet $(x_\mu)_{\mu\in M}$, and $\lim_\mu x_\mu = \eta \in \partial_\gamma\mathbb{R}^d$ by the assumption of unboundedness. So $\lim_\mu f(x_\mu) = \bar{f}(\eta) = 0$ for a subnet of a given subnet, and therefore $\lim_\lambda f(x_\lambda) = 0$. This means $f \in C_0(\mathbb{R}^d)$. □

We next study uniformly elliptic operators H_a with a symbol in $a \in C_h^\infty(\mathbb{R}^d, \mathbb{C}^{d\times d})$. By definition a has a continuous extension to $h\mathbb{R}^d$. Thus for $\eta \in \partial_h\mathbb{R}^d$ the symbol $\bar{a}(\eta) = \lim_{x\to\eta} a(x)$ is well-defined, and by Lemma 2.13 $H_{\bar{a}(\eta)}$ is a differential operator with constant coefficients. Let k^η denote the reproducing kernel of $\mathrm{PW}_\Omega(H_{\bar{a}(\eta)})$. We show that the mapping $x \mapsto T_{-x}k_x$ has a continuous extension to $h\mathbb{R}^d$.

Proposition 6.3. *Let the symbol $a \in C_h^\infty(\mathbb{R}^d, \mathbb{C}^{d\times d})$ be the symbol of the operator H_a and let k_x be the reproducing kernel of $\mathrm{PW}_\Omega(H_a)$. Set $K(x) = T_{-x}k_x \in L^2(\mathbb{R}^d)$. Then K extends to a continuous function from $h\mathbb{R}^d$ to $L^2(\mathbb{R}^d)$ by setting $K(\eta) = k_0^\eta$ for $\eta \in h\mathbb{R}^d$. In particular, the diagonal $k(x,x) = \|k_x\|_2^2$ is a slowly oscillating function.*

Proof. By Proposition 2.3 the centered reproducing kernel K is continuous from $\mathbb{R}^d$ to $L^2(\mathbb{R}^d)$. To show that $K \in C_h(\mathbb{R}^d, L^2)$, we need to extend K to the Higson corona $\partial_h\mathbb{R}^d$.

This is accomplished by means of Theorem 4.2. Let $(x_\lambda) \subseteq \mathbb{R}^d$ be an unbounded net such that $x_\lambda \to \eta \in \partial_h\mathbb{R}^d$. Since $a \in C_h(\mathbb{R}^d, \mathbb{C}^{d\times d})$, there is a continuous function $\bar{a} \in C(h\mathbb{R}^d, \mathbb{C}^{d\times d})$ such that $\lim_\lambda a(x_\lambda) = \bar{a}(\eta)$. Furthermore, for $x \in \mathbb{R}^d$ arbitrary, $x + x_\lambda \to \eta$ by Lemma 6.1, and this fact implies the pointwise convergence

$$\lim_\lambda T_{-x_\lambda} a(x) = \bar{a}(\eta).$$

Clearly, the spectrum of the (constant-coefficient) operator $H_{\bar{a}(\eta)}$ is continuous and does not contain any eigenvalues. The assumptions of Theorem 4.2 are thus satisfied.

To formulate its conclusion, denote the reproducing kernel of $\mathrm{PW}_\Omega(H_{\bar{a}(\eta)})$ by k^η. Then by Theorem 4.2

$$\lim_\lambda T_{-x_\lambda} k_{x_\lambda} = k_0^\eta$$

in the L^2-norm, and this holds for every net (x_λ) with $x_\lambda \to \eta$. Thus we must take the limiting function to be

$$K(\eta) = k_0^\eta = (2\pi)^{-d/2}\mathscr{F}^{-1}(\chi_{\Sigma_\Omega^{\bar{a}(\eta)}}),$$

with the explicit formula for the kernel given by (6-2).

It remains to be shown that the limiting kernel K is continuous on $\partial_h\mathbb{R}^d$. Let $\eta_\lambda \to \eta \in \partial_h\mathbb{R}^d$. Then with the definition of Σ_Ω^b and (6-1) we obtain

$$\begin{aligned}\|k_0^{\eta_\lambda} - k_0^\eta\|_2^2 &= (2\pi)^{-d}\|\chi_{\Sigma_\Omega^{\bar{a}(\eta_\lambda)}} - \chi_{\Sigma_\Omega^{\bar{a}(\eta)}}\|_2^2 \\ &= (2\pi)^{-d}(|\Sigma_\Omega^{\bar{a}(\eta_\lambda)}| + |\Sigma_\Omega^{\bar{a}(\eta)}| - 2|\Sigma_\Omega^{\bar{a}(\eta_\lambda)} \cap \Sigma_\Omega^{\bar{a}(\eta)}|).\end{aligned}$$

As $a \in C_h(\mathbb{R}^d, \mathbb{C}^{d\times d})$, $\bar{a}$ is continuous on $\partial_h\mathbb{R}^d$, and this expression tends to 0, whence K is also continuous on the corona $\partial_h\mathbb{R}^d$. □

6C. *Geometric densities for slowly oscillating symbols.* In order to obtain values for the critical densities $D_\mu^\pm(S)$ we need the averaged traces

$$\mathrm{tr}_\mu^-(k) = \liminf_{r\to\infty} \inf_{x\in\mathbb{R}^d} \frac{1}{\mu(B_r(x))} \int_{B_r(x)} k(z,z)\,dz$$

and $\mathrm{tr}_\mu^+(k)$. For the comparison of averaged traces we will need the following well-known fact, whose proof is supplied in Appendix A for completeness.

For $f \in C_b(\mathbb{R}^d)$ set

$$\mathrm{tr}^-(f) = \liminf_{r\to\infty} \inf_{y\in\mathbb{R}^d} \frac{1}{|B_r(y)|} \int_{B_r(y)} f(x)\,dx,$$

and define $\mathrm{tr}^+(f)$ similarly with sup instead of inf.

Lemma 6.4. *Assume that $f, g \in C_b(\mathbb{R}^d)$ and $\lim_{x\to\infty} |f(x) - g(x)| = 0$. Then*

$$\mathrm{tr}^-(f) = \mathrm{tr}^-(g) \quad \textit{and} \quad \mathrm{tr}^+(f) = \mathrm{tr}^+(g).$$

Proposition 6.5. *If the symbol a is in $C_h^\infty(\mathbb{R}^d, \mathbb{C}^{d\times d})$, then the trace of the reproducing kernel satisfies*

$$\lim_{r\to\infty} \sup_{y\in\mathbb{R}^d} \left| \frac{1}{|B_r(y)|} \int_{B_r(y)} (k(x,x) - \frac{|B_1|}{(2\pi)^d} \frac{\Omega^{d/2}}{\det a(x)^{1/2}})\, dx \right| = 0. \tag{6-3}$$

Equivalently, using the Borel measure $\nu(B) = \int_B \det a(x)^{-1/2}\, dx$,

$$\lim_{r\to\infty} \sup_{y\in\mathbb{R}^d} \left| \frac{1}{\nu(B_r(y))} \int_{B_r(y)} k(x,x)\, dx - \frac{|B_1|}{(2\pi)^d} \Omega^{d/2} \right| = 0. \tag{6-4}$$

Consequently, the averaged trace is

$$\mathrm{tr}_\nu^+(k) = \mathrm{tr}_\nu^-(k) = \frac{|B_1|}{(2\pi)^d} \Omega^{d/2}. \tag{6-5}$$

Proof. We apply Lemma 6.4 to the functions $f(x) = k(x,x)$ and $g(x) = (|B_1|\Omega^{d/2}/(2\pi)^d) \det a(x)^{-1/2}$. Then $k(x,x)$ is bounded by Proposition 2.2 and continuous by Proposition 2.3. Likewise $\det a(x)^{-1/2}$ is bounded and continuous by elliptic regularity. By assumption on a and Proposition 6.3 both functions are in $C_h(\mathbb{R}^d)$ and thus possess the limits $\bar{a}(\eta)$ and $\|K(\eta)\|_2^2$; in particular for a this means that

$$\lim_{x_\lambda\to\eta} \det a(x_\lambda)^{-1/2} = \det \bar{a}(\eta)^{-1/2}.$$

Using the notation of Section 6A and Proposition 6.3, we obtain

$$\lim_{x_\lambda\to\eta} \|k_{x_\lambda}\|_2^2 = \lim_{x_\lambda\to\eta} \|T_{-x_\lambda} k_{x_\lambda}\|_2^2 = \|k_0^\eta\|_2^2 = |\Sigma_\Omega^{\bar{a}(\eta)}| = \frac{|B_1|\Omega^{d/2}}{(2\pi)^d} \det \bar{a}(\eta)^{-1/2}.$$

We conclude that both f and g have the same limit function, and therefore $f - g \in C_0(\mathbb{R}^d)$ by means of Proposition 6.2. Lemma 6.4 now yields (6-3). Equation (6-4) follows after multiplying with $|B_r(y)|/\nu(B_r(y))$ and taking limits. Finally, (6-5) is a direct consequence of (6-4). □

Equation (6-4) allows us to state our main result on geometric Beurling densities for operators with slowly oscillating symbols in a simple form. In order to do so we need an elementary result on the relation between density and a change of measure.

Lemma 6.6. *Let $d\mu = h\, dx$ for a positive, continuous function h on $\mathbb{R}^d$, bounded above and below, $0 < c \le h(z) \le C$ for all $z \in \mathbb{R}^d$. Then the dimension-free density condition*

$$D_0^-(S) = \liminf_{r\to\infty} \inf_{x\in\mathbb{R}^d} \frac{\#(S \cap B_r(x))}{\int_{B_r(x)} k(y,y)\, dy} \ge 1$$

holds, if and only if

$$D_\mu^-(S) \ge \mathrm{tr}_\mu^-(k).$$

Similarly

$$D_0^+(S) \le 1 \quad \textit{if and only if} \quad D_\mu^+(S) \le \mathrm{tr}_\mu^-(k).$$

Proof. The inequality $D_0^-(S) \geq 1$ means that for all $\varepsilon > 0$ there is an $r_\varepsilon > 0$ such that for all $r > r_\varepsilon$

$$\#(S \cap B_r(x)) \geq (1-\varepsilon) \int_{B_r(x)} k(y,y)\,dy,$$

or equivalently,

$$\frac{\#(S \cap B_r(x))}{\mu(B_r(x))} \geq (1-\varepsilon) \frac{1}{\mu(B_r(x))} \int_{B_r(x)} k(y,y)\,dy.$$

Written in terms of the Beurling density, this is

$$D_\mu^-(S) \geq \liminf_{r\to\infty} \inf_{x\in\mathbb{R}^d} \frac{\int_{B_r(x)} k(y,y)\,dy}{\mu(B_r(x))} = \mathrm{tr}_\mu^-(k).$$

The converse is obtained by reading the argument backwards. □

As a direct consequence we obtain the main result on geometric Beurling densities for uniformly elliptic operators with slowly oscillating symbols. This is Theorem C of the Introduction.

Theorem 6.7. *Assume that $H_a = -\sum_{j,k=1}^d \partial_j a_{jk} \partial_k$ is uniformly elliptic with symbol $a \in C_h^\infty(\mathbb{R}^d, \mathbb{C}^{d\times d})$. Let $\mathrm{PW}_\Omega(H_a) = \chi_{[0,\Omega]}(H_a) L^2(\mathbb{R}^d)$ be the corresponding Paley–Wiener space and set $d\nu(x) = \det(a(x))^{-1/2}\,dx$.*

- *If $S \subseteq \mathbb{R}^d$ is a set of stable sampling for $\mathrm{PW}_\Omega(H_a)$ then*

$$D_\nu^-(S) \geq \frac{|B_1|}{(2\pi)^d} \Omega^{d/2}.$$

- *If $S \subseteq \mathbb{R}^d$ is a set of interpolation for $\mathrm{PW}_\Omega(H_a)$, then*

$$D_\nu^+(S) \leq \frac{|B_1|}{(2\pi)^d} \Omega^{d/2}.$$

Proof. We only verify the first assertion. By Corollary 3.2, if S is a set of stable sampling, then $D_0^-(S) \geq 1$. By Lemma 6.6 this is equivalent to

$$D_\nu^-(S) \geq \mathrm{tr}_\nu^-(k).$$

The averaged trace $\mathrm{tr}_\nu^-(k)$ was computed in (6-5) to be $(|B_1|/(2\pi)^d)\Omega^{d/2}$. □

Example 6.8. We consider some special cases of Theorem 6.7.

(i) *Asymptotically constant symbols.* Assume that $a \in C_b^\infty(\mathbb{R}^d, \mathbb{C}^{d\times d})$ and $\lim_{x\to\infty} a(x) = b$. Then it is straightforward to verify that $a \in C_h^\infty(\mathbb{R}^d, \mathbb{C}^{d\times d})$ and $D_\nu^\pm(S) = (\det b)^{1/2} D^\pm(S)$. Thus we may use the original Beurling density, and Theorem 6.7 implies that a sampling set $S \subseteq \mathbb{R}^d$ for $\mathrm{PW}_\Omega(H_a)$ must have density

$$D^-(S) = (\det b)^{-1/2} D_\nu^-(S) \geq (\det b)^{-1/2} \frac{|B_1|}{(2\pi)^d} \Omega^{d/2} = \frac{|\Sigma_\Omega^b|}{(2\pi)^d},$$

and a set of interpolation S in $\mathrm{PW}_\Omega(H_a)$ must satisfy $D^+(S) \leq |\Sigma_\Omega^b|/(2\pi)^d$. This is Corollary D of the Introduction. As was to be expected, this coincides with the critical density for the classical Paley–Wiener space $\mathrm{PW}_\Omega(H_b)$; see [Landau 1967].

(ii) *Symbols with radial limits.* Let us consider the class of symbols that possess radial limits at ∞. We say that $a \in C_b^\infty(\mathbb{R}^d, \mathbb{C}^{d\times d})$ is spherically continuous, if it possesses radial limits in the following sense. There exists a continuous matrix function $b \in C(S^{d-1}, \mathbb{C}^{d\times d})$ such that

$$\lim_{r\to\infty} \sup_{\eta\in S^{d-1}} \|a(r\eta) - b(\eta)\|_X = 0.$$

A 3ε-argument shows that these symbols are slowly oscillating. Consequently Theorem 6.7 holds for spherically continuous symbols in $C_b^\infty(\mathbb{R}^d, \mathbb{C}^{d\times d})$. Spherically continuous symbols are related to another compactification, the spherical compactification with corona S^{d-1}. In contrast to the Higson compactification, it is metrizable, but it is much smaller. See [Cordes 1979] for its use in partial differential equations.

6D. *Variable bandwidth in dimension $d = 1$.* Let H_a be the differential operator

$$H_a f = -\frac{d}{dx}\Big(a\frac{d}{dx} f\Big)$$

on $L^2(\mathbb{R})$. This is a Sturm–Liouville operator on $\mathbb{R}$, and the ellipticity assumption amounts to the conditions $\inf_{x\in\mathbb{R}} a(x) > 0$ and $a \in C_b^\infty(\mathbb{R}^d)$. In [Gröchenig and Klotz 2017] we argued that the spectral subspaces of H_a can be interpreted as spaces of locally variable bandwidth. Intuitively, the quantity $a(x)^{-1/2}$ is a measure for the bandwidth in a neighborhood of x. We apply Theorem 6.7 to H_a. The relevant measure is $d\nu(x) = a^{-1/2}(x)\,dx$, and $\nu(I) = \int_I a(x)^{-1/2}\,dx$ for $I \subseteq \mathbb{R}$. Then we have the following necessary density condition for functions of variable bandwidth (Corollary E of the Introduction).

Corollary 6.9. *Assume that $a \in C_b^\infty(\mathbb{R})$ and $\lim_{x\to\pm\infty} a'(x) = 0$. Let $\mathrm{PW}_\Omega(H_a)$ be the Paley–Wiener space associated to H_a.*

(i) *If S is a sampling set for $\mathrm{PW}_\Omega(H_a)$, then*

$$D_\nu^-(S) = \liminf_{r\to\infty} \inf_{x\in\mathbb{R}} \frac{\#(S\cap[x-r, x+r])}{\nu([x-r, x+r])} \geq \frac{\Omega^{1/2}}{\pi}. \tag{6-6}$$

(ii) *If S is a set of interpolation for $\mathrm{PW}_\Omega(H_a)$, then $D_\nu^+(S) \leq \Omega^{1/2}/\pi$.*

Arguing as in Lemma 6.6, equation (6-6) says that for $\varepsilon > 0$ and r large enough we have

$$\#(S\cap[x-r, x+r]) \geq \left(\frac{\Omega^{1/2}}{\pi} - \varepsilon\right) \int_{x-r}^{x+r} a(y)^{-1/2}\,dy.$$

Thus the number of samples in an interval $[x-r, x+r]$ is determined by $a(x)^{-1/2}$, which is in line with our interpretation of $a^{-1/2}$ as the local bandwidth.

Corollary 6.9 is precisely the formulation of the necessary density conditions in [Gröchenig and Klotz 2017]. However, the main result of that work was proved under the restrictive assumption that a is constant outside an interval $[-R, R]$. The proof there dwelt heavily on the scattering theory of one-dimensional Schrödinger operators. The method of this paper yields a significantly more general result with a completely different method of proof. Corollary 6.9 was our dream that motivated this work.

Finally we remark that the density conditions of Theorem 6.7 suggest that the Paley–Wiener spaces $\mathrm{PW}_\Omega(H_a)$ associated to a uniformly elliptic differential operator may be taken as an appropriate generalization of variable bandwidth to higher dimensions.

7. Outlook

We have proved necessary density conditions for sampling and interpolation in spectral subspaces of uniformly elliptic partial differential operators with slowly oscillating coefficients. These spectral subspaces may be taken as a suitable generalization of the notion of variable bandwidth to higher dimensions. The emphasis has been on a new method that combines elements from limit operators, regularity theory and heat kernel estimates, and the use of compactifications.

Clearly one can envision manifold extensions of our results and methods. Theorem 3.1 is stated for a significantly larger class of operators and symbols. For instance, it could be applied to higher-order partial differential operators or to Schrödinger operators and to symbols with less smoothness or to almost periodic symbols. However, the spectral theory of such operators is more involved and one needs to find conditions that prevent their limit operators from having a point spectrum at the ends of the spectral interval. As these questions belong to spectral theory rather than sampling theory, we plan to pursue them in a separate publication.

In a different direction one may consider the graph Laplacian on an infinite graph or even a metric measure space endowed with a kernel that satisfies Gaussian estimates [Coulhon et al. 2012]. While many steps of our proofs remain in place, this set-up opens numerous new questions.

Finally several hidden connections beg to be explored. The identity (6-3) resembles the famous Weyl formula for the asymptotic density of eigenvalues in a spectral interval [Hörmander 1968]. This observation invites the comparison of the Beurling density with the density of states in spectral theory. We plan to investigate some of these issues in future work.

Appendix A: Averaged traces

For completeness we provide the proof of Lemma 6.4. Recall that

$$\mathrm{tr}^-(f) = \liminf_{r\to\infty} \inf_{y\in\mathbb{R}^d} \frac{1}{|B_r(y)|} \int_{B_r(y)} f(x)\,dx.$$

If $f, g \in C_b(\mathbb{R}^d)$ *and* $\lim_{|x|\to\infty} |f(x) - g(x)| = 0$, *then*

$$\mathrm{tr}^-(f) = \mathrm{tr}^-(g) \quad \text{and} \quad \mathrm{tr}^+(f) = \mathrm{tr}^+(g).$$

Proof. Set $h = f - g$, then $\lim_{x\to\infty} h(x) = 0$. We split the relevant averages as

$$\frac{1}{|B_r(y)|} \int_{B_r(y)} |h(x)|\,dx = \frac{1}{|B_r(y)|} \left[\int_{B_r(y)\cap B_R^c(0)} + \int_{B_r(y)\cap B_R(0)} \right] |h(x)|\,dx = (I) + (II).$$

Given $\varepsilon > 0$, there exists an $R_\varepsilon > 0$ such that $\sup_{|x|\geq R_\varepsilon} |h(x)| < \varepsilon/2$. So,

$$(I) < \frac{|B_r(y) \cap B^c_{R_\varepsilon}(0)|}{|B_r(y)|} \frac{\varepsilon}{2} < \frac{\varepsilon}{2},$$

independent of y. For the second term observe that

$$(II) < \|h\|_\infty \frac{|B_r(y) \cap B_{R_\varepsilon}(0)|}{|B_r(y)|} \leq \|h\|_\infty \frac{|B_{R_\varepsilon}|}{|B_r(y)|} < \frac{\varepsilon}{2}$$

for $r > \left(\frac{2}{\varepsilon}\|h\|_\infty\right)^{1/d} R_\varepsilon$. Consequently,

$$\lim_{r\to\infty} \sup_{y\in\mathbb{R}^d} \frac{1}{|B_r(y)|} \int_{B_r(y)} |f - g| = 0.$$

It follows that

$$\mathrm{tr}^-(f) \leq \liminf_{r\to\infty} \inf_{y\in\mathbb{R}^d} \frac{1}{|B_r(y)|} \int_{B_r(y)} g + \lim_{r\to\infty} \sup_{y\in\mathbb{R}^d} \frac{1}{|B_r(y)|} \int_{B_r(y)} |f - g| = \mathrm{tr}^-(g).$$

Interchanging f and g yields equality. The equality $\mathrm{tr}^+(f) = \mathrm{tr}^+(g)$ is proved in the same way. □

Appendix B: The lower bound for the reproducing kernel

We verify the lower bound for $\|k_x\|_2$ in the proof of Proposition 2.2. This fact is proved in [Coulhon et al. 2012, Lemma 3.19(a)]. For completeness we reproduce that proof with some necessary modifications and adjustments. The idea is to relate the reproducing kernel to the heat kernel of e^{-tH_a} via functional calculus.

We write k^Ω for the reproducing kernel of $\mathrm{PW}_\Omega\,(H_a)$. For a bounded, nonnegative Borel function $F \geq 0$ with support in $[0, \Omega]$ we define $k^F_x = F(H_a)k^\Omega_x$ and the corresponding integral kernel $k^F(x, y) := F(H_a)(x, y) := k^F_x(y)$. The last expression is well-defined, as $k^F_x \in \mathrm{PW}_\Omega(H_a)$. The kernel $k^F(x, y)$ is symmetric, because $F(H_a)$ is self-adjoint:

$$k^F_x(y) = \langle F(H_a)k^\Omega_x, k^\Omega_y\rangle = \langle k^\Omega_x, F(H_a)k^\Omega_y\rangle = \overline{\langle F(H_a)k^\Omega_y, k^\Omega_x\rangle} = \overline{k^F_y(x)}.$$

Consequently, $F(H_a)$ is an integral operator. For $f \in L^2(\mathbb{R}^d)$

$$F(H_a)f(x) = \langle F(H_a)f, k^\Omega_x\rangle = \langle f, F(H_a)k^\Omega_x\rangle = \langle f, k^F_x\rangle = \int_{\mathbb{R}^d} k^F(y, x)f(y)\,dy.$$

If $0 \leq G \leq F$, then

$$0 \leq k^G(x, x) \leq k^F(x, x) \tag{B-1}$$

for all $x \in \mathbb{R}^d$. For the proof observe that $F - G \geq 0$ implies that

$$k^F(x, x) = \langle F(H_a)k^\Omega_x, k^\Omega_x\rangle \geq \langle G(H_a)k^\Omega_x, k^\Omega_x\rangle = k^G(x, x).$$

The heat operator e^{-tH_a} is bounded and has a kernel $p_t(x, y)$ that satisfies *on diagonal estimates*. There are positive constants c, C such that for all $x \in \mathbb{R}^d$ and $t > 0$

$$ct^{-d/2} \le p_t(x, x) \le Ct^{-d/2}.$$

This is well known; see, e.g., [Ouhabaz 2006].

Claim. *We have* $0 < c < k^{\Omega}(x, x) < C$ *for all* $x \in \mathbb{R}^d$.

Proof. As $\chi_\Omega(u) \le e \cdot e^{-u/\Omega}\chi_{[0,\infty)}(u)$ and $\chi_{[0,\infty)}(H_a) = \mathrm{Id}$, we obtain

$$k^{\Omega}(x, x) = \chi_{[0,\Omega]}(H_a)(x, x) \le ee^{-\Omega^{-1}H_a}(x, x) \le C\Omega^{d/2}, \tag{B-2}$$

which gives an explicit upper bound for $\|k_x\|_2^2$ in Proposition 2.2.

For the proof of the lower bound, we use a dyadic decomposition:

$$\begin{aligned}\chi_{[0,T)}(u)e^{-tu} &\le \chi_{[0,\infty)}(u)e^{-tu} = \chi_{[0,\Omega]}(u)e^{-tu} + \sum_{k\ge 0}\chi_{(2^k\Omega,2^{k+1}\Omega]}(u)e^{-tu}\\ &\le \chi_{[0,\Omega]}(u) + \sum_{k\ge 0}\chi_{[0,2^{k+1}\Omega]}(u)e^{-t2^k\Omega}, \quad t > 0.\end{aligned}$$

One can verify that this inequality remains true as an operator inequality

$$\chi_{[0,T]}(H_a)e^{-tH_a} \le \chi_{[0,\Omega]}(H_a) + \sum_{k\ge 0}\chi_{[0,2^{k+1}\Omega]}(H_a)e^{-t2^k\Omega},$$

with strong convergence of the sum, and every term is an integral operator. By (B-1) and (B-2) the operator inequality can be transferred to a corresponding inequality of the diagonals of the integral kernel as follows:

$$\begin{aligned}(\chi_{[0,T]}(H_a)e^{-tH_a})(x, x) &\le \chi_{[0,\Omega]}(H_a)(x, x) + \sum_{k\ge 0}\chi_{[0,2^{k+1}\Omega]}(H_a)(x, x)e^{-t2^k\Omega}\\ &\le \chi_{[0,\Omega]}(H_a)(x, x) + C\Omega^{d/2}\sum_{k\ge 0}2^{(k+1)d/2}\,e^{-t2^k\Omega}.\end{aligned}$$

In [Coulhon et al. 2012, equation (3.46)] it is shown that $p_t(x, y) = \lim_{T\to\infty}(\chi_{[0,T]}(H_a)e^{-tH_a})(x, y)$, consequently

$$ct^{-d/2} \le p_t(x, x) \le \chi_{[0,\Omega]}(H_a)(x, x) + C\Omega^{d/2}\sum_{k\ge 0}2^{(k+1)d/2}\,e^{-t2^k\Omega}.$$

We choose $t = 2^r/\Omega$ for $r \in \mathbb{N}$ to be specified later. Then

$$\begin{aligned}c\Omega^{d/2}2^{-rd/2} &\le \chi_{[0,\Omega]}(H_a)(x, x) + C\Omega^{d/2}\sum_{k\ge 0}e^{-2^k2^r}2^{(k+1)d/2}\\ &= \chi_{[0,\Omega]}(H_a)(x, x) + C2^{d/2}\Omega^{d/2}2^{-rd/2}\sum_{k\ge 0}e^{-2^{k+r}}2^{(k+r)d/2}\\ &\le \chi_{[0,\Omega]}(H_a)(x, x) + C2^{d/2}\Omega^{d/2}2^{-rd/2}\sum_{k\ge r}e^{-2^k}2^{kd/2}.\end{aligned}$$

Hence,

$$\Omega^{d/2}2^{-rd/2}\left(c - C'2^{d/2}\sum_{k\geq r} e^{-2^k}2^{kd/2}\right) \leq \chi_{[0,\Omega]}(H_a)(x,x) = k^{\Omega}(x,x).$$

For $r \in \mathbb{N}$ sufficiently large, this implies the lower bound for $k^{\Omega}(x,x)$. □

References

[Agmon 1965] S. Agmon, *Lectures on elliptic boundary value problems*, Van Nostrand Math. Stud. **2**, Van Nostrand, Princeton, NJ, 1965. MR Zbl

[Aronszajn 1950] N. Aronszajn, "Theory of reproducing kernels", *Trans. Amer. Math. Soc.* **68** (1950), 337–404. MR Zbl

[Beurling 1966] A. Beurling, "Local harmonic analysis with some applications to differential operators", pp. 109–125 in *Some recent advances in the basic sciences*, *I* (New York, 1962–1964), edited by A. Gelbart, Yeshiva Univ., New York, 1966. MR

[Beurling 1989] A. Beurling, *The collected works of Arne Beurling*, *II*: *Harmonic analysis*, Birkhäuser, Boston, 1989. MR Zbl

[Cordes 1979] H. O. Cordes, *Elliptic pseudo-differential operators: an abstract theory*, Lecture Notes in Math. **756**, Springer, 1979. MR Zbl

[Coulhon et al. 2012] T. Coulhon, G. Kerkyacharian, and P. Petrushev, "Heat kernel generated frames in the setting of Dirichlet spaces", *J. Fourier Anal. Appl.* **18**:5 (2012), 995–1066. MR Zbl

[Davies and Georgescu 2013] E. B. Davies and V. Georgescu, "C^*-algebras associated with some second order differential operators", *J. Operator Theory* **70**:2 (2013), 437–450. MR Zbl

[DLMF 2020] F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain (editors), "NIST digital library of mathematical functions", electronic reference, Nat. Inst. Standards Tech., 2020, available at http://dlmf.nist.gov. Release 1.0.28.

[Engelking 1977] R. Engelking, *General topology*, Monografie Mat. **60**, PWN, Warsaw, 1977. MR Zbl

[Feichtinger 1984] H. G. Feichtinger, "Compactness in translation invariant Banach spaces of distributions and compact multipliers", *J. Math. Anal. Appl.* **102**:2 (1984), 289–327. MR Zbl

[Feichtinger and Pesenson 2004] H. Feichtinger and I. Pesenson, "Recovery of band-limited functions on manifolds by an iterative algorithm", pp. 137–152 in *Wavelets*, *frames and operator theory* (College Park, MD, 2003), edited by C. Heil et al., Contemp. Math. **345**, Amer. Math. Soc., Providence, RI, 2004. MR Zbl

[Feichtinger et al. 2016] H. G. Feichtinger, H. Führ, and I. Z. Pesenson, "Geometric space-frequency analysis on manifolds", *J. Fourier Anal. Appl.* **22**:6 (2016), 1294–1355. MR Zbl

[Filbir and Mhaskar 2011] F. Filbir and H. N. Mhaskar, "Marcinkiewicz–Zygmund measures on manifolds", *J. Complexity* **27**:6 (2011), 568–596. MR Zbl

[Führ et al. 2017] H. Führ, K. Gröchenig, A. Haimi, A. Klotz, and J. L. Romero, "Density of sampling and interpolation in reproducing kernel Hilbert spaces", *J. Lond. Math. Soc.* (2) **96**:3 (2017), 663–686. MR Zbl

[Gamelin 1969] T. W. Gamelin, *Uniform algebras*, Prentice-Hall, Englewood Cliffs, NJ, 1969. MR Zbl

[Georgescu 2011] V. Georgescu, "On the structure of the essential spectrum of elliptic operators on metric spaces", *J. Funct. Anal.* **260**:6 (2011), 1734–1765. MR Zbl

[Georgescu 2018] V. Georgescu, "On the essential spectrum of elliptic differential operators", *J. Math. Anal. Appl.* **468**:2 (2018), 839–864. MR Zbl

[Grafakos 2004] L. Grafakos, *Classical and modern Fourier analysis*, Pearson, Upper Saddle River, NJ, 2004. MR Zbl

[Gröchenig and Klotz 2010] K. Gröchenig and A. Klotz, "Noncommutative approximation: inverse-closed subalgebras and off-diagonal decay of matrices", *Constr. Approx.* **32**:3 (2010), 429–466. MR Zbl

[Gröchenig and Klotz 2017] K. Gröchenig and A. Klotz, "What is variable bandwidth?", *Comm. Pure Appl. Math.* **70**:11 (2017), 2039–2083. MR Zbl

[Gröchenig et al. 2019] K. Gröchenig, A. Haimi, J. Ortega-Cerdà, and J. L. Romero, "Strict density inequalities for sampling and interpolation in weighted spaces of holomorphic functions", *J. Funct. Anal.* **277**:12 (2019), art. id. 108282. MR Zbl

[Heil 2011] C. Heil, *A basis theory primer*, expanded ed., Birkhäuser, New York, 2011. MR Zbl

[Hörmander 1968] L. Hörmander, "The spectral function of an elliptic operator", *Acta Math.* **121** (1968), 193–218. MR Zbl

[Kahane 1962] J.-P. Kahane, "Pseudo-périodicité et séries de Fourier lacunaires", *Ann. Sci. École Norm. Sup.* (3) **79** (1962), 93–150. MR Zbl

[Landau 1967] H. J. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions", *Acta Math.* **117** (1967), 37–52. MR

[Matei and Meyer 2010] B. Matei and Y. Meyer, "Simple quasicrystals are sets of stable sampling", *Complex Var. Elliptic Equ.* **55**:8-10 (2010), 947–964. MR Zbl

[Olevskiĭ and Ulanovskii 2008] A. Olevskiĭ and A. Ulanovskii, "Universal sampling and interpolation of band-limited signals", *Geom. Funct. Anal.* **18**:3 (2008), 1029–1052. MR Zbl

[Ortega-Cerdà and Pridhnani 2012] J. Ortega-Cerdà and B. Pridhnani, "Beurling–Landau's density on compact manifolds", *J. Funct. Anal.* **263**:7 (2012), 2102–2140. MR Zbl

[Ouhabaz 2006] E. M. Ouhabaz, "Sharp Gaussian bounds and L^p-growth of semigroups associated with elliptic and Schrödinger operators", *Proc. Amer. Math. Soc.* **134**:12 (2006), 3567–3575. MR Zbl

[Pesenson 1998] I. Pesenson, "Sampling of Paley–Wiener functions on stratified groups", *J. Fourier Anal. Appl.* **4**:3 (1998), 271–281. MR Zbl

[Pesenson 1999] I. Pesenson, "A reconstruction formula for band limited functions in $L_2(\mathbb{R}^d)$", *Proc. Amer. Math. Soc.* **127**:12 (1999), 3593–3600. MR Zbl

[Pesenson 2000] I. Pesenson, "A sampling theorem on homogeneous manifolds", *Trans. Amer. Math. Soc.* **352**:9 (2000), 4257–4269. MR Zbl

[Pesenson 2001] I. Pesenson, "Sampling of band-limited vectors", *J. Fourier Anal. Appl.* **7**:1 (2001), 93–100. MR Zbl

[Pesenson and Zayed 2009] I. Pesenson and A. I. Zayed, "Paley–Wiener subspace of vectors in a Hilbert space with applications to integral transforms", *J. Math. Anal. Appl.* **353**:2 (2009), 566–582. MR Zbl

[Rabinovich et al. 2004a] V. Rabinovich, S. Roch, and B. Silbermann, *Limit operators and their applications in operator theory*, Oper. Theory Adv. Appl. **150**, Birkhäuser, Basel, 2004. MR Zbl

[Rabinovich et al. 2004b] V. S. Rabinovich, S. Roch, and J. Roe, "Fredholm indices of band-dominated operators", *Integral Equations Operator Theory* **49**:2 (2004), 221–238. MR Zbl

[Roe 2003] J. Roe, *Lectures on coarse geometry*, Univ. Lect. Ser. **31**, Amer. Math. Soc., Providence, RI, 2003. MR Zbl

[Rudin 1973] W. Rudin, *Functional analysis*, McGraw-Hill, New York, 1973. MR Zbl

[Seip 2004] K. Seip, *Interpolation and sampling in spaces of analytic functions*, Univ. Lect. Ser. **33**, Amer. Math. Soc., Providence, RI, 2004. MR Zbl

[Shannon 1948] C. E. Shannon, "A mathematical theory of communication", *Bell System Tech. J.* **27** (1948), 379–423, 623–656. MR Zbl

[Shteinberg 2000] B. Y. Shteinberg, "Compactification of a locally compact group and the Noethericity of convolution operators with coefficients on quotient groups", pp. 209–223 in *Proceedings of the St. Petersburg Mathematical Society*, *VI*, edited by N. N. Uraltseva, Amer. Math. Soc. Transl. Ser. 2 **199**, Amer. Math. Soc., Providence, RI, 2000. MR Zbl

[Shubin 1992] M. A. Shubin, "Spectral theory of elliptic operators on noncompact manifolds", pp. 35–108 in *Méthodes semi-classiques*, *I* (Nantes, France, 1991), Astérisque **207**, Soc. Math. France, Paris, 1992. MR Zbl

[Špakula and Willett 2017] J. Špakula and R. Willett, "A metric approach to limit operators", *Trans. Amer. Math. Soc.* **369**:1 (2017), 263–308. MR Zbl

[Teschl 2009] G. Teschl, *Mathematical methods in quantum mechanics: with applications to Schrödinger operators*, Grad. Stud. Math. **99**, Amer. Math. Soc., Providence, RI, 2009. MR Zbl

[Wendland 2005] H. Wendland, *Scattered data approximation*, Cambridge Monogr. Appl. Computat. Math. **17**, Cambridge Univ. Press, 2005. MR Zbl

[Zimmer 1990] R. J. Zimmer, *Essential results of functional analysis*, Univ. Chicago Press, 1990. MR Zbl

Received 20 Sep 2021. Revised 7 Jun 2022. Accepted 11 Jul 2022.

KARLHEINZ GRÖCHENIG: karlheinz.groechenig@univie.ac.at
Faculty of Mathematics, University of Vienna, Vienna, Austria

ANDREAS KLOTZ: andreas.klotz@univie.ac.at
Faculty of Mathematics, University of Vienna, Vienna, Austria

ANALYSIS AND PDE
Vol. 17 (2024), No. 2, pp. 617–680
DOI: 10.2140/apde.2024.17.617

ON BLOWUP FOR THE SUPERCRITICAL QUADRATIC WAVE EQUATION

ELEK CSOBO, IRFAN GLOGIĆ AND BIRGIT SCHÖRKHUBER

We study singularity formation for the quadratic wave equation in the energy supercritical case, i.e., for $d \geq 7$. We find in closed form a new, nontrivial, radial, self-similar blow-up solution u^* which exists for all $d \geq 7$. For $d = 9$, we study the stability of u^* without any symmetry assumptions on the initial data and show that there is a family of perturbations which lead to blowup via u^*. In similarity coordinates, this family represents a codimension-1 Lipschitz manifold modulo translation symmetries. The stability analysis relies on delicate spectral analysis for a non-self-adjoint operator. In addition, in $d = 7$ and $d = 9$, we prove nonradial stability of the well-known ODE blow-up solution. Also, for the first time we establish persistence of regularity for the wave equation in similarity coordinates.

1. Introduction

In this paper, we are concerned with the quadratic wave equation

$$(\partial_t^2 - \Delta_x)u(t,x) = u(t,x)^2, \tag{1-1}$$

where $(t,x) \in I \times \mathbb{R}^d$, for some interval $I \subset \mathbb{R}$ containing zero.

It is well known that in all space dimensions (1-1) admits solutions that blow up in finite time, starting from smooth and compactly supported initial data. This follows from a classical result by Levine [1974], which provides an open set of such initial data. However, Levine's argument is indirect, and therefore

Glogić is supported by the Austrian Science Fund FWF Projects P 34378 and P 30076. This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project-ID 258734477 and SFB 1173.
MSC2020: 35B44.
Keywords: nonlinear wave equation, supercritical, blowup, singularity formation, self-similar solution, stability.

does not give insight into the blow-up profile. A more concrete example can be produced using the well-known *ODE solution*

$$u_T^{\mathrm{ODE}}(t,x) := \frac{6}{(T-t)^2}, \quad T > 0. \tag{1-2}$$

By truncating the initial data $(u_T^{\mathrm{ODE}}(0,\cdot), \partial_t u_T^{\mathrm{ODE}}(0,\cdot))$ outside a ball of radius larger than T and using finite speed of propagation, one constructs smooth and compactly supported initial data that lead to blowup at $t = T$. What is more, invariance of (1-1) under the rescaling

$$u(t,x) \mapsto u_\lambda(t,x) := \lambda^{-2} u\left(\frac{t}{\lambda}, \frac{x}{\lambda}\right), \quad \lambda > 0, \tag{1-3}$$

allows one to look for *self-similar* blow-up solutions of the form

$$u(t,x) = \frac{1}{(T-t)^2}\phi\left(\frac{x}{T-t}\right).$$

Note that (1-2) is a self-similar solution with trivial profile $\phi \equiv 6$. We note that the rescaling (1-3) leaves invariant the energy norm $\dot{H}^1(\mathbb{R}^d) \times L^2(\mathbb{R}^d)$ of $(u(t,\cdot), \partial_t u(t,\cdot))$ precisely when $d = 6$, in which case (1-1) is referred to as *energy critical*. In this case, it can be easily shown that in addition to (1-2) no other radial and smooth self-similar solutions to (1-1) exist; see [Kavian and Weissler 1990]. However, in the *energy supercritical* case, i.e., for $d \geq 7$, numerics [Kycia 2011] indicate that in addition to (1-2) there are nontrivial, radial, globally defined, smooth, and decaying similarity profiles. In fact, for $d = 7$, there are infinitely many of them, all of which are positive, as proven by Dai and Duyckaerts [2021]. A similar result is expected to hold for all $7 \leq d \leq 15$; see [Kycia 2011].

From the point of view of the Cauchy problem for (1-1), the relevant similarity profiles appear to be the trivial one (1-2) and its first nontrivial "excitation". Namely, numerical work on supercritical power nonlinearity wave equations in the radial case [Bizoń et al. 2004; Glogić et al. 2020] yields evidence that generic blowup is described by the ODE profile, while the threshold separating generic blowup from global existence is given by the stable manifold of the first excited profile; see also [Bizoń 2001]. The first step in showing such genericity results would be to establish stability of the ODE profile and show that its first excitation is codimension-1 stable (which indicates that the stable manifold splits the phase space locally into two connected components). The only result so far for (1-1) in this direction is by Donninger and the third author [Donninger and Schörkhuber 2017], who proved radial stability of u_T for all odd $d \geq 7$. In this paper, we exhibit in closed form what appears to be the first excitation of (1-2) for every $d \geq 7$. Namely, we have the following self-similar solution to (1-1):

$$u^*(t,x) := \frac{1}{t^2} U\left(\frac{|x|}{t}\right), \tag{1-4}$$

where

$$U(\rho) = \frac{c_1 - c_2\rho^2}{(c_3 + \rho^2)^2}, \tag{1-5}$$

with

$$c_1 = \tfrac{4}{25}((3d-8)d_0 + 8d^2 - 56d + 48), \quad c_2 = \tfrac{4}{5}d_0, \quad c_3 = \tfrac{1}{15}(3d - 18 + d_0),$$

and $d_0 = \sqrt{6(d-1)(d-6)}$. We note that $c_3 > 0$ when $d \geq 7$, and thus $U \in C^\infty[0,\infty)$. To the best of our knowledge, this solution has not been known before, and with the intent of studying threshold behavior, the main object of this paper is to show a variant of codimension-1 stability of u^*.

Note that U has precisely one zero at $\rho^* = \rho^*(d) > 2$. In particular, this profile is not positive and therefore not a member of the family of self-similar profiles constructed in [Dai and Duyckaerts 2021]. However, it is strictly positive inside the backward light cone of the blow-up point $(0,0)$. Hence, in this local sense u^* provides a solution to the more frequently studied focusing equation

$$(\partial_t^2 - \Delta_x)u(t,x) = |u(t,x)|u(t,x). \tag{1-6}$$

What is more, as an outcome of our stability analysis we get that small perturbations of both the ODE profile and u^* stay positive under the evolution of (1-1) and therefore yield solutions to (1-6) as well.

1A. *Main results.*

Preliminaries. By action of symmetries, the solution (1-4) gives rise to a $(2d+1)$-parameter family of (in general nonradial) blow-up solutions. Namely, (1-1) is invariant under spacetime translations

$$S_{T,x_0}(t,x) := (t-T, x-x_0)$$

for $T > 0$, $x_0 \in \mathbb{R}^d$, time reflections

$$R(t,x) := (-t,x),$$

as well as Lorentz boosts, which we write in terms of hyperbolic rotations as

$$\Lambda(a) := \Lambda^d(a^d) \circ \Lambda^{d-1}(a^{d-1}) \circ \cdots \circ \Lambda^1(a^1),$$

where $a \in \mathbb{R}^d$ and $\Lambda^j(a^j)$ for $j = 1,\dots,d$ are given by

$$\begin{cases} t \mapsto t\cosh(a^j) + x^j \sinh(a^j), \\ x^j \mapsto t\sinh(a^j) + x^j\cosh(a^j), \\ x^k \mapsto x^k \quad (k \neq j). \end{cases}$$

We then let

$$\Lambda_{T,x_0}(a) := R \circ \Lambda(a) \circ S_{T,x_0}, \tag{1-7}$$

and thereby obtain the following $(2d+1)$-parameter family of solutions to (1-1):

$$u^*_{T,x_0,a}(t,x) := u^* \circ \Lambda_{T,x_0}(a)(t,x).$$

We note that, for

$$(t',x') := \Lambda_{T,x_0}(a)(t,x),$$

we have

$$|x'|^2 - t'^2 = |x-x_0|^2 - (T-t)^2. \tag{1-8}$$

Furthermore, for $\xi, a \in \mathbb{R}^d$, we set[1]

$$\gamma(\xi,a) := A_0(a) - A_j(a)\xi^j, \tag{1-9}$$

[1] For simplicity, we use Einstein's summation convention throughout the paper.

where

$$\begin{aligned} A_0(a) &:= \cosh(a^1)\cosh(a^2)\cdots\cosh(a^d),\\ A_1(a) &:= \sinh(a^1)\cosh(a^2)\cdots\cosh(a^d),\\ A_2(a) &:= \sinh(a^2)\cosh(a^3)\cdots\cosh(a^d),\\ &\vdots\\ A_d(a) &:= \sinh(a^d). \end{aligned}$$

Then, it is easy to check that

$$t' = (T-t)\gamma\left(\frac{x-x_0}{T-t}, a\right) \quad \text{and} \quad x'^j = (t-T)\partial_{a^j}\gamma\left(\frac{x-x_0}{T-t}, a\right)B^j(a) \tag{1-10}$$

for $j = 1, \dots, d$, where

$$B^j(a) = \prod_{i=j+1}^{d} \cosh(a^i)^{-1}.$$

Now, by using relations (1-8) and (1-10) we find more explicitly that

$$u^*_{T,x_0,a}(t,x) = \frac{1}{(T-t)^2}U_a\left(\frac{x-x_0}{T-t}\right), \tag{1-11}$$

with $U_a : \mathbb{R}^d \to \mathbb{R}$ given by

$$U_a(\xi) = \frac{(c_1-c_2)\gamma(\xi,a)^2 + c_2(1-|\xi|^2)}{((1+c_3)\gamma(\xi,a)^2 + |\xi|^2 - 1)^2}. \tag{1-12}$$

Note that for $a = 0$, we have $U_0(\xi) = U(|\xi|)$ with U being the radial profile in (1-5). Also, since $c_1 > c_2$ for all $d \geq 7$, there exists a positive constant $c_0 = c_0(d)$ such that

$$U_a \geq c_0 > 0 \quad \text{on } \mathbb{B}^d \tag{1-13}$$

for all $a \in \mathbb{R}^d$, where $\mathbb{B}^d$ denotes the open unit ball in $\mathbb{R}^d$. In summary, we have that, for $a \in \mathbb{R}^d$, $x_0 \in \mathbb{R}^d$, and $T > 0$, (1-1) admits an explicit solution (1-11), which starts off smooth, blows up at $x = x_0$ as $t \to T^-$, and is strictly positive on the backward light cone

$$\mathcal{C}_{T,x_0} := \bigcup_{t\in[0,T)} \{t\} \times \overline{\mathbb{B}}^d_{T-t}(x_0)$$

of the blow-up point (T, x_0) — see Section 1C for the notation — which makes it a solution inside $\mathcal{C}_{T,x_0}$ to (1-6) as well. Furthermore, simply by scaling we have that, for $k \in \mathbb{N}_0$,

$$\left\| U_a\left(\frac{\cdot - x_0}{T-t}\right)\right\|_{\dot{H}^k(\mathbb{B}^d_{T-t}(x_0))} \simeq (T-t)^{\frac{d}{2}-k}, \tag{1-14}$$

and hence

$$\|u^*_{T,x_0,a}(t,\cdot)\|_{\dot{H}^k(\mathbb{B}^d_{T-t}(x_0))} \simeq (T-t)^{\frac{d}{2}-2-k},$$

which implies that the solution blows up in local homogeneous Sobolev seminorms of order $k > s_c = \frac{1}{2}d - 2$. Here, s_c denotes the critical regularity, i.e., $\dot{H}^{s_c}(\mathbb{R}^d)$ is left-invariant under the rescaling (1-3).

Conditional stability of blowup via u^*. The main goal of this paper is to investigate stability of blowup governed by u^*. For $T = 1$, $x_0 = 0$, and $a = 0$, the blow-up initial data are given by

$$u^*_{1,0,0}(0,x) = U(|x|) \quad \text{and} \quad \partial_t u^*_{1,0,0}(0,x) = 2U(|x|) + |x|U'(|x|).$$

We can now formulate the following stability result, where we restrict ourselves to the case $d = 9$.

Theorem 1.1. *Let $d = 9$. Define functions $h_j : \mathbb{R}^9 \to \mathbb{R}$, $j = 1, 2$, by*

$$h_1(x) = \frac{1}{(7+5|x|^2)^3} \quad \text{and} \quad h_2(x) = \frac{35-5|x|^2}{(7+5|x|^2)^4}. \tag{1-15}$$

There exist constants $M > 0$, $\delta > 0$, and $\omega > 0$ such that, for all real-valued $(f, g) \in C^\infty(\overline{\mathbb{B}}^9_2) \times C^\infty(\overline{\mathbb{B}}^9_2)$ satisfying

$$\|(f,g)\|_{H^6(\mathbb{B}^9_2)\times H^5(\mathbb{B}^9_2)} \le \frac{\delta}{M},$$

the following holds: There are parameters $a \in \overline{\mathbb{B}}^9_{M\delta/\omega}$, $x_0 \in \overline{\mathbb{B}}^9_\delta$, $T \in [1-\delta, 1+\delta]$, and $\alpha \in [-\delta, \delta]$ depending Lipschitz-continuously on (f, g) such that, for initial data

$$u(0,\cdot) = U(|\cdot|) + f + \alpha h_1 \quad \text{and} \quad \partial_t u(0,\cdot) = 2U(|\cdot|) + |\cdot|U'(|\cdot|) + g + \alpha h_2, \tag{1-16}$$

there exists a unique solution $u \in C^\infty(\mathcal{C}_{T,x_0})$ to (1-1). Furthermore, this solution blows up at (T, x_0) and can be written as

$$u(t,x) = \frac{1}{(T-t)^2}\left[U_a\left(\frac{x-x_0}{T-t}\right) + \varphi(t,x)\right],$$

where $\|\varphi(t,\cdot)\|_{L^\infty(\mathbb{B}^9_{T-t}(x_0))} \lesssim (T-t)^\omega$ and

$$(T-t)^{k-\frac{9}{2}}\|\varphi(t,\cdot)\|_{\dot{H}^k(\mathbb{B}^9_{T-t}(x_0))} \lesssim (T-t)^\omega$$

for $k = 0, \dots, 5$. In particular,

$$\begin{aligned} (T-t)^{k-\frac{5}{2}}\|u(t,\cdot) - u^*_{T,x_0,a}(t,\cdot)\|_{\dot{H}^k(\mathbb{B}^9_{T-t}(x_0))} &\lesssim (T-t)^\omega, \\ (T-t)^{k-\frac{5}{2}}\|\partial_t u(t,\cdot) - \partial_t u^*_{T,x_0,a}(t,\cdot)\|_{\dot{H}^{k-1}(\mathbb{B}^9_{T-t}(x_0))} &\lesssim (T-t)^\omega \end{aligned} \tag{1-17}$$

for $k = 1, \dots, 5$. Moreover, u is strictly positive on $\mathcal{C}_{T,x_0}$, and hence the statement above applies to (1-6) as well.

We note that the normalizing factor on the left-hand side of (1-17) appears naturally and corresponds to the behavior of the blow-up solution we perturbed around; see (1-14).

Some further remarks on the result are in order.

Remark 1.2. The proof of Theorem 1.1 relies on stability analysis in similarity coordinates, in which the above set of perturbations has a codimension-1 interpretation. More precisely, we construct a Lipschitz manifold which is of codimension 11, where ten codimensions are related to instabilities caused by translation symmetries of the equation and the remaining codimension is characterized by (h_1, h_2). This is elaborated on in Section 2; see in particular Propositions 2.1 and 2.4. We believe that this manifold

gives rise to a proper codimension-1 manifold in a suitable physical data space. However, by the local nature of our approach and the presence of translation symmetries, this is not entirely clear.

Remark 1.3 (Regularity of the initial data). It is only the transformation from similarity coordinates to physical coordinates that induces the higher-regularity assumption on the data, from which we can easily deduce the Lipschitz-dependence on the blow-up parameters. We nonetheless believe that this can be optimized by a more refined analysis.

Remark 1.4 (Persistence of regularity). While persistence of regularity is standard for the wave equation in physical coordinates, it has not yet been considered for the local problem in similarity coordinates. In fact, all of the related works so far, such as [Chatzikaleas and Donninger 2019; Donninger and Schörkhuber 2016; Glogić and Schörkhuber 2021], are based on a notion of strong solutions in similarity coordinates. In this paper, we close this gap and rigorously prove regularity of solutions for smooth initial data. Our proof relies on estimates for the free wave evolution in similarity coordinates in arbitrarily high Sobolev spaces; see Proposition 3.1 on page 629.

Remark 1.5 (Generalization to other space dimensions). Large parts of the proof of Theorem 1.1 can be generalized to other odd space dimensions. However, the analysis of the underlying spectral problem is quite delicate and only for $d = 9$ we are able to solve it rigorously. Nevertheless, from numerical computations, we have strong evidence that the situation is analogous in other space dimensions in the sense that the linearization has exactly one *genuine* unstable eigenvalue.

Stable ODE blowup without symmetry. For both (1-1) and (1-6), stability of the ODE blow-up solution under small radial perturbations has been proven by Donninger and the third author [Donninger and Schörkhuber 2017] in all odd space dimensions $d \geq 7$. By exploiting the framework of the proof of Theorem 1.1, we generalize the result from that paper to nonradial perturbations in dimensions $d = 7$ and $d = 9$.

Before we state the result, we apply the symmetry transformations (1-7) to the ODE profile (1-2) to obtain the following family of blow-up solutions to both (1-1) and (1-6):

$$u^{\mathrm{ODE}}_{T,x_0,a}(t,x) := \frac{1}{(T-t)^2}\kappa_a\left(\frac{x-x_0}{T-t}\right), \tag{1-18}$$

where

$$\kappa_a(\xi) = 6\gamma(\xi,a)^{-2}. \tag{1-19}$$

To shorten the notation, we write $\mathcal{C}_T := \mathcal{C}_{T,0}$ for the backward light cone with vertex $(T,0)$.

Theorem 1.6. *Let $d \in \{7,9\}$. There are constants $C > 0$, $\delta > 0$, and $\omega > 0$ such that, for any real-valued $(f,g) \in C^\infty(\overline{\mathbb{B}}^d_2) \times C^\infty(\overline{\mathbb{B}}^d_2)$ satisfying*

$$\|(f,g)\|_{H^{(d+3)/2}(\mathbb{B}^d_2)\times H^{(d+1)/2}(\mathbb{B}^d_2)} \leq \frac{\delta}{C}, \tag{1-20}$$

the following holds: There exist parameters $a \in \overline{\mathbb{B}}^d_{C\delta/\omega}$ and $T \in [1-\delta, 1+\delta]$ depending Lipschitz continuously on (f,g) such that, for initial data

$$u(0,\cdot) = 6 + f \quad \text{and} \quad \partial_t u(0,\cdot) = 12 + g,$$

there exists a unique solution $u \in C^\infty(\mathcal{C}_T)$ *to* (1-1). *This solution blows up at* $(T, 0)$ *and can be written as*

$$u(t,x) = \frac{1}{(T-t)^2}\left[\kappa_a\left(\frac{x}{T-t}\right) + \varphi(t,x)\right],$$

where φ *satisfies* $\|\varphi(t,\cdot)\|_{L^\infty(\mathbb{B}^d_{T-t})} \lesssim (T-t)^\omega$ *and*

$$(T-t)^{k-\frac{d}{2}}\|\varphi(t,\cdot)\|_{\dot{H}^k(\mathbb{B}^d_{T-t})} \lesssim (T-t)^\omega$$

for $k = 0, \dots, \frac{1}{2}(d+1)$. *In particular,*

$$\begin{aligned}(T-t)^{k-\frac{d}{2}+2}\|u(t,\cdot) - u^{\mathrm{ODE}}_{T,0,a}(t,\cdot)\|_{\dot{H}^k(\mathbb{B}^d_{T-t})} &\lesssim (T-t)^\omega,\\ (T-t)^{k-\frac{d}{2}+2}\|\partial_t u(t,\cdot) - \partial_t u^{\mathrm{ODE}}_{T,0,a}(t,\cdot)\|_{\dot{H}^{k-1}(\mathbb{B}^d_{T-t})} &\lesssim (T-t)^\omega\end{aligned} \tag{1-21}$$

for $k = 1, \dots, \frac{1}{2}(d+1)$. *Furthermore,* u *is strictly positive and the statement above therefore applies to* (1-6) *as well.*

We note that due to the invariance of $u^{\mathrm{ODE}}_{1,0,0}$ under spatial translations the blow-up location $x_0 = 0$ does not change under small perturbations.

Remark 1.7. Stability of the ODE blow-up solution for energy supercritical wave equations outside radial symmetry was established in $d = 3$ by Donninger and the third author [Donninger and Schörkhuber 2016]. For the cubic wave equation, the corresponding result was obtained by Chatzikaleas and Donninger [2019] in $d = 5, 7$. Compared to these works, one important improvement in Theorem 1.6 is the regularity of the solution which allows for the classical interpretation. Furthermore, we prove Lipschitz dependence of the blow-up time and the blow-up point on the initial data. Finally, from a technical perspective, the adapted inner product defined in Section 3 is simpler than the corresponding expressions in [Chatzikaleas and Donninger 2019] and can easily be generalized.

1B. ***Related results.*** Wave equations with focusing power nonlinearities provide the simplest possible models for the study of nonlinear wave dynamics and have been investigated intensively in the past decades. Consequently, local well-posedness and the behavior of solutions for small initial data are by now well understood; see, e.g., [Lindblad and Sogge 1995]. Concerning global dynamics for large initial data, substantial progress has been made more recently for energy critical problems. This includes fundamental works on the characterization of the threshold between finite-time blowup and dispersion in terms of the well-known stationary ground state solution; see [Kenig and Merle 2008; Krieger et al. 2015].

In contrast, large data results for energy supercritical equations are rare. For various models, the ODE blowup is known to provide a stable blow-up mechanism and Theorem 1.6 further extends these results; see Remark 1.7. In [Bizoń et al. 2007], nontrivial self-similar solutions are constructed for odd supercritical nonlinearities in dimension 3, and [Dai and Duyckaerts 2021] provides a generalization to $d \geq 4$. Also, in the three-dimensional case, large global solutions were obtained for a supercritical nonlinearity in [Krieger and Schlag 2017]. Finally, for $d \geq 11$ and large enough nonlinearities, manifolds of codimension greater than or equal to two have been constructed in [Collot 2018] that lead to non-self-similar blowup in finite time.

In the description of threshold dynamics for energy supercritical wave equations, self-similar solutions appear to play the key role. This has been observed numerically for power-type nonlinearities [Bizoń et al. 2004; Glogić et al. 2020], but also for more physically relevant models such as wave maps [Biernat et al. 2017; Bizoń et al. 2000] or the Yang–Mills equation in equivariant symmetry [Bizoń and Tabor 2001; Bizoń 2002]. We note that the latter reduces essentially to a radial quadratic wave equation in $d \geq 7$, hence (1-1) provides a toy model. From an analytic point of view, threshold phenomena for energy supercritical wave equations are entirely unexplored. Moreover, results analogous to the energy critical case seem completely out of reach.

However, very recently, the first *explicit* candidate for a self-similar threshold solution has been found by the second and third authors in [Glogić and Schörkhuber 2021] for the focusing cubic wave equation in all supercritical space dimensions $d \geq 5$. In $d = 7$, by the conformal symmetry of the linearized equation, the genuine unstable direction could be given in closed form, see also [Glogić et al. 2020], which allowed for a rigorous stability analysis. Interestingly, the same effect occurs for the quadratic wave equation and the new self-similar solution (1-4) in $d = 9$, which explains the specific choice of the space dimension in Theorem 1.1. In view of our results, we conjecture that the self-similar profile U given in (1-5) plays an important role in the threshold dynamics for (1-1) and (1-6).

In the proofs of Theorems 1.1 and 1.6 we build on methods developed in earlier works, in particular, [Donninger and Schörkhuber 2016; Glogić and Schörkhuber 2021]. However, several aspects, in particular the spectral analysis, are specific to the problem and rather delicate. Furthermore, we add important generalizations such as the preservation of regularity, which improves the statements of these earlier works. The presentation of our results is completely self-contained and all necessary details are provided in the proofs.

1C. ***Notation.*** Throughout the whole paper the Einstein summation convention is in force, i.e., we sum over repeated upper and lower indices, where latin indices run from 1 to d. We write $\mathbb{N}$ for the natural numbers $\{1, 2, 3, \dots\}$ and $\mathbb{N}_0 := \{0\} \cup \mathbb{N}$. Furthermore, $\mathbb{R}^+ := \{x \in \mathbb{R} : x > 0\}$. Also, $\overline{\mathbb{H}}$ stands for the closed complex right half-plane. By $\mathbb{B}^d_R(x_0)$ we denote the open ball of radius $R > 0$ in $\mathbb{R}^d$ centered at $x_0 \in \mathbb{R}^d$. The unit ball is abbreviated by $\mathbb{B}^d := \mathbb{B}^d_1(0)$, and $\mathbb{S}^{d-1} := \partial\mathbb{B}^d$. The notation $a \lesssim b$ means $a \leq Cb$ for an absolute constant $C > 0$, and we write $a \simeq b$ if $a \lesssim b$ and $b \lesssim a$. If $a \leq C_\varepsilon b$ for a constant $C_\varepsilon > 0$ depending on some parameter ε, we write $a \lesssim_\varepsilon b$.

By $L^2(\mathbb{B}^d_R(x_0))$ and $H^k(\mathbb{B}^d_R(x_0))$, $k \in \mathbb{N}_0$, we denote the Lebesgue and Sobolev spaces, respectively, obtained from the completion of $C^\infty(\mathbb{B}^d_R(x_0))$ with respect to the usual norm

$$\|u\|^2_{H^k(\mathbb{B}^d_R(x_0))} := \sum_{|\alpha| \leq k} \|\partial^\alpha u\|^2_{L^2(\mathbb{B}^d_R(x_0))},$$

with $\alpha \in \mathbb{N}_0^d$ denoting a multi-index and $\partial^\alpha u = \partial_1^{\alpha_1} \cdots \partial_d^{\alpha_d} u$, where $\partial_i u(x) = \partial_{x_j} u(x)$. For vector-valued functions, we use boldface letters, e.g., $\boldsymbol{f} = (f_1, f_2)$ and we sometime write $[\boldsymbol{f}]_1 := f_1$ to extract a single component. Throughout the paper, $W(f, g)$ denotes the Wronskian of two functions $f, g \in C^1(I)$, $I \subset \mathbb{R}$, where we use the convention $W(f, g) = fg' - f'g$, with f' denoting the first derivative. On a Hilbert space $\mathcal{H}$ we denote by $\mathcal{B}(\mathcal{H})$ the set of bounded linear operators. For a closed linear operator $(L, \mathcal{D}(L))$

on $\mathcal{H}$, we define the resolvent set $\rho(L)$ as the set of all $\lambda \in \mathbb{C}$ such that $R_L(\lambda) := (\lambda - L)^{-1}$ exists as a bounded operator on the whole underlying space. Furthermore, the spectrum of L is defined as $\sigma(L) := \mathbb{C} \setminus \rho(L)$ and the point spectrum is denoted by $\sigma_p(L) \subset \sigma(L)$.

Spherical harmonics. Fix a dimension $d \geq 3$. For $\ell \in \mathbb{N}_0$, an eigenfunction for the Laplace–Beltrami operator on $\mathbb{S}^{d-1}$ with eigenvalue $\ell(\ell + d - 2)$ is called a spherical harmonic function of degree ℓ. For each $\ell \in \mathbb{N}$, we denote by $M_{d,\ell}$ the number of linearly independent spherical harmonics of degree ℓ, and for $\Omega_\ell := \{1, \dots, M_{d,\ell}\}$ we designate by $\{Y_{\ell,m} : m \in \Omega_\ell\}$ a set of orthonormal spherical harmonics, i.e.,

$$\int_{\mathbb{S}^{d-1}} Y_{\ell,m}(\omega)\overline{Y_{\ell,m'}(\omega)}\, d\sigma(\omega) = \delta_{mm'}.$$

Obviously, one has $\Omega_0 = \{1\}$ and $\Omega_1 = \{1, \dots, d\}$, and we can take $Y_{0,1}(\omega) = c_1$ and $Y_{1,m}(\omega) = \tilde{c}_m \omega_m$ for suitable normalization constants $c_1, \tilde{c}_m \in \mathbb{R}$. For $g \in C^\infty(\mathbb{S}^{d-1})$, we define $\mathcal{P}_\ell : L^2(\mathbb{S}^{d-1}) \to L^2(\mathbb{S}^{d-1})$ by

$$\mathcal{P}_\ell g(\omega) := \sum_{m \in \Omega_\ell} (g \,|\, Y_{\ell,m})_{L^2(\mathbb{S}^{d-1})} Y_{\ell,m}(\omega).$$

It is well known, see, e.g., [Atkinson and Han 2012], that $\mathcal{P}_\ell$ defines a self-adjoint projection on $L^2(\mathbb{S}^{d-1})$ and that $\lim_{n\to\infty} \| g - \sum_{\ell=0}^n \mathcal{P}_\ell g \|_{L^2(\mathbb{S}^{d-1})} = 0$. This can be extended to Sobolev spaces, in particular, $\lim_{n\to\infty} \| g - \sum_{\ell=0}^n \mathcal{P}_\ell g \|_{H^k(\mathbb{S}^{d-1})} = 0$ for all $g \in C^\infty(\mathbb{S}^{d-1})$, see, e.g., [Donninger and Schörkhuber 2016], Lemma A.1. Furthermore, given $f \in C^\infty(\overline{\mathbb{B}}^d_R)$, by setting

$$[P_\ell f](x) := \sum_{m \in \Omega_\ell} (f(|x|\cdot) \,|\, Y_{\ell,m})_{L^2(\mathbb{S}^{d-1})} Y_{\ell,m}\left(\frac{x}{|x|}\right), \tag{1-22}$$

we have that (see for example Lemma A.2 in [Donninger and Schörkhuber 2016])

$$\lim_{n\to\infty} \left\| f - \sum_{\ell=0}^n P_\ell f \right\|_{H^k(\mathbb{B}^d_R)} = 0. \tag{1-23}$$

2. The stability problem in similarity coordinates

In this section we formulate (1-1) in similarity variables. The advantage of the new setting is the fact that self-similar solutions become time-independent and stability of finite-time blowup turns into asymptotic stability of static solutions. Then we state the main results in the new coordinate system.

Given $T > 0$ and $x_0 \in \mathbb{R}^d$, we define *similarity coordinates*

$$\tau := -\log(T - t) + \log T \quad \text{and} \quad \xi := \frac{x - x_0}{T - t}.$$

Note that in (τ, ξ), the backward light cone $\mathcal{C}_{T,x_0}$ corresponds to the infinite cylinder

$$\mathcal{Z} := \bigcup_{\tau \geq 0} \{\tau\} \times \overline{\mathbb{B}}^d.$$

Furthermore, by setting

$$\psi(\tau, \xi) := T^2 e^{-2\tau} u(T - Te^{-\tau}, Te^{-\tau}\xi + x_0),$$

(1-1) transforms into

$$(\partial_\tau^2 + 5\partial_\tau + 2\xi\cdot\nabla\partial_\tau + (\xi\cdot\nabla)^2 - \Delta + 5\xi\cdot\nabla + 6)\psi(\tau,\xi) = \psi(\tau,\xi)^2. \tag{2-1}$$

To get a first-order formulation we define

$$\psi_1(\tau,\xi) := \psi(\tau,\xi) \quad \text{and} \quad \psi_2(\tau,\xi) := \partial_\tau\psi(\tau,\xi) + \xi\cdot\nabla\psi(\tau,\xi) + 2\psi(\tau,\xi), \tag{2-2}$$

and let $\Psi(\tau) = (\psi_1(\tau,\cdot), \psi_2(\tau,\cdot))$, by means of which (2-1) can be written as

$$\partial_\tau\Psi(\tau) = \widetilde{\boldsymbol{L}}\Psi(\tau) + \boldsymbol{F}(\Psi(\tau)), \tag{2-3}$$

where

$$\widetilde{\boldsymbol{L}}\boldsymbol{u}(\xi) = \begin{pmatrix} -\xi\cdot\nabla u_1(\xi) - 2u_1(\xi) + u_2(\xi) \\ \Delta u_1(\xi) - \xi\cdot\nabla u_2(\xi) - 3u_2(\xi) \end{pmatrix} \quad \text{and} \quad \boldsymbol{F}(\boldsymbol{u}) = \begin{pmatrix} 0 \\ u_1^2 \end{pmatrix}$$

for $\boldsymbol{u} = (u_1, u_2)$. Note that in the new variables, the solutions $u^*_{T,x_0,a}$ and $u^{\mathrm{ODE}}_{T,x_0,a}$ become static. Namely, every $a \in \mathbb{R}^d$ yields smooth, positive, and τ-independent solutions

$$\boldsymbol{U}_a = (U_{1,a}, U_{2,a}) \quad \text{and} \quad \boldsymbol{\kappa}_a = (\kappa_{1,a}, \kappa_{2,a})$$

of (2-3) given by

$$U_{1,a}(\xi) = U_a(\xi), \quad U_{2,a}(\xi) = \xi\cdot\nabla U_a(\xi) + 2U_a(\xi),$$
$$\kappa_{1,a}(\xi) = \kappa_a(\xi), \quad \kappa_{2,a}(\xi) = \xi\cdot\nabla\kappa_a(\xi) + 2\kappa_a(\xi).$$

We study (2-3) for small perturbations of $\boldsymbol{U}_a$ and $\boldsymbol{\kappa}_a$ in the Hilbert space

$$\mathcal{H} := H^{\frac{d+1}{2}}(\mathbb{B}^d) \times H^{\frac{d-1}{2}}(\mathbb{B}^d)$$

equipped with the standard norm

$$\|\boldsymbol{u}\|^2 := \|u_1\|^2_{H^{(d+1)/2}(\mathbb{B}^d)} + \|u_2\|^2_{H^{(d-1)/2}(\mathbb{B}^d)}.$$

Also, write $\mathcal{B}_R := \{\boldsymbol{u} \in \mathcal{H} : \|\boldsymbol{u}\| \le R\}$.

In Proposition 3.1 on page 629 we show that, for $d \in \{7, 9\}$, the operator

$$\widetilde{\boldsymbol{L}} : C^\infty(\overline{\mathbb{B}}^d) \times C^\infty(\overline{B}^d) \subset \mathcal{H} \to \mathcal{H},$$

which describes the free wave evolution in similarity coordinates, is closable and its closure, which we denote by

$$\boldsymbol{L} : \mathcal{D}(\boldsymbol{L}) \subset \mathcal{H} \to \mathcal{H},$$

generates a strongly continuous one-parameter semigroup $(\boldsymbol{S}(\tau))_{\tau\ge 0} \subset \mathcal{B}(\mathcal{H})$. By using the Sobolev embedding, it is easy to see that the nonlinearity satisfies

$$\|\boldsymbol{F}(\boldsymbol{u})\| = \|u_1^2\|_{H^{(d-1)/2}(\mathbb{B}^d)} \le \|u_1^2\|_{H^{(d+1)/2}(\mathbb{B}^d)} \lesssim \|u_1\|^2_{H^{(d+1)/2}(\mathbb{B}^d)} \lesssim \|\boldsymbol{u}\|^2$$

for all $\boldsymbol{u} \in \mathcal{H}$; hence $\boldsymbol{F}$ is well defined on $\mathcal{H}$.

2A. ***Stability of U_a.*** The key to proving Theorem 1.1 is the following result, which establishes, for $d = 9$, conditional orbital asymptotic stability of the family of static solutions $\{\boldsymbol{U}_a : a \in \mathbb{R}^9\}$.

Proposition 2.1. *Let $d = 9$. There are constants $C > 0$ and $\omega > 0$ such that the following holds. For all sufficiently small $\delta > 0$ there exists a codimension-11 Lipschitz manifold $\mathcal{M} = \mathcal{M}_{\delta,C} \subset \mathcal{B}_{\delta/C}$ with $\mathbf{0} \in \mathcal{M}$ such that, for any $\Phi_0 \in \mathcal{M}$, there are $\Psi \in C([0,\infty), \mathcal{H})$ and $a \in \overline{\mathbb{B}}^9_{\delta/\omega}$ such that*

$$\Psi(\tau) = \boldsymbol{S}(\tau)(\boldsymbol{U}_0 + \Phi_0) + \int_0^\tau \boldsymbol{S}(\tau - \sigma)\boldsymbol{F}(\Psi(\sigma))\, d\sigma \tag{2-4}$$

and

$$\|\Psi(\tau) - \boldsymbol{U}_a\| \lesssim \delta e^{-\omega\tau}$$

for all $\tau \geq 0$.

The number of codimensions in Proposition 2.1 is related to the number of unstable eigenvalues of the linearization around $\boldsymbol{U}_a$ and the dimension of the corresponding eigenspaces; see Section 5. In fact, ten of these instabilities are caused by the translation symmetries of the problem, and can be controlled by choosing appropriately the blow-up parameters (T, x_0). There is, therefore, only one genuine unstable direction. Next, we state a persistence of regularity result for solutions to (2-4).

Proposition 2.2. *If the initial data Φ_0 from Proposition 2.1 is in $C^\infty(\overline{\mathbb{B}}^9) \times C^\infty(\overline{\mathbb{B}}^9)$ then the corresponding solution Ψ of (2-3) belongs to $C^\infty(\mathcal{Z}) \times C^\infty(\mathcal{Z})$. In particular, Ψ satisfies (2-3) in the classical sense.*

Remark 2.3. That this proposition is not vacuous, i.e., that there exists $\Phi_0 \in \mathcal{M} \cap (C^\infty(\overline{\mathbb{B}}^9) \times C^\infty(\overline{\mathbb{B}}^9))$, follows from Proposition 2.4.

The proofs of Propositions 2.1 and 2.2 are provided in Section 7D.

In order to derive Theorem 1.1 from the above results we prescribe in physical variables initial data of the form

$$u(0,\cdot) = u^*_{1,0,0}(0,\cdot) + f \quad \text{and} \quad \partial_t u(0,\cdot) = \partial_t u^*_{1,0,0}(0,\cdot) + g \tag{2-5}$$

for free functions (f, g) defined on a suitably large ball centered at the origin. In similarity variables, this transforms into initial data $\Psi(0) = \boldsymbol{U}_0 + \Phi_0$ for (2-3), with

$$\Phi_0 = \Upsilon((f,g), T, x_0), \tag{2-6}$$

where

$$\Upsilon((f,g), T, x_0) := \mathcal{R}((f,g), T, x_0) + \mathcal{R}(\boldsymbol{U}_0, T, x_0) - \mathcal{R}(\boldsymbol{U}_0, 1, 0) \tag{2-7}$$

and

$$\mathcal{R}((f_1, f_2), T, x_0) = \begin{pmatrix} T^2 f_1(T\cdot + x_0) \\ T^3 f_2(T\cdot + x_0) \end{pmatrix}.$$

The next statement asserts that, for all small (f, g), there is a choice of parameters x_0, T, and α for which $\Upsilon((f + \alpha h_1, g + \alpha h_2), T, x_0)$ belongs to the manifold $\mathcal{M}$ from Proposition 2.1.

Proposition 2.4. *Let (h_1, h_2) be defined as in (1-15). There exists $M > 0$ such that, for all sufficiently small $\delta > 0$, the following holds. For any $(f, g) \in H^6(\mathbb{B}_2^9) \times H^5(\mathbb{B}_2^9)$ satisfying*

$$\|(f, g)\|_{H^6(\mathbb{B}_2^9)\times H^5(\mathbb{B}_2^9)} \leq \frac{\delta}{M^2},$$

there are $x_0 \in \overline{\mathbb{B}}^9_{\delta/M}$, $T \in [1-\delta/M, 1+\delta/M]$, and $\alpha \in [-\delta/M, \delta/M]$ depending Lipschitz continuously on (f, g) such that

$$\Upsilon((f + \alpha h_1, g + \alpha h_2), T, x_0) \in \mathcal{M}_{\delta,C},$$

where $\mathcal{M}_{\delta,C}$ is the manifold from Proposition 2.1.

Theorem 1.1 is then obtained by transforming the results of Propositions 2.1, 2.2, and 2.4 back to coordinates (t, x).

Remark 2.5. We note that when proving stability of the ODE blow-up solution for $d \in \{7, 9\}$ similar results are obtained. In fact, the proof implies the existence of a Lipschitz manifold $\mathcal{N}$ of codimension $d+1$ in the Hilbert space $\mathcal{H}$, according to $d+1$ directions of instability induced by translation invariance. A result similar to Proposition 2.4 guarantees that for *any* small enough data (f, g) one can suitably adjust the blow-up time T and the blow-up point x_0 such that $\Upsilon((f, g), T, x_0) \in \mathcal{N}$, which gives Theorem 1.6 on stable blowup. This point of view further justifies using codimension-1 terminology to describe the stability of u^*.

Time-evolution for small perturbations: modulation ansatz. In the following, we assume that $a = a(\tau)$, $a(0) = 0$, and $\lim_{\tau\to\infty} a(\tau) = a_\infty$. Inserting the ansatz

$$\Psi(\tau) = \boldsymbol{U}_{a(\tau)} + \Phi(\tau) \tag{2-8}$$

into (2-3) we obtain

$$\partial_\tau \Phi(\tau) = (\widetilde{\boldsymbol{L}} + \boldsymbol{L}'_{a(\tau)})\Phi(\tau) + \boldsymbol{F}(\Phi(\tau)) - \partial_\tau \boldsymbol{U}_{a(\tau)},$$

with

$$\boldsymbol{L}'_{a(\tau)}\boldsymbol{u} = \begin{pmatrix} 0 \\ V_{a(\tau)} u_1 \end{pmatrix} \quad \text{and} \quad V_a(\xi) = 2U_a(\xi).$$

In the following, we define

$$\boldsymbol{G}_{a(\tau)}(\Phi(\tau)) := [\boldsymbol{L}'_{a(\tau)} - \boldsymbol{L}'_{a_\infty}]\Phi(\tau) + \boldsymbol{F}(\Phi(\tau))$$

and study the evolution equation

$$\partial_\tau \Phi(\tau) = [\widetilde{\boldsymbol{L}} + \boldsymbol{L}'_{a_\infty}]\Phi(\tau) + \boldsymbol{G}_{a(\tau)}(\Phi(\tau)) - \partial_\tau \boldsymbol{U}_{a(\tau)}, \tag{2-9}$$

with initial data $\Phi(0) = \boldsymbol{u} \in \mathcal{H}$. This naturally splits into three parts: First, in Section 3, we study the time evolution governed by $\widetilde{\boldsymbol{L}}$ using semigroup theory. In Section 4, we analyze the linearized problem, where we consider $\widetilde{\boldsymbol{L}} + \boldsymbol{L}'_{a_\infty}$ as a (compact) perturbation of the free evolution and investigate the underlying spectral problem, restricting to $d = 9$. Resolvent bounds allow us to transfer the spectral information to suitable growth estimates for the linearized time evolution. The nonlinear problem will be analyzed in integral form in Section 7, using modulation theory and fixed-point arguments. Also, we

prove Propositions 2.1–2.4 and, based on this, Theorem 1.1. In Section 8 we give the main arguments to prove Theorem 1.6.

3. The free wave evolution in similarity variables

In this section we prove well-posedness of the linear version of (2-3) in $\mathcal{H}$. In other words, we show that the (closure of the) operator $\widetilde{\boldsymbol{L}}$ generates a strongly continuous one-parameter semigroup of bounded operators on $\mathcal{H}$. What is more, in view of the regularity result Proposition 2.2, we consider the evolution in Sobolev spaces of arbitrarily high integer order. In Section 4 we then restrict the problem again to $\mathcal{H}$.

For $k \geq 1$, let

$$\mathcal{H}_k := H^k(\mathbb{B}^d) \times H^{k-1}(\mathbb{B}^d)$$

be equipped with the standard norm denoted by $\|\cdot\|_{H^k(\mathbb{B}^d)\times H^{k-1}(\mathbb{B}^d)}$. We set

$$\mathcal{D}(\widetilde{\boldsymbol{L}}) := C^\infty(\overline{\mathbb{B}}^d) \times C^\infty(\overline{\mathbb{B}}^d)$$

and consider the densely defined operator

$$\widetilde{\boldsymbol{L}} : \mathcal{D}(\widetilde{\boldsymbol{L}}) \subset \mathcal{H}_k \to \mathcal{H}_k.$$

We now state the central result of this section.

Proposition 3.1. *Let $d \in \{7, 9\}$ and $k \geq 3$. The operator $\widetilde{\boldsymbol{L}} : \mathcal{D}(\widetilde{\boldsymbol{L}}) \subset \mathcal{H}_k \to \mathcal{H}_k$ is closable and its closure $\boldsymbol{L}_k : \mathcal{D}(\boldsymbol{L}_k) \subset \mathcal{H}_k \to \mathcal{H}_k$ generates a strongly continuous semigroup $\boldsymbol{S}_k : [0, \infty) \to \mathcal{B}(\mathcal{H}_k)$ which satisfies*

$$\|\boldsymbol{S}_k(\tau)\boldsymbol{u}\|_{H^k(\mathbb{B}^d)\times H^{k-1}(\mathbb{B}^d)} \leq M_k e^{-\frac{1}{2}\tau} \|\boldsymbol{u}\|_{H^k(\mathbb{B}^d)\times H^{k-1}(\mathbb{B}^d)} \tag{3-1}$$

for all $\boldsymbol{u} \in \mathcal{H}_k$, all $\tau \geq 0$, and some $M_k > 1$. Furthermore, the following holds for the spectrum of $\boldsymbol{L}_k$:

$$\sigma(\boldsymbol{L}_k) \subset \left\{z \in \mathbb{C} : \operatorname{Re} z \leq -\tfrac{1}{2}\right\}, \tag{3-2}$$

and the resolvent has the bound

$$\|\boldsymbol{R}_{\boldsymbol{L}_k}(\lambda)\boldsymbol{f}\|_{H^k(\mathbb{B}^d)\times H^{k-1}(\mathbb{B}^d)} \leq \frac{M_k}{\operatorname{Re}\lambda + \frac{1}{2}} \|\boldsymbol{f}\|_{H^k(\mathbb{B}^d)\times H^{k-1}(\mathbb{B}^d)}$$

for $\lambda \in \mathbb{C}$ with $\operatorname{Re}\lambda > -\frac{1}{2}$ and $\boldsymbol{f} \in \mathcal{H}_k$.

Remark 3.2. We prove Proposition 3.1 via the Lumer–Phillips Theorem. By using the standard inner product on $\mathcal{H}_k$, one can easily prove existence of the semigroup $(\boldsymbol{S}_k(\tau))_{\tau\geq 0}$, but in order to show that it decays exponentially and to prove the growth bound (3-1) in particular, we need to introduce an appropriate equivalent inner product. The necessity for such an approach will become apparent in the proof of Lemma 3.4 in the Appendix. We note that, for $d = 9$, the restriction on k is optimal within the class of integer Sobolev spaces. In particular, for scaling reasons exponential decay cannot be expected at lower integer regularities. For $d = 7$, a similar statement can be obtained for $k = 2$.

For $d \in \{7, 9\}$ and $k \geq 3$ we define the sesquilinear form

$$(\cdot \,|\, \cdot)_{\mathcal{H}_k} : (C^\infty(\overline{\mathbb{B}}^d) \times C^\infty(\overline{\mathbb{B}}^d))^2 \to \mathbb{C}, \quad (\boldsymbol{u} \,|\, \boldsymbol{v})_{\mathcal{H}_k} = \sum_{j=1}^{k} (\boldsymbol{u} \,|\, \boldsymbol{v})_j,$$

where

$$\begin{aligned}
(\boldsymbol{u} \,|\, \boldsymbol{v})_1 &= \int_{\mathbb{S}^{d-1}} \partial_i u_1(\omega) \overline{\partial^i v_1(\omega)}\, d\sigma(\omega) + \int_{\mathbb{S}^{d-1}} u_1(\omega) \overline{v_1(\omega)}\, d\sigma(\omega) + \int_{\mathbb{S}^{d-1}} u_2(\omega) \overline{v_2(\omega)}\, d\sigma(\omega), \\
(\boldsymbol{u} \,|\, \boldsymbol{v})_2 &= \int_{\mathbb{B}^d} \partial_i \Delta u_1(\xi) \overline{\partial^i \Delta v_1(\xi)}\, d\xi + \int_{\mathbb{B}^d} \partial_i \partial_j u_2(\xi) \overline{\partial^i \partial^j v_2(\xi)}\, d\xi + \int_{\mathbb{S}^{d-1}} \partial_i u_2(\omega) \overline{\partial^i v_2(\omega)}\, d\sigma(\omega), \\
(\boldsymbol{u} \,|\, \boldsymbol{v})_3 &= 4 \int_{\mathbb{B}^d} \partial_i \partial_j \partial_k u_1(\xi) \overline{\partial^i \partial^j \partial^k v_1(\xi)}\, d\xi + 4 \int_{\mathbb{B}^d} \partial_i \partial_j u_2(\xi) \overline{\partial^i \partial^j v_2(\xi)}\, d\xi \\
&\qquad + 4 \int_{\mathbb{S}^{d-1}} \partial_i \partial_j u_1(\omega) \overline{\partial^i \partial^j v_1(\omega)}\, d\sigma(\omega),
\end{aligned}$$

and for $j \geq 4$ we use the standard $\dot{H}^j(\mathbb{B}^d) \times \dot{H}^{j-1}(\mathbb{B}^d)$ inner product

$$(\boldsymbol{u} \,|\, \boldsymbol{v})_j = (u_1 \,|\, v_1)_{\dot{H}^j(\mathbb{B}^d)} + (u_2 \,|\, v_2)_{\dot{H}^{j-1}(\mathbb{B}^d)}. \tag{3-3}$$

We then set

$$\|\boldsymbol{u}\|_{\mathcal{H}_k} := \sqrt{(\boldsymbol{u} \,|\, \boldsymbol{u})_{\mathcal{H}_k}}.$$

For brevity, we will use the notation $(\cdot \,|\, \cdot)_j = \|\cdot\|_j^2$, $j = 1, \dots, k$, for different parts of $(\cdot \,|\, \cdot)_{\mathcal{H}_k}$.

Lemma 3.3. *Let $d \in \{7, 9\}$ and $k \geq 3$. We have*

$$\|\boldsymbol{u}\|_{\mathcal{H}_k} \simeq \|\boldsymbol{u}\|_{H^k(\mathbb{B}^d) \times H^{k-1}(\mathbb{B}^d)}$$

for all $\boldsymbol{u} \in C^\infty(\overline{\mathbb{B}}^d) \times C^\infty(\overline{\mathbb{B}}^d)$. In particular, $\|\cdot\|_{\mathcal{H}_k}$ defines an equivalent norm on $\mathcal{H}_k$.

Proof. Note that it suffices to prove

$$\|\boldsymbol{u}\|^2_{H^3(\mathbb{B}^d) \times H^2(\mathbb{B}^d)} \lesssim \sum_{j=1}^{3} \|\boldsymbol{u}\|_j^2 \lesssim \|\boldsymbol{u}\|^2_{H^3(\mathbb{B}^d) \times H^2(\mathbb{B}^d)}. \tag{3-4}$$

The first estimate in (3-4) follows from the fact that

$$\|u\|^2_{L^2(\mathbb{B}^d)} \lesssim \|\nabla u\|^2_{L^2(\mathbb{B}^d)} + \|u\|^2_{L^2(\mathbb{S}^{d-1})}$$

for all $u \in C^\infty(\overline{\mathbb{B}}^d)$, which is a simple consequence of the identity

$$\begin{aligned}
\int_{\mathbb{S}^{d-1}} |u(\omega)|^2\, d\sigma(\omega) &= \int_{\mathbb{B}^d} \operatorname{div}(\xi |u(\xi)|^2)\, d\xi \\
&= \int_{\mathbb{B}^d} (d\,|u(\xi)|^2 + \xi^i u(\xi) \overline{\partial_i u(\xi)} + \xi^i \overline{u(\xi)} \partial_i u(\xi))\, d\xi.
\end{aligned} \tag{3-5}$$

Using this, it is easy to see that

$$\|u\|_{H^2(\mathbb{B}^d)} \lesssim \int_{\mathbb{B}^d} \partial_i \partial_j u(\xi) \overline{\partial^i \partial^j u(\xi)}\, d\xi + \int_{\mathbb{S}^{d-1}} \partial_i u(\omega) \overline{\partial^i u(\omega)}\, d\sigma(\omega) + \int_{\mathbb{S}^{d-1}} |u(\omega)|^2\, d\sigma(\omega)$$

for all $u \in C^\infty(\overline{\mathbb{B}}^d)$. Similar bounds imply the first inequality in (3-4). Another consequence of (3-5) is the trace theorem, which asserts that

$$\int_{\mathbb{S}^{d-1}} |u(\omega)|^2 \, d\sigma(\omega) \lesssim \|u\|^2_{H^1(\mathbb{B}^d)}$$

for all $u \in C^\infty(\overline{\mathbb{B}}^d)$; using this, it is straightforward to obtain the second inequality in (3-4). Hence we obtain the claimed estimates in Lemma 3.3 for all $\boldsymbol{u} \in C^\infty(\overline{\mathbb{B}}^d) \times C^\infty(\overline{\mathbb{B}}^d)$ and, by density, we extend this to all of $\mathcal{H}_k$. □

Now we turn to proving Proposition 3.1. As the first auxiliary result, we have the following dissipation property of $\widetilde{\boldsymbol{L}}$.

Lemma 3.4. *Let $d \in \{7, 9\}$ and $k \geq 3$. Then*

$$\operatorname{Re}(\widetilde{\boldsymbol{L}}\boldsymbol{u} \,|\, \boldsymbol{u})_{\mathcal{H}_k} \leq -\tfrac{1}{2}\|\boldsymbol{u}\|^2_{\mathcal{H}_k}$$

for all $\boldsymbol{u} \in \mathcal{D}(\widetilde{\boldsymbol{L}})$.

The proof is provided in the Appendix. To apply the Lumer–Phillips theorem, we also need the following density property of $\widetilde{\boldsymbol{L}}$.

Lemma 3.5. *Let $d \in \{7, 9\}$ and $k \geq 3$. There exists $\lambda > -\frac{1}{2}$ such that* $\operatorname{ran}(\lambda - \widetilde{\boldsymbol{L}})$ *is dense in $\mathcal{H}_k$.*

Proof. Let $d \in \{7, 9\}$ and $k \geq 3$. We prove the statement by showing that there exists a λ such that, given $\boldsymbol{f} \in \mathcal{H}_k$ and $\varepsilon > 0$, there is some $\boldsymbol{f}_\varepsilon$ in the ε-neighborhood of $\boldsymbol{f}$ for which the equation $(\lambda - \widetilde{\boldsymbol{L}})\boldsymbol{u} = \boldsymbol{f}_\varepsilon$ admits a solution in $\mathcal{D}(\widetilde{\boldsymbol{L}})$. First, by density, there is $\tilde{\boldsymbol{f}} \in C^\infty(\overline{\mathbb{B}}^d) \times C^\infty(\overline{\mathbb{B}}^d)$ for which $\|\tilde{\boldsymbol{f}} - \boldsymbol{f}\|_{H^k(\mathbb{B}^d)\times H^{k-1}(\mathbb{B}^d)} < \frac{1}{2}\varepsilon$. Then, for $n \in \mathbb{N}$, we define $\boldsymbol{f}_n := (f_{1,n}, f_{2,n})$ with

$$f_{1,n} = \sum_{\ell=0}^{n} P_\ell \tilde{f}_1 \quad \text{and} \quad f_{2,n} = \sum_{\ell=0}^{n} P_\ell \tilde{f}_2,$$

where the P_ℓ are the projection operators defined in (1-22). Furthermore, according to (1-23) there exists an index $N \in \mathbb{N}$ for which $\|\boldsymbol{f}_N - \tilde{\boldsymbol{f}}\|_{H^k(\mathbb{B}^d)\times H^{k-1}(\mathbb{B}^d)} < \frac{1}{2}\varepsilon$. It is therefore sufficient to consider

$$(\lambda - \widetilde{\boldsymbol{L}})\boldsymbol{u} = \boldsymbol{f}_N \tag{3-6}$$

and produce a solution $\boldsymbol{u} \in \mathcal{D}(\widetilde{\boldsymbol{L}})$. First, we rewrite (3-6) as a system of equations in u_1 and u_2:

$$-(\delta^{ij} - \xi^i\xi^j)\partial_i\partial_j u_1(\xi) + 2(\lambda+3)\xi^i\partial_i u_1(\xi) + (\lambda+3)(\lambda+2)u_1(\xi) = g_N(\xi), \tag{3-7}$$

$$u_2(\xi) = \xi^i\partial_i u_1(\xi) + (\lambda+1)u_1(\xi) - f_{1,N}(\xi), \tag{3-8}$$

where

$$g_N(\xi) = \xi^i\partial_i f_{1,N}(\xi) + (\lambda+3)f_{1,N}(\xi) + f_{2,N}(\xi).$$

We now treat the case $d = 9$, for which we choose $\lambda = \frac{5}{2}$. With this choice, (3-7) reads as

$$-(\delta^{ij} - \xi^i\xi^j)\partial_i\partial_j u_1(\xi) + 11\xi^i\partial_i u_1(\xi) + \tfrac{99}{4}u_1(\xi) = g_N(\xi). \tag{3-9}$$

Note that g_N is a finite linear combination of spherical harmonics, and this allows us to decompose the PDE (3-9) which is posed on $\mathbb{B}^9$ into a finite number of ODEs posed on the interval $(0,1)$. To this end, we switch to spherical coordinates $\rho = |\xi|$ and $\omega = \xi/|\xi|$. In particular, the relevant differential expressions transform in the following way:

$$\begin{aligned}\xi^i \partial_i u(\xi) &= \rho \partial_\rho u(\rho\omega),\\ \xi^i \xi^j \partial_i \partial_j u(\xi) &= \rho^2 \partial_\rho^2 u(\rho\omega),\\ \partial^i \partial_i u(\xi) &= \left(\partial_\rho^2 + \frac{8}{\rho}\partial_\rho + \frac{1}{\rho^2}\Delta_\omega^{\mathbb{S}^8}\right) u(\rho\omega).\end{aligned}$$

Consequently, (3-9) becomes

$$\left(-(1-\rho^2)\partial_\rho^2 + \left(-\frac{8}{\rho} + 11\rho\right)\partial_\rho + \frac{99}{4} - \frac{1}{\rho^2}\Delta_\omega^{\mathbb{S}^8}\right) u(\rho\omega) = g_N(\rho\omega). \tag{3-10}$$

Now we take the decomposition of the right-hand side of (3-10) into spherical harmonics:

$$g_N(\rho\omega) = \sum_{\ell=0}^{N} \sum_{m\in\Omega_\ell} g_{\ell,m}(\rho) Y_{\ell,m}(\omega)$$

for some $g_{\ell,m} \in C^\infty[0,1]$. Then by inserting the ansatz

$$u_1(\rho\omega) = \sum_{\ell=0}^{N} \sum_{m\in\Omega_\ell} u_{\ell,m}(\rho) Y_{\ell,m}(\omega) \tag{3-11}$$

into (3-10), we obtain the system of ODEs

$$\left(-(1-\rho^2)\partial_\rho^2 + \left(-\frac{8}{\rho} + 11\rho\right)\partial_\rho + \frac{\ell(\ell+7)}{\rho^2} + \frac{99}{4}\right) u_{\ell,m}(\rho) = g_{\ell,m}(\rho) \tag{3-12}$$

for $\ell = 0, \dots, N$ and $m \in \Omega_\ell$. For later convenience, we first set $v_{\ell,m}(\rho) = \rho^3 u_{\ell,m}(\rho)$ and thereby transform (3-12) into

$$\left(-(1-\rho^2)\partial_\rho^2 + \left(-\frac{2}{\rho} + 5\rho\right)\partial_\rho + \frac{(\ell+4)(\ell+3)}{\rho^2} + \frac{15}{4}\right) v_{\ell,m}(\rho) = \rho^3 g_{\ell,m}(\rho). \tag{3-13}$$

Then, by means of a further change of variables $v_{\ell,m}(\rho) = \rho^{\ell+3} w_{\ell,m}(\rho^2)$, we turn the homogeneous version of (3-13) into a hypergeometric equation in its canonical form:

$$z(1-z)w''_{\ell,m}(z) + (c-(a+b+1)z)w'_{\ell,m}(z) - abw_{l,m}(z) = 0, \tag{3-14}$$

where

$$a = \tfrac{1}{4}(9+2\ell), \quad b = a + \tfrac{1}{2} = \tfrac{1}{4}(11+2\ell), \quad \text{and} \quad c = 2a = \tfrac{1}{2}(9+2\ell).$$

Equation (3-14) admits the two solutions

$$\phi_{0,\ell}(z) = {}_2F_1\big(a, a+\tfrac{1}{2}, 2a, z\big) \quad \text{and} \quad \phi_{1,\ell}(z) = {}_2F_1\big(a, a+\tfrac{1}{2}, \tfrac{3}{2}, 1-z\big),$$

which are analytic around $z=0$ and $z=1$, respectively; see [DLMF 2010]. In fact, the functions $\phi_{0,\ell}$ and $\phi_{1,\ell}$ can be expressed in closed form as

$$\phi_{0,\ell}(z)=\frac{1}{\sqrt{1-z}}\left(\frac{2}{1+\sqrt{1-z}}\right)^{\frac{7}{2}+\ell},$$
$$\phi_{1,\ell}(z)=\sqrt{1-z}\left(\left(\frac{1}{1-\sqrt{1-z}}\right)^{\frac{7}{2}+\ell}-\left(\frac{1}{1+\sqrt{1-z}}\right)^{\frac{7}{2}+\ell}\right);$$

see [DLMF 2010, pp. 386-387]. Now by undoing the change of variables from above, we get solutions $\psi_{\ell,0}=\rho^{\ell+3}\phi_{0,\ell}(\rho^2)$ and $\psi_{\ell,1}=\rho^{\ell+3}\phi_{1,\ell}(\rho^2)$ to the homogeneous version of (3-13). Furthermore, the Wronskian is $W(\psi_{0,\ell},\psi_{1,\ell})(\rho)=C_\ell(1-\rho^2)^{-3/2}\rho^{-2}$ for some nonzero constant C_ℓ. Then, by the variation of constants formula we obtain a solution to (3-13) on $(0,1)$:

$$\begin{aligned}v_{\ell,m}(\rho)&=-\psi_{\ell,0}(\rho)\int_\rho^1\frac{\psi_{\ell,1}(s)}{W(\psi_{\ell,0},\psi_{\ell,1})(s)}\frac{s^3g_{\ell,m}(s)}{1-s^2}\,ds-\psi_{\ell,1}(\rho)\int_0^\rho\frac{\psi_{\ell,0}(s)}{W(\psi_{\ell,0},\psi_{\ell,1})(s)}\frac{s^3g_{\ell,m}(s)}{1-s^2}\,ds\\&=-\psi_{\ell,0}(\rho)\int_\rho^1\psi_{\ell,1}(s)\sqrt{1-s}h_{\ell,m}(s)\,ds-\psi_{\ell,1}(\rho)\int_0^\rho\psi_{\ell,0}(s)\sqrt{1-s}h_{\ell,m}(s)\,ds,\end{aligned}\tag{3-15}$$

where $h_{\ell,m}\in C^\infty[0,1]$. Obviously $v_{\ell,m}\in C^\infty(0,1)$. We claim that $v_{\ell,m}\in C^\infty(0,1]$. To see this, we note that, at $\rho=1$, the set of Frobenius indices of (3-13) is $\{-\frac{1}{2},0\}$. Hence near $\rho=1$, there is another solution, linearly independent of $\psi_{\ell,1}$, which has the form $(1-\rho)^{-1/2}\psi_{\ell,2}(\rho)$, where $\psi_{\ell,2}$ is analytic at $\rho=1$. Hence

$$\psi_{\ell,0}(\rho)=c_{\ell,1}\psi_{\ell,1}(\rho)+c_{\ell,2}\frac{\psi_{\ell,2}(\rho)}{\sqrt{1-\rho}}\tag{3-16}$$

for some constants $c_{\ell,1}$ and $c_{\ell,2}$. Now, by letting

$$\alpha_{\ell,m}:=\int_0^1\psi_{\ell,0}(s)\sqrt{1-s}h_{\ell,m}(s)\,ds$$

and inserting (3-16) into (3-15), we get

$$\begin{aligned}v_{\ell,m}(\rho)=-c_{\ell,2}\frac{\psi_{\ell,2}(\rho)}{\sqrt{1-\rho}}\int_\rho^1\psi_{\ell,1}(s)\sqrt{1-s}h_{\ell,m}(s)\,ds&\\-\alpha_{\ell,m}\psi_{\ell,1}(\rho)+c_{\ell,2}\psi_{\ell,1}(\rho)\int_\rho^1\psi_{\ell,2}(s)h_{\ell,m}(s)\,ds.&\end{aligned}$$

The second and the third term above are obviously smooth up to $\rho=1$; for the first term, the square root factors in fact cancel out, as can easily be seen via the substitution $s=\rho+(1-\rho)t$, and smoothness of $v_{\ell,m}$ up to $\rho=1$ follows. Consequently, the function u_1 defined in (3-11) belongs to $C^\infty(\overline{\mathbb{B}}^9\setminus\{0\})$, and it solves (3-9) in the classical sense away from zero. Furthermore, from (3-15) one can check that $u_{\ell,m}$ and $u'_{\ell,m}$ are bounded near zero, and hence $u_1\in H^1(\mathbb{B}^9)$. In particular, u_1 solves (3-9) in the weak sense on $\mathbb{B}^9$, and since the right-hand side is a smooth function, we conclude that $u_1\in C^\infty(\mathbb{B}^9)$ by elliptic regularity. Consequently, $u_1\in C^\infty(\overline{\mathbb{B}}^9)$, and therefore $u_2\in C^\infty(\overline{\mathbb{B}}^9)$ according to (3-8). In conclusion, $\boldsymbol{u}:=(u_1,u_2)\in\mathcal{D}(\widetilde{\boldsymbol{L}})$ solves (3-6).

For $d = 7$, the same proof can be repeated by choosing $\lambda = \frac{3}{2}$. Namely, by taking the decomposition of the functions into spherical harmonics and by introducing the new variable

$$\tilde{v}_{\ell,m}(\rho) = \rho^2 u_{\ell,m}(\rho),$$

the problem is reduced to

$$\left(-(1-\rho^2)\partial_\rho^2 + \left(-\frac{2}{\rho} + 5\rho\right)\partial_\rho + \frac{(\ell+3)(\ell+2)}{\rho^2} + \frac{15}{4}\right)\tilde{v}_{\ell,m}(\rho) = \rho^2 g_{\ell,m}(\rho),$$

which is the same as (3-13) up to a shift in ℓ and the weight on the right-hand side. Hence the same reasoning applies. □

Proof of Proposition 3.1. Based on Lemmas 3.4 and 3.5, the Lumer–Phillips theorem (see [Engel and Nagel 2000, p. 83, Theorem 3.15]) together with Lemma 3.3 implies that $\widetilde{\boldsymbol{L}}$ is closable in $\mathcal{H}_k$, and that its closure $\boldsymbol{L}_k$ generates a semigroup $(\boldsymbol{S}_k(\tau))_{\tau\geq 0}$ for which (3-1) holds. The rest of the proposition follows from standard semigroup theory results; see, e.g., [Engel and Nagel 2000, p. 55, Theorem 1.10]. □

We conclude this section by proving certain restriction properties of the semigroups $(\boldsymbol{S}_k(\tau))_{\tau\geq 0}$. This will be crucial in showing persistence of regularity for the nonlinear equation.

Lemma 3.6. *Let $d \in \{7, 9\}$ and $k \geq 3$. For any $j \in \mathbb{N}$, the semigroup $(\boldsymbol{S}_{k+j}(\tau))_{\tau\geq 0}$ is the restriction of $(\boldsymbol{S}_k(\tau))_{\tau\geq 0}$ to $\mathcal{H}_{k+j}$, i.e.,*

$$\boldsymbol{S}_{k+j}(\tau) = \boldsymbol{S}_k(\tau)|_{\mathcal{H}_{k+j}}$$

for all $\tau \geq 0$. In particular, we have the growth bound

$$\|\boldsymbol{S}_k(\tau)\boldsymbol{u}\|_{H^{k+j}(\mathbb{B}^d)\times H^{k+j-1}(\mathbb{B}^d)} \lesssim_j e^{-\frac{1}{2}\tau}\|\boldsymbol{u}\|_{H^{k+j}(\mathbb{B}^d)\times H^{k+j-1}(\mathbb{B}^d)}$$

for all $\boldsymbol{u} \in \mathcal{H}_{k+j}$ and all $\tau \geq 0$.

Proof. Let $d \in \{7, 9\}$ and $k \geq 3$. We prove the claim only for $j = 1$, as the general case follows from the arbitrariness of k. The crucial ingredients of the proof are continuity of the embedding $\mathcal{H}_{k+1} \hookrightarrow \mathcal{H}_k$ and the fact that $\mathcal{D}(\widetilde{\boldsymbol{L}})$ is a core for both $\boldsymbol{L}_k$ and $\boldsymbol{L}_{k+1}$. First, we prove that $\boldsymbol{L}_{k+1}$ is a restriction of $\boldsymbol{L}_k$; more precisely we show

$$\mathcal{D}(\boldsymbol{L}_{k+1}) \subset \mathcal{D}(\boldsymbol{L}_k) \quad \text{and} \quad \boldsymbol{L}_{k+1}\boldsymbol{u} = \boldsymbol{L}_k\boldsymbol{u} \tag{3-17}$$

for all $\boldsymbol{u} \in \mathcal{D}(\boldsymbol{L}_{k+1})$. For $\boldsymbol{u} \in \mathcal{D}(\widetilde{\boldsymbol{L}})$, from the definition of $\boldsymbol{L}_{k+1}$ and $\boldsymbol{L}_k$ it follows that

$$\boldsymbol{u} \in \mathcal{D}(\boldsymbol{L}_{k+1}) \cap \mathcal{D}(\boldsymbol{L}_k) \quad \text{and} \quad \boldsymbol{L}_{k+1}\boldsymbol{u} = \boldsymbol{L}_k\boldsymbol{u} = \widetilde{\boldsymbol{L}}\boldsymbol{u}.$$

Let now $\boldsymbol{u} \in \mathcal{D}(\boldsymbol{L}_{k+1})$. Since $(\boldsymbol{L}_{k+1}, \mathcal{D}(\boldsymbol{L}_{k+1}))$ is closed, there exists a sequence $(\boldsymbol{u}_n)_{n\in\mathbb{N}} \subset \mathcal{D}(\widetilde{\boldsymbol{L}})$ such that

$$\boldsymbol{u}_n \xrightarrow{\mathcal{H}_{k+1}} \boldsymbol{u} \quad \text{and} \quad \widetilde{\boldsymbol{L}}\boldsymbol{u}_n \xrightarrow{\mathcal{H}_{k+1}} \boldsymbol{L}_{k+1}\boldsymbol{u}.$$

From the embedding $\mathcal{H}_{k+1} \hookrightarrow \mathcal{H}_k$ we infer

$$\boldsymbol{u}_n \xrightarrow{\mathcal{H}_k} \boldsymbol{u} \quad \text{and} \quad \widetilde{\boldsymbol{L}}\boldsymbol{u}_n \xrightarrow{\mathcal{H}_k} \boldsymbol{L}_{k+1}\boldsymbol{u},$$

and by the closedness of $\boldsymbol{L}_k$ it follows that $\boldsymbol{u} \in \mathcal{D}(\boldsymbol{L}_k)$ and $\boldsymbol{L}_{k+1}\boldsymbol{u} = \boldsymbol{L}_k\boldsymbol{u}$. Now let $\lambda \in \rho(\boldsymbol{L}_{k+1}) \cap \rho(\boldsymbol{L}_k)$. From (3-17) we get that $\boldsymbol{R}_{\boldsymbol{L}_{k+1}}(\lambda) = \boldsymbol{R}_{\boldsymbol{L}_k}(\lambda)|_{\mathcal{H}_{k+1}}$. Now, given $\boldsymbol{u} \in \mathcal{H}_{k+1}$, we get by the Post–Widder inversion formula (see [Engel and Nagel 2000, p. 223, Corollary 5.5]) and the embedding $\mathcal{H}_{k+1} \hookrightarrow \mathcal{H}_k$ that, for every $\tau > 0$,

$$\boldsymbol{S}_{k+1}(\tau)\boldsymbol{u} = \lim_{n\to\infty}\left[\frac{n}{\tau}\boldsymbol{R}_{\boldsymbol{L}_{k+1}}\left(\frac{n}{\tau}\right)\right]^n \boldsymbol{u} = \lim_{n\to\infty}\left[\frac{n}{\tau}\boldsymbol{R}_{\boldsymbol{L}_k}\left(\frac{n}{\tau}\right)\right]^n \boldsymbol{u} = \boldsymbol{S}_k(\tau)\boldsymbol{u}.$$

This proves that $(\boldsymbol{S}_{k+1}(\tau))_{\tau\geq 0}$ is the restriction of $(\boldsymbol{S}_k(\tau))_{\tau\geq 0}$ to $\mathcal{H}_{k+1}$. As a result, from Proposition 3.1 we have

$$\|\boldsymbol{S}_k(\tau)\boldsymbol{u}\|_{H^{k+1}(\mathbb{B}^d)\times H^k(\mathbb{B}^d)} = \|\boldsymbol{S}_{k+1}(\tau)\boldsymbol{u}\|_{H^{k+1}(\mathbb{B}^d)\times H^k(\mathbb{B}^d)} \lesssim e^{-\frac{1}{2}\tau}\|\boldsymbol{u}\|_{H^{k+1}(\mathbb{B}^d)\times H^k(\mathbb{B}^d)}$$

for all $\boldsymbol{u} \in \mathcal{H}_{k+1}$ and all $\tau \geq 0$. □

4. Linearization around a self-similar solution: preliminaries on the structure of the spectrum

From now on, for fixed $d \in \{7, 9\}$, we work solely in the Sobolev space $H^{(d+1)/2}(\mathbb{B}^d) \times H^{(d-1)/2}(\mathbb{B}^d)$, which we earlier denoted by $\mathcal{H}_{(d+1)/2}$. To abbreviate the notation, we write

$$\mathcal{H} := \mathcal{H}_{(d+1)/2}.$$

We also denote by $(\boldsymbol{S}(\tau))_{\tau\geq 0}$ and $\boldsymbol{L} : \mathcal{D}(\boldsymbol{L}) \subset \mathcal{H} \to \mathcal{H}$ the corresponding semigroup $(\boldsymbol{S}_k(\tau))_{\tau\geq 0}$ and its generator $\boldsymbol{L}_k$, respectively, for $k = \frac{1}{2}(d+1)$.

With an eye towards studying the flow near the orbit $\{\boldsymbol{U}_a : a \in \mathbb{R}^d\}$ — see the section on page 628 — in this section we describe some general properties of the underlying linear operator

$$\widetilde{\boldsymbol{L}} + \boldsymbol{L}'_a, \quad \boldsymbol{L}'_a\boldsymbol{u} := \begin{pmatrix} 0 \\ V_a u_1 \end{pmatrix},$$

where

$$V_a(\xi) := 2U_a(\xi), \tag{4-1}$$

with U_a given in (1-12).

Remark 4.1. We emphasize that the results of this section apply to any smooth $V_a : \overline{\mathbb{B}}^d \to \mathbb{R}$ that depends smoothly on the parameter a. Obviously, such potentials arise in the linearization around smooth self-similar profiles.

Proposition 4.2. *Fix $d \in \{7, 9\}$. For every $a \in \mathbb{R}^d$, the operator $\boldsymbol{L}'_a : \mathcal{H} \to \mathcal{H}$ is compact, and the operator*

$$\boldsymbol{L}_a := \boldsymbol{L} + \boldsymbol{L}'_a, \quad \mathcal{D}(\boldsymbol{L}_a) := \mathcal{D}(\boldsymbol{L}) \subset \mathcal{H} \to \mathcal{H},$$

generates a strongly continuous semigroup $\boldsymbol{S}_a : [0, \infty) \to \mathcal{B}(\mathcal{H})$. Furthermore, given $\delta > 0$, there is $K > 0$ such that

$$\|\boldsymbol{L}_a - \boldsymbol{L}_b\| \leq K|a - b|$$

for all $a, b \in \overline{\mathbb{B}}^d_\delta$.

Proof. The compactness of $\boldsymbol{L}_a'$ follows from the smoothness of V_a and the compactness of the embedding $H^{(d+1)/2}(\mathbb{B}^d) \hookrightarrow H^{(d-1)/2}(\mathbb{B}^d)$. The fact that $\boldsymbol{L}_a$ generates a semigroup is a consequence of the bounded perturbation theorem; see, e.g., [Engel and Nagel 2000, p. 158]. For the Lipschitz dependence on the parameter a, we first note that by the fundamental theorem of calculus we have

$$V_a(\xi) - V_b(\xi) = (a^j - b^j) \int_0^1 \partial_{\alpha_j} V_{\alpha(s)}(\xi)\, ds \tag{4-2}$$

for $\alpha(s) = b + s(a-b)$. This implies that, given $\delta > 0$, we have

$$\|V_a - V_b\|_{\dot{H}^k(\mathbb{B}^d)} \lesssim_k |a-b| \tag{4-3}$$

for all $a, b \in \overline{\mathbb{B}}^d_\delta$. In particular,

$$\|V_a - V_b\|_{W^{(d-1)/2,\infty}(\mathbb{B}^d)} \lesssim |a-b|,$$

and we thus have

$$\|(V_a - V_b)u\|_{H^{(d-1)/2}(\mathbb{B}^d)} \lesssim |a-b|\|u\|_{H^{(d-1)/2}(\mathbb{B}^d)} \lesssim |a-b|\|u\|_{H^{(d+1)/2}(\mathbb{B}^d)}$$

for all $u \in C^\infty(\overline{\mathbb{B}}^d)$ and all $a, b \in \overline{\mathbb{B}}^d_\delta$, which implies the claim. □

Next, we show that the unstable spectrum of $\boldsymbol{L}_a : \mathcal{D}(\boldsymbol{L}_a) \subset \mathcal{H} \to \mathcal{H}$ consists of isolated eigenvalues and is confined to a compact region. This is achieved by proving bounds on the resolvent and using compactness of the perturbation.

Proposition 4.3. *Fix $d \in \{7, 9\}$. Let $\varepsilon > 0$ and $\delta > 0$. Then there are constants $\kappa > 0$ and $c > 0$ such that*

$$\|\boldsymbol{R}_{\boldsymbol{L}_a}(\lambda)\| \le c \tag{4-4}$$

for all $a \in \overline{\mathbb{B}}^d_\delta$ and for all $\lambda \in \mathbb{C}$ satisfying $\operatorname{Re}\lambda \ge -\frac{1}{2} + \varepsilon$ and $|\lambda| \ge \kappa$. Furthermore, if $\lambda \in \sigma(\boldsymbol{L}_a)$ with $\operatorname{Re}\lambda > -\frac{1}{2}$, then λ is an isolated eigenvalue.

Proof. Let $\lambda \in \mathbb{C}$ with $\operatorname{Re}\lambda > -\frac{1}{2}$. Then Proposition 3.1 implies that $\lambda \in \rho(\boldsymbol{L})$, and we therefore have the identity

$$\lambda - \boldsymbol{L}_a = [1 - \boldsymbol{L}_a' \boldsymbol{R}_{\boldsymbol{L}}(\lambda)](\lambda - \boldsymbol{L}). \tag{4-5}$$

In what follows we prove that, for suitably chosen λ, the Neumann series $\sum_{k=0}^\infty [\boldsymbol{L}_a' \boldsymbol{R}_{\boldsymbol{L}}(\lambda)]^k$ converges. According to (4-5), this yields

$$\boldsymbol{R}_{\boldsymbol{L}_a}(\lambda) = \boldsymbol{R}_{\boldsymbol{L}}(\lambda) \sum_{k=0}^\infty [\boldsymbol{L}_a' \boldsymbol{R}_{\boldsymbol{L}}(\lambda)]^k,$$

and then (4-4) follows from Proposition 3.1. First, observe that, given $\delta > 0$, we have

$$\|\boldsymbol{L}_a' \boldsymbol{R}_{\boldsymbol{L}}(\lambda)\boldsymbol{f}\| = \|V_a[\boldsymbol{R}_{\boldsymbol{L}}(\lambda)\boldsymbol{f}]_1\|_{H^{(d-1)/2}(\mathbb{B})^d} \lesssim \|[\boldsymbol{R}_{\boldsymbol{L}}(\lambda)\boldsymbol{f}]_1\|_{H^{(d-1)/2}(\mathbb{B}^d)} \tag{4-6}$$

for all $a \in \overline{\mathbb{B}}^d_\delta$ and all $\boldsymbol{f} \in \mathcal{H}$. Now, given $\boldsymbol{f} \in \mathcal{H}$, let $\boldsymbol{u} = \boldsymbol{R}_{\boldsymbol{L}}(\lambda)\boldsymbol{f}$. Since $(\lambda - \boldsymbol{L})\boldsymbol{u} = \boldsymbol{f}$, from the first component of this equation, we get

$$\xi^j \partial_j u_1(\xi) + (\lambda + 2)u_1(\xi) - u_2(\xi) = f_1(\xi)$$

in the weak sense on the ball $\mathbb{B}^d$. Consequently,

$$\|u_1\|_{H^{(d-1)/2}(\mathbb{B}^d)} \lesssim \frac{1}{|\lambda+2|}(\|u_1\|_{H^{(d+1)/2}(\mathbb{B}^d)} + \|u_2\|_{H^{(d-1)/2}(\mathbb{B}^d)} + \|f_1\|_{H^{(d-1)/2}(\mathbb{B}^d)}).$$

Then Proposition 3.1 implies that, given $\varepsilon > 0$,

$$\|[\boldsymbol{R}_{\boldsymbol{L}}(\lambda)\boldsymbol{f}]_1\|_{H^{(d-1)/2}(\mathbb{B}^d)} \lesssim |\lambda|^{-1}(\|\boldsymbol{R}_{\boldsymbol{L}}(\lambda)\boldsymbol{f}\| + \|\boldsymbol{f}\|) \lesssim |\lambda|^{-1}\|\boldsymbol{f}\|$$

for all $\lambda \in \mathbb{C}$ with $\operatorname{Re}\lambda \geq -\frac{1}{2} + \varepsilon$ and all $\boldsymbol{f} \in \mathcal{H}$. Together with (4-6), this gives

$$\|\boldsymbol{L}_a'\boldsymbol{R}_{\boldsymbol{L}}(\lambda)\boldsymbol{f}\| \lesssim |\lambda|^{-1}\|\boldsymbol{f}\|,$$

and the uniform bound (4-4) holds for some $c > 0$ when we restrict to $|\lambda| \geq \kappa$ for suitably large κ. The second statement follows from the compactness of $\boldsymbol{L}_a'$. Indeed, if $\operatorname{Re}\lambda > -\frac{1}{2}$ then $\lambda \in \rho(\boldsymbol{L})$, and according to (4-5) we have that $\lambda \in \sigma(\boldsymbol{L}_a)$ only if $1 - \boldsymbol{L}_a'\boldsymbol{R}_{\boldsymbol{L}}(\lambda)$ is not a bounded invertible operator, which is equivalent to 1 being an eigenvalue of the compact operator $\boldsymbol{L}_a'\boldsymbol{R}_{\boldsymbol{L}}(\lambda)$, which according to (4-5) implies that λ is an eigenvalue of $\boldsymbol{L}_a$. The fact that λ is isolated follows from the analytic Fredholm theorem (see [Simon 2015, Theorem 3.14.3, p. 194]) applied to the mapping $\lambda \mapsto \boldsymbol{L}_a'\boldsymbol{R}_{\boldsymbol{L}}(\lambda)$ defined on $\mathbb{H}_{-1/2} = \{\lambda \in \mathbb{C} : \operatorname{Re}\lambda > -\frac{1}{2}\}$. □

Remark 4.4. The previous proposition implies that there are finitely many unstable spectral points of $\boldsymbol{L}_a$, i.e., the ones belonging to $\overline{\mathbb{H}} := \{\lambda \in \mathbb{C} : \operatorname{Re}\lambda \geq 0\}$, all of which are eigenvalues. This can actually be abstractly shown just by using the compactness of $\boldsymbol{L}_a'$; see [Glogić 2022, Theorem B.1]. We nonetheless need Proposition 4.3 as it allows us later on to reduce the spectral analysis of $\boldsymbol{L}_a$ for all small a to the case $a = 0$; see Section 5C.

Note that the eventual presence of unstable spectral points of $\boldsymbol{L}_a$ prevents decay of the associated semigroup $(\boldsymbol{S}_a(\tau))_{\tau\geq 0}$ on the whole space $\mathcal{H}$. What is more, since $\boldsymbol{L}_a'$ is compact, a spectral mapping theorem for the unstable spectrum holds (see [Glogić 2022, Theorem B.1]), and hence eventual growing modes of $(\boldsymbol{S}_a(\tau))_{\tau\geq 0}$ are completely determined by the unstable spectrum of $\boldsymbol{L}_a$ and the associated eigenspaces. Therefore, in what follows we turn to spectral analysis of $\boldsymbol{L}_a$. First, we show an important result which relates solvability of the spectral equation $(\lambda - \boldsymbol{L}_a)\boldsymbol{u} = 0$ for $a = 0$, $\lambda \in \overline{\mathbb{H}}$, to the existence of *smooth* solutions to a certain ordinary differential equation. We note that, for $a = 0$, the potential V_a is radial; more precisely,

$$V_0(\xi) = 2U_0(\xi) = 2U(|\xi|) =: V(|\xi|),$$

with U given in (1-5).

Proposition 4.5. *Fix $d \in \{7, 9\}$. Let $\lambda \in \mathbb{C}$ with $\operatorname{Re}\lambda \geq 0$. Then $\lambda \in \sigma(\boldsymbol{L}_0)$ if and only if there are $\ell \in \mathbb{N}_0$ and $f \in C^\infty[0,1]$ such that*

$$\mathcal{T}_\ell^{(d)}(\lambda)f(\rho) := (1-\rho^2)f''(\rho) + \left(\frac{d-1}{\rho} - 2(\lambda+3)\rho\right)f'(\rho) - \left((\lambda+2)(\lambda+3) + \frac{\ell(\ell+d-2)}{\rho^2} - V(\rho)\right)f(\rho) = 0 \quad (4\text{-}7)$$

for all $\rho \in (0,1)$.

Proof. Let $\lambda \in \overline{\mathbb{H}} \cap \sigma(\boldsymbol{L}_0)$. By Proposition 4.3, λ is an eigenvalue, and hence there is a nontrivial $\boldsymbol{u} \in \mathcal{D}(\boldsymbol{L}_0)$ satisfying $(\lambda - \boldsymbol{L}_0)\boldsymbol{u} = 0$. By a straightforward calculation, we get that the components u_1 and u_2 satisfy the equations

$$-(\delta^{ij} - \xi^i \xi^j)\partial_i \partial_j u_1(\xi) + 2(\lambda+3)\xi^j \partial_j u_1(\xi) + (\lambda+3)(\lambda+2)u_1(\xi) - V_0(\xi)u_1(\xi) = 0 \tag{4-8}$$

and

$$u_2(\xi) = \xi^j \partial_j u_1(\xi) + (\lambda+2)u_1(\xi) \tag{4-9}$$

weakly on $\mathbb{B}^d$. Since $u_1 \in H^{(d+1)/2}(\mathbb{B}^d)$, we get by elliptic regularity that $u_1 \in C^\infty(\mathbb{B}^d)$. Furthermore, we may take the decomposition of u_1 into spherical harmonics:

$$u_1(\xi) = \sum_{\ell=0}^{\infty} \sum_{m \in \Omega_\ell} (u_1(|\xi|\cdot) \,|\, Y_{\ell,m})_{L^2(\mathbb{S}^{d-1})} Y_{\ell,m}\left(\frac{\xi}{|\xi|}\right) = \sum_{\ell=0}^{\infty} \sum_{m \in \Omega_\ell} u_{\ell,m}(\rho) Y_{\ell,m}(\omega), \tag{4-10}$$

where $\rho = |\xi|$ and $\omega = \xi/|\xi|$. To be precise, the expansion above holds in $H^k(\mathbb{B}^d_{1-\epsilon})$ for any $k \in \mathbb{N}$ and $\epsilon > 0$; see (1-22) and (1-23). Since the potential V_0 is radially symmetric, (4-8) decouples by means of (4-10) into a system of infinitely many ODEs:

$$\mathcal{T}_\ell^{(d)}(\lambda) u_{\ell,m}(\rho) = 0, \tag{4-11}$$

posed on the interval $(0,1)$, where the operator $\mathcal{T}_\ell^{(d)}(\lambda)$ is given by (4-7). Since u_1 is nontrivial, there are indices $\ell \in \mathbb{N}_0$ and $m \in \Omega_\ell$ such that $u_{\ell,m}$ is nonzero and satisfies (4-11). Furthermore, since $u_1 \in C^\infty(\mathbb{B}^d) \cap H^{(d+1)/2}(\mathbb{B}^d)$, we have that $u_{\ell,m} \in C^\infty[0,1) \cap H^{(d+1)/2}\big(\frac{1}{2}, 1\big)$. Now we prove that $u_{\ell,m}$ is smooth up to $\rho = 1$. Note that $\rho = 1$ is a regular singular point of (4-11), and the corresponding set of Frobenius indices is $\{0, 2-\lambda\}$ when $d = 9$, and $\{0, 1-\lambda\}$ when $d = 7$. In the first case, if $\lambda \notin \{0,1,2\}$, then $u_{\ell,m}$ is either analytic or behaves like $(1-\rho)^{2-\lambda}$ near $\rho = 1$. If $\lambda \in \{0,1,2\}$, then the nonanalytic behavior can be described by $(1-\rho)^2 \log(1-\rho)$, $(1-\rho)\log(1-\rho)$, or $\log(1-\rho)$. In each case, singularity can be excluded by the requirement that $u_{\ell,m} \in H^5\big(\frac{1}{2}, 1\big)$. This implies that $u_{\ell,m}$ belongs to $C^\infty[0,1]$ and solves (4-7) on $(0,1)$. The same reasoning applies to the case $d = 7$. Implication in the other direction is now obvious. □

Remark 4.6. Note that Frobenius theory implies that smooth solutions f from Proposition 4.5 are in fact analytic on $[0,1]$, in the sense that they can be extended to an analytic function on an open interval that contains $[0,1]$. Consequently, determining the unstable spectrum of $\boldsymbol{L}_0$ amounts to solving the connection problem for a family of ODEs. We note that the connection problem is so far completely resolved only for hypergeometric equations, i.e., the ones with three regular singular points, while the ODE (4-7) has six of them. In fact, their number can, by a suitable change of variables, be reduced to four, but this nonetheless renders the standard ODE theory useless. Nevertheless, by building on the techniques developed recently to treat such problems (see [Costin et al. 2016; 2017; Glogić 2018; Glogić and Schörkhuber 2021]), for $d = 9$, we are able to solve the connection problem for (4-7) and we thereby provide in the following section a complete characterization of the unstable spectrum of $\boldsymbol{L}_0$.

5. Spectral analysis for perturbations around U_a: the case $d = 9$

From now on we restrict ourselves to $d = 9$.

5A. ***Analysis of the spectral ODE.*** In this section we investigate the ODE (4-7) for $d = 9$, and for convenience we shorten the notation by letting $\mathcal{T}_\ell(\lambda) := \mathcal{T}_\ell^{(9)}(\lambda)$, i.e., we have

$$\mathcal{T}_\ell(\lambda) f(\rho) := (1-\rho^2) f''(\rho) + \left(\frac{8}{\rho} - 2(\lambda+3)\rho\right) f'(\rho) - \left((\lambda+2)(\lambda+3) + \frac{\ell(\ell+7)}{\rho^2} - V(\rho)\right) f(\rho),$$

where the potential is given by

$$V(\rho) = \frac{480(7-\rho^2)}{(7+5\rho^2)^2}.$$

Now, in view of Proposition 4.5, given $\ell \in \mathbb{N}_0$, we define the set

$$\Sigma_\ell := \{\lambda \in \overline{\mathbb{H}} : \text{there exists } f_\ell(\cdot;\lambda) \in C^\infty[0,1] \text{ satisfying } \mathcal{T}_\ell(\lambda) f_\ell(\cdot;\lambda) = 0 \text{ on } (0,1)\}.$$

The central result of our spectral analysis is the following proposition.

Proposition 5.1. *The structure of Σ_ℓ is as follows*:

(1) *For $\ell = 0$, we have $\Sigma_0 = \{1, 3\}$, with corresponding solutions*

$$f_0(\rho;1) = \frac{1-\rho^2}{(7+5\rho^2)^3} \quad \textit{and} \quad f_0(\rho;3) = \frac{1}{(7+5\rho^2)^3},$$

which are unique up to a constant multiple.

(2) *For $\ell = 1$, we have $\Sigma_1 = \{0, 1\}$, and the corresponding solutions are*

$$f_1(\rho;0) = \frac{\rho(7-3\rho^2)}{(7+5\rho^2)^3} \quad \textit{and} \quad f_1(\rho;1) = \frac{\rho(77-5\rho^2)}{(7+5\rho^2)^3}.$$

(3) *For all $\ell \geq 2$, we have $\Sigma_\ell = \varnothing$.*

To prove this proposition, we use an adaptation of the ODE techniques devised in [Costin et al. 2016; 2017; Glogić 2018; Glogić and Schörkhuber 2021]. We will therefore occasionally refer to these works throughout the proof. Also, we found it convenient to split the proof into two cases: $\ell \in \{0, 1\}$ and $\ell \geq 2$.

Proof of Proposition 5.1 for $\ell \in \{0, 1\}$. For a detailed heuristic discussion of our approach we refer the reader to [Glogić and Schörkhuber 2021, Section 4.1]. Namely, the first step is to transform $\mathcal{T}_\ell(\lambda) f(\rho) = 0$ to an "isospectral" equation with four regular singular points. For this, we let $x = \rho^2$, and we define the new dependent variable y via

$$f(\rho) = \rho^\ell \left(\tfrac{7}{5} + \rho^2\right)^{-3} y(\rho^2).$$

This yields the following equation in its canonical Heun form (see [DLMF 2010]):

$$y''(x) + \left(\frac{\gamma(\ell)}{x} + \frac{\delta(\lambda)}{x-1} - \frac{6}{x-\mu}\right) y'(x) + \frac{\alpha(\ell,\lambda)\beta(\ell,\lambda)x - q(\ell,\lambda)}{x(x-1)(x-\mu)} y(x) = 0, \tag{5-1}$$

with singularities at $x \in \{0, 1, \mu, \infty\}$, where $\mu = -\frac{7}{5}$, $\gamma(\ell) = \frac{1}{2}(9 + 2\ell)$, $\delta(\lambda) = \lambda - 1$,

$$\alpha(\ell, \lambda) = \tfrac{1}{2}(\lambda - 3 + \ell), \quad \beta(\ell, \lambda) = \tfrac{1}{2}(\lambda - 4 + \ell),$$
$$q(\ell, \lambda) = -\tfrac{1}{20}(7(\lambda - 3)(\lambda + 8) + 7\ell^2 + (14\lambda + 95)\ell).$$

By Frobenius' theory, any $y \in C^\infty[0, 1]$ that solves (5-1) on $(0, 1)$ is in fact analytic on the closed interval $[0, 1]$. Furthermore, the Frobenius indices of (5-1) at $x = 0$ are $s_1 = 0$ and $s_2 = -\frac{1}{2}(7 + 2\ell)$. Therefore, for every $\lambda \in \mathbb{C}$ there is a unique solution (up to a constant multiple) to (5-1), which is analytic at $x = 0$. Furthermore, this solution has a power series expansion of the form

$$y_{\ell,\lambda}(x) = \sum_{n=0}^{\infty} a_n(\ell, \lambda) x^n, \quad a_0(\ell, \lambda) = 1. \tag{5-2}$$

To determine the coefficients a_n, we insert the ansatz (5-2) into (5-1) and obtain the recurrence relation

$$a_{n+2}(\ell, \lambda) = A_n(\ell, \lambda) a_{n+1}(\ell, \lambda) + B_n(\ell, \lambda) a_n(\ell, \lambda), \tag{5-3}$$

where

$$A_n(\ell, \lambda) = \frac{7\lambda(\lambda + 9) + 7\ell^2 + \ell(8n + 14\lambda + 103) + 8n^2 + 4(7\lambda + 34)n - 40}{14(n + 2)(2\ell + 2n + 11)} \tag{5-4}$$

and

$$B_n(\ell, \lambda) = \frac{5(\lambda + \ell + 2n - 4)(\lambda + \ell + 2n - 3)}{14(n + 2)(2\ell + 2n + 11)}, \tag{5-5}$$

with the initial condition

$$a_{-1}(\ell, \lambda) = 0 \quad \text{and} \quad a_0(\ell, \lambda) = 1. \tag{5-6}$$

Now, note that $\lambda \in \Sigma_\ell$ precisely when the radius of convergence of the series (5-2) is larger than 1. To analyze this radius, we resort to results from the theory of difference equations with variable coefficients. Namely, since

$$\lim_{n\to\infty} A_n(\ell, \lambda) = \tfrac{2}{7} \quad \text{and} \quad \lim_{n\to\infty} B_n(\ell, \lambda) = \tfrac{5}{7},$$

the so-called characteristic equation of (5-3) is

$$t^2 - \tfrac{2}{7}t - \tfrac{5}{7} = 0,$$

and according to Poincaré's theorem (see, for example, [Elaydi 2005, p. 343], or [Glogić and Schörkhuber 2021, Appendix A]) we have that either $a_n(\ell, \lambda) = 0$ eventually in n or

$$\lim_{n\to\infty} \frac{a_{n+1}(\ell, \lambda)}{a_n(\ell, \lambda)} = 1 \quad \text{or} \quad \lim_{n\to\infty} \frac{a_{n+1}(\ell, \lambda)}{a_n(\ell, \lambda)} = -\tfrac{5}{7}.$$

To explore this further, we treat cases $\ell = 0$ and $\ell = 1$ separately.

<u>The case $\ell = 0$</u>. First, we observe that in this case there are explicit polynomial solutions for $\lambda = 1$ and $\lambda = 3$, given by

$$y_{0,1}(x) = 1 - x \quad \text{and} \quad y_{0,3}(x) = 1, \tag{5-7}$$

respectively. These in turn correspond to $f_0(\cdot;1)$ and $f_0(\cdot;3)$, stated in Proposition 5.1. So we have that $\{1,3\}\subset\Sigma_0$. We now show the reverse inclusion. Let $\lambda\in\overline{\mathbb{H}}\setminus\{1,3\}$. Since $\ell=0$, from (5-4) and (5-5) we have

$$A_n(0,\lambda)=\frac{7\lambda(\lambda+9)+8n^2+4(7\lambda+34)n-40}{14(n+2)(2n+11)}\quad\text{and}\quad B_n(0,\lambda)=\frac{5(\lambda+2n-4)(\lambda+2n-3)}{14(n+2)(2n+11)}.$$

Now, note that the assumption that $a_n(0,\lambda)=0$ eventually in n contradicts the initial condition (5-6), as follows by backward substitution. Consequently, we have that either

$$\lim_{n\to\infty}\frac{a_{n+1}(0,\lambda)}{a_n(0,\lambda)}=1, \tag{5-8}$$

or

$$\lim_{n\to\infty}\frac{a_{n+1}(0,\lambda)}{a_n(0,\lambda)}=-\tfrac{5}{7}. \tag{5-9}$$

We prove that (5-8) holds, from which it follows that the radius of convergence of the series (5-2) (that is when $\ell=0$) is 1, and therefore $\lambda\notin\Sigma_0$. To that end, we first compute

$$a_2(0,\lambda)=\tfrac{1}{5544}(\lambda-3)(\lambda-1)(7\lambda^2+126\lambda+680)$$

and

$$a_3(0,\lambda)=\tfrac{1}{3027024}(\lambda-3)(\lambda-1)(49\lambda^4+1519\lambda^3+18494\lambda^2+84224\lambda+46080).$$

Then we define

$$r_2(0,\lambda):=\frac{a_3(0,\lambda)}{a_2(0,\lambda)},$$

where the common factor $(\lambda-3)(\lambda-1)$ (which is an artifact of the existence of the polynomial solutions (5-7)) is canceled, and consequently, according to (5-3), for $n\geq 2$, we let

$$r_{n+1}(0,\lambda)=A_n(0,\lambda)+\frac{B_n(0,\lambda)}{r_n(0,\lambda)}. \tag{5-10}$$

To show (5-8), our strategy is the following. For (5-10) we construct an approximate solution $\tilde{r}_n$ (which we also call a *quasi-solution*) for which $\lim_{n\to\infty}\tilde{r}_n(0,\lambda)=1$ and which is provably close enough to r_n so as to rule out (5-9). The quasi-solution we use is

$$\tilde{r}_n(0,\lambda):=\frac{\lambda^2}{2(2n+9)(n+1)}+\frac{\lambda(4n+9)}{2(2n+9)(n+1)}+\frac{2n+2}{2n+9}. \tag{5-11}$$

We have elaborated on constructing such expressions in [Glogić and Schörkhuber 2021, Section 4.2.2] and in [Costin et al. 2016, Section 4.1]; one can also check [Glogić 2018, Sections 2.6.3 and 2.7.2]. Concerning (5-11), suffice it to say here that we chose a quadratic polynomial in λ with rational coefficients in n so as to emulate the behavior of $r_n(0,\lambda)$ for both large and small values of the participating parameters. To show that the quasi-solution indeed resembles $r_n(0,\lambda)$, we define the relative difference function

$$\delta_n(0,\lambda):=\frac{r_n(0,\lambda)}{\tilde{r}_n(0,\lambda)}-1 \tag{5-12}$$

and show that it is small uniformly in λ and n. To this end we substitute (5-12) into (5-10) and thereby derive the recurrence relation for δ_n:

$$\delta_{n+1}(0,\lambda) = \varepsilon_n(0,\lambda) - C_n(0,\lambda)\frac{\delta_n(0,\lambda)}{1+\delta_n(0,\lambda)}, \tag{5-13}$$

where

$$\varepsilon_n(0,\lambda) = \frac{A_n(0,\lambda)\tilde{r}_n(0,\lambda) + B_n(0,\lambda)}{\tilde{r}_n(0,\lambda)\tilde{r}_{n+1}(0,\lambda)} - 1 \quad \text{and} \quad C_n(0,\lambda) = \frac{B_n(0,\lambda)}{\tilde{r}_n(0,\lambda)\tilde{r}_{n+1}(0,\lambda)}. \tag{5-14}$$

We have the following result.

Lemma 5.2. *For all $n \geq 6$ and $\lambda \in \overline{\mathbb{H}}$, the following estimates hold*:

$$|\delta_6(0,\lambda)| \leq \frac{1}{5}, \quad |\varepsilon_n(0,\lambda)| \leq \frac{3}{140} + \frac{23}{40n}, \quad \textit{and} \quad |C_n(0,\lambda)| \leq \frac{5}{7} - \frac{23}{10n}. \tag{5-15}$$

Note that from (5-15) and (5-13), by a simple induction we infer that $|\delta_n(0,\lambda)| \leq \frac{1}{5}$ for all $n \geq 6$. This then via (5-12) and the fact that $\lim_{n\to\infty} \tilde{r}_n(0,\lambda) = 1$ excludes (5-9), and we are done. It therefore remains to prove the preceding lemma.

Proof. First we show that for $n \geq 6$ the functions $\delta_6(0,\cdot)$, $\varepsilon_n(0,\cdot)$, and $C_n(0,\cdot)$ are analytic in $\overline{\mathbb{H}}$. This, based on (5-12) and (5-14), follows from the fact that the zeros of $\tilde{r}_n(0,\cdot)$ and the poles of $r_6(0,\cdot)$ are all contained in the (open) left half-plane. This is immediate for $\tilde{r}_n(0,\cdot)$ as it is a quadratic polynomial with two negative zeros. As for the zeros of the denominator of $r_6(0,\lambda)$, which is a polynomial of degree 10, this, although it can be proven by elementary means, can be straightforwardly checked by the Routh–Hurwitz stability criterion; see [Glogić and Schörkhuber 2021, Section A.2]. Furthermore, being rational functions, $\delta_6(0,\cdot)$, $\varepsilon_n(0,\cdot)$, and $C_n(0,\cdot)$ are all polynomially bounded in $\overline{\mathbb{H}}$. Therefore, to prove the lemma, it is enough to establish the estimates (5-15) on the imaginary axis only as they can be then extended to all of $\overline{\mathbb{H}}$ by the Phragmen–Lindelöf principle (in its sectorial form); see, e.g., [Titchmarsh 1939, p. 177].

In the following we prove only the third estimate in (5-15), as the first two are shown similarly. We proceed with writing $C_{n+6}(0,\lambda)$ (note the shift in the index) as the ratio of two polynomials $P_1(n,\lambda)$ and $P_2(n,\lambda)$, both of which belong to $\mathbb{Z}[n,\lambda]$. Then, for $t \in \mathbb{R}$, we have the following representation on the imaginary line:

$$|P_j(n,it)|^2 = Q_j(n,t^2)$$

for $j \in \{1,2\}$, where $Q_1(n,t^2) \in \mathbb{Z}[n,t^2]$ and $Q_2(n,t^2) \in \mathbb{N}_0[n,t^2]$. Now the desired estimate is equivalent to

$$\frac{Q_1(n,t^2)}{Q_2(n,t^2)} \leq \left(\frac{5}{7} - \frac{23}{10(n+6)}\right)^2,$$

which is in turn equivalent to

$$(50n+139)^2 Q_2(n,t^2) - (70(n+6))^2 Q_1(n,t^2) \geq 0.$$

Finally, the last inequality trivially holds as the polynomial on the left (when expanded) has manifestly positive coefficients. □

The case $\ell = 1$. We proceed similarly to the previous case, and we therefore only provide the relevant expressions. For $\lambda = 0$ and $\lambda = 1$, we have explicit polynomial solutions

$$y_{1,0}(x) = 1 - \tfrac{3}{7}x \quad \text{and} \quad y_{1,1}(x) = 1 - \tfrac{5}{77}x,$$

respectively, which correspond to $f_1(\cdot\,;0)$ and $f_1(\cdot\,;1)$ from the statement of the proposition. Therefore $\{0,1\} \subset \Sigma_1$, and we proceed by showing that there are no additional elements in Σ_1. Let $\lambda \in \overline{\mathbb{H}} \setminus \{0,1\}$. For $\ell = 1$, the series (5-2) yields a solution to (5-1), which is analytic at $x = 0$. According to (5-3), we have

$$a_{n+2}(1,\lambda) = A_n(1,\lambda)a_{n+1}(1,\lambda) + B_n(1,\lambda)a_n(1,\lambda), \tag{5-16}$$

where

$$A_n(1,\lambda) = \frac{7(\lambda+1)(\lambda+10) + 8n^2 + 4(7\lambda+36)n}{14(n+2)(2n+13)}$$

and

$$B_n(1,\lambda) = \frac{5(\lambda+2n-3)(\lambda+2n-2)}{14(n+2)(2n+13)}.$$

Since

$$a_2(1,\lambda) = \tfrac{1}{8008}\lambda(\lambda-1)(7\lambda^2 + 133\lambda + 786)$$

and

$$a_3(1,\lambda) = \tfrac{1}{720720}\lambda(\lambda-1)(7\lambda^4 + 238\lambda^3 + 3263\lambda^2 + 17828\lambda + 22476),$$

we define

$$r_2(1,\lambda) := \frac{a_3(1,\lambda)}{a_2(1,\lambda)},$$

where the common linear factors are canceled, and according to (5-16) we define r_n for $n \geq 2$ by the recurrence

$$r_{n+1}(1,\lambda) = A_n(1,\lambda) + \frac{B_n(1,\lambda)}{r_n(1,\lambda)}.$$

As a quasi-solution, we let

$$\tilde{r}_n(1,\lambda) := \frac{\lambda^2}{2(2n+11)(n+1)} + \frac{(4n+11)\lambda}{2(2n+11)(n+1)} + \frac{n+1}{n+4},$$

and analogously to the previous case we define $\delta_n(1,\lambda)$, $\varepsilon_n(1,\lambda)$, and $C_n(1,\lambda)$. Also, by the same method as above, we establish the following result.

Lemma 5.3. *For $n = 5$, we have $|\delta_5(1,\lambda)| \leq \frac{1}{5}$. Furthermore, for every $n \geq 5$,*

$$|\varepsilon_n(1,\lambda)| \leq \frac{3}{140} + \frac{5}{8(n+1)} \quad \textit{and} \quad |C_n(1,\lambda)| \leq \frac{5}{7} - \frac{5}{2(n+1)} \tag{5-17}$$

uniformly for all $\lambda \in \overline{\mathbb{H}}$. Consequently, $|\delta_n(1,\lambda)| \leq \frac{1}{5}$ for all $n \geq 5$ and $\lambda \in \overline{\mathbb{H}}$. This implies $\lim_{n\to\infty} r_n(1,\lambda) = 1$.

Proof of Proposition 5.1 for $\ell \geq 2$. Since the parameter ℓ is now free, the analysis is more complicated. Namely, in addition to having to emulate the global behavior in ℓ as well, a quasi-solution also has to approximate the actual solution well enough so as to, with an additional parameter ℓ, obey the estimates analogous to (5-17). We note that a similar problem was treated by the second and the third authors in [Glogić and Schörkhuber 2021, Sections 4.2.1 and 4.2.2], and we closely follow their approach. First, we introduce the change of variable $x = 12\rho^2/(5\rho^2+7)$, by means of which the singular points $\rho = 0$ and $\rho = 1$ remain fixed, while the remaining finite singularity (which corresponds to $\rho = \infty$) is now further away from the unit disk at $x = \frac{12}{5}$. Furthermore, by applying also the transformation

$$f(\rho) = x^{\frac{\ell}{2}}\left(\tfrac{12}{5} - x\right)^{\frac{\lambda+3}{2}} \tilde{y}(x)$$

to $\mathcal{T}_\ell(\lambda) f(\rho) = 0$, we arrive at a Heun equation for $\tilde{y}$:

$$\tilde{y}''(x) + \left(\frac{\tilde{\gamma}(\ell)}{x} + \frac{\tilde{\delta}(\lambda)}{x-1} + \frac{\epsilon}{x-\tilde{\mu}}\right)\tilde{y}'(x) + \frac{\tilde{\alpha}(\ell,\lambda)\tilde{\beta}(\ell,\lambda)x - \tilde{q}(\ell,\lambda)}{x(x-1)(x-\mu)}\tilde{y}(x) = 0, \tag{5-18}$$

where $\tilde{\mu} = \frac{12}{5}$, $\tilde{\gamma}(\ell) = \frac{1}{2}(9+2\ell)$, $\tilde{\delta}(\lambda) = \lambda - 1$, $\epsilon = \frac{3}{2}$, $\tilde{\alpha}(\lambda) = \frac{1}{2}(\lambda - 3 + \ell)$, $\tilde{\beta}(\lambda) = \frac{1}{2}(\lambda + 11 + \ell)$, and

$$\tilde{q}(\ell,\lambda) = \tfrac{1}{20}(17\ell^2 + 2\ell(55+12\lambda) - 7\lambda^2 + 80\lambda - 303).$$

The Frobenius indices of (5-18) at $x = 0$ are $s_1 = 0$ and $s_2 = -\frac{1}{2}(7+2\ell)$. Therefore, we consider the (normalized) analytic solution at $x = 0$:

$$\tilde{y}(x) = \sum_{n=0}^{\infty} \tilde{a}_n(\ell,\lambda)x^n, \quad \tilde{a}_0(\ell,\lambda) = 1. \tag{5-19}$$

Inserting (5-19) into (5-18) yields

$$\tilde{a}_{n+2}(\ell,\lambda) = \tilde{A}_n(\ell,\lambda)\tilde{a}_{n+1}(\ell,\lambda) + \widetilde{B}_n(\ell,\lambda)\tilde{a}_n(\ell,\lambda), \tag{5-20}$$

with

$$\tilde{A}_n(\ell,\lambda) = \frac{68n^2 + (48\lambda + 68\ell + 356)n + 7\lambda^2 + 17\ell^2 + 24\lambda\ell + 128\lambda + 178\ell - 15}{24(n+2)(2n+2\ell+11)}$$

and

$$\widetilde{B}_n(\ell,\lambda) = \frac{-5(2n+\lambda+\ell+11)(2n+\lambda+\ell-3)}{24(n+2)(2n+2\ell+11)},$$

supplied with the initial condition $\tilde{a}_{-1}(\ell,\lambda) = 0$ and $\tilde{a}_0(\ell,\lambda) = 1$. Now, $\lim_{n\to\infty} \tilde{A}_n(\ell,\lambda) = \frac{17}{12}$ and $\lim_{n\to\infty} \widetilde{B}_n(\ell,\lambda) = -\frac{5}{12}$, and consequently the characteristic equation of (5-3) is $t^2 - \frac{17}{12}t + \frac{5}{12} = 0$, with solutions $t_1 = \frac{5}{12}$ and $t_2 = 1$. Hence, for

$$\hat{r}_n(\ell,\lambda) := \frac{\tilde{a}_{n+1}(\ell,\lambda)}{\tilde{a}_n(\ell,\lambda)},$$

either $\tilde{a}_n(\ell,\lambda) = 0$ eventually in n or

$$\lim_{n\to\infty} \hat{r}_n(\ell,\lambda) = 1 \tag{5-21}$$

or

$$\lim_{n\to\infty} \hat{r}_n(\ell,\lambda) = \tfrac{5}{12}. \tag{5-22}$$

Now, for $\lambda \in \overline{\mathbb{H}}$, similarly to the previous cases, we exclude the first option by backward substitution. Then, from (5-20), we derive the recurrence relation for $\hat{r}_n$

$$\hat{r}_{n+1}(\ell,\lambda) = \tilde{A}_n(\ell,\lambda) + \frac{\widetilde{B}_n(\ell,\lambda)}{\hat{r}_n(\ell,\lambda)}, \tag{5-23}$$

along with the initial condition $r_0(\ell,\lambda) = A_{-1}(\ell,\lambda)$. For a quasi-solution to (5-23) we use

$$R_n(\ell,\lambda) := \frac{7\lambda^2}{24(n+1)(2n+2\ell+9)} + \frac{\lambda(6n+3\ell+10)}{3(n+1)(2n+2\ell+9)} + \frac{17\ell}{48(n+1)} + \frac{n-1}{n+1}.$$

Again, for the exact way of constructing such quasi-solutions we refer the reader to [Glogić and Schörkhuber 2021, Section 4.2.2] or [Glogić 2018, Section 2.7.2]. Thereupon we set

$$\tilde{\delta}_n(\ell,\lambda) := \frac{\hat{r}_n(\ell,\lambda)}{R_n(\ell,\lambda)} - 1 \tag{5-24}$$

to obtain

$$\tilde{\delta}_{n+1}(\ell,\lambda) = \tilde{\varepsilon}_n(\ell,\lambda) - \widetilde{C}_n(\ell,\lambda)\frac{\tilde{\delta}_n(\ell,\lambda)}{1+\tilde{\delta}_n(\ell,\lambda)},$$

where

$$\tilde{\varepsilon}_n(\ell,\lambda) = \frac{\tilde{A}_n(\ell,\lambda)R_n(\ell,\lambda) + \widetilde{B}_n(\ell,\lambda)}{R_n(\ell,\lambda)R_{n+1}(\ell,\lambda)} - 1 \quad \text{and} \quad \widetilde{C}_n(\ell,\lambda) = \frac{\widetilde{B}_n(\ell,\lambda)}{R_n(\ell,\lambda)R_{n+1}(\ell,\lambda)}.$$

Now, similarly to the previous cases, we establish the following lemma.

Lemma 5.4. *For all $\ell \geq 2$, $n \geq 3$, and $\lambda \in \overline{\mathbb{H}}$, the following estimates hold*:

$$|\tilde{\delta}_3(\ell,\lambda)| \leq \tfrac{1}{3}, \quad |\tilde{\varepsilon}_n(\ell,\lambda)| \leq \tfrac{1}{8}, \quad \textit{and} \quad |\widetilde{C}_n(\ell,\lambda)| \leq \tfrac{5}{12}.$$

As a consequence, $|\tilde{\delta}_n(\ell,\lambda)| \leq \frac{1}{3}$ for all $n \geq 3$.

From this lemma, (5-24), and the fact that $\lim_{n\to\infty} R_n(\ell,\lambda) = 1$, we exclude (5-22) and we therefore have $\lim_{n\to\infty} \tilde{r}_n(\ell,\lambda) = 1$. Hence, given $\lambda \in \overline{\mathbb{H}}$, there are no solutions to (5-18) which are analytic on $[0,1]$, and consequently $\Sigma_\ell = \varnothing$.

5B. *The spectrum of $\boldsymbol{L}_0$.* With the results from above at hand, we can provide a complete description of the unstable spectrum of $\boldsymbol{L}_0$.

Proposition 5.5. *There exists $\omega_0 \in \left(0, \frac{1}{2}\right]$ such that*

$$\sigma(\boldsymbol{L}_0) \cap \{\lambda \in \mathbb{C} : \operatorname{Re}\lambda > -\omega_0\} = \{\lambda_0, \lambda_1, \lambda_2\}, \tag{5-25}$$

where $\lambda_0 = 0$, $\lambda_1 = 1$, and $\lambda_2 = 3$ are eigenvalues. The geometric eigenspace of λ_2 is spanned by $\boldsymbol{h}_0 = (h_{0,1}, h_{0,2})$, where

$$h_{0,1}(\xi) = \frac{1}{(7+5|\xi|^2)^3} \quad \textit{and} \quad h_{0,2}(\xi) = \xi^i \partial_i h_{0,1}(\xi) + 5h_{0,1}(\xi). \tag{5-26}$$

Moreover, the geometric eigenspaces of λ_1 *and* λ_0 *are spanned by* $\{\boldsymbol{g}_0^{(k)}\}_{k=0}^9 = \{(g_{0,1}^{(k)}, g_{0,2}^{(k)})\}_{k=0}^9$ *and* $\{\boldsymbol{q}_0^{(j)}\}_{j=1}^9 = \{(q_{0,1}^{(j)}, q_{0,2}^{(j)})\}_{j=1}^9$, *respectively, where we have in closed form*

$$\begin{aligned} g_{0,1}^{(0)}(\xi) &= \frac{1-|\xi|^2}{(7+5|\xi|^2)^3}, \qquad g_{0,2}^{(0)}(\xi) = \xi^i \partial_i g_{0,1}^{(0)}(\xi) + 3g_{0,1}^{(0)}(\xi), \\ g_{0,1}^{(j)}(\xi) &= \frac{\xi^j(77-5|\xi|^2)}{(7+5|\xi|^2)^3}, \quad g_{0,2}^{(j)}(\xi) = \xi^i \partial_i g_{0,1}^{(j)}(\xi) + 3g_{0,1}^{(j)}(\xi) \end{aligned} \tag{5-27}$$

for $j = 1, \dots, 9$ *as well as*

$$q_{0,1}^{(j)}(\xi) = \frac{\xi^j(7-3|\xi|^2)}{(7+5|\xi|^2)^3} \quad \textit{and} \quad q_{0,2}^{(j)}(\xi) = \xi^i \partial_i q_{0,1}^{(j)}(\xi) + 2q_{0,1}^{(j)}(\xi). \tag{5-28}$$

Remark 5.6. Recall that $\boldsymbol{U}_a$ solves the stationary equation $\boldsymbol{L}\boldsymbol{U}_a + \boldsymbol{F}(\boldsymbol{U}_a) = 0$. By the chain rule we get, for any $k = 1, \dots, d$,

$$(\boldsymbol{L} + \boldsymbol{F}'(\boldsymbol{U}_a))\partial_{a^k}\boldsymbol{U}_a = \boldsymbol{L}_a \partial_{a^k}\boldsymbol{U}_a = 0.$$

This implies that $\partial_{a^k}\boldsymbol{U}_a$ is an eigenvector of $\boldsymbol{L}_a$ with eigenvalue $\lambda = 0$. In particular, a direct calculation shows that $q_{0,1}^{(j)}(\xi) = c\,\partial_{a^j} U_a(\xi)|_{a=0}$.

Proof. From Propositions 4.3, 4.5, and 5.1 we deduce the existence of $\omega_0 \in \left(0, \frac{1}{2}\right]$ for which (5-25) holds. To determine the eigenspaces, we do the following. First, in view of Proposition 5.1, if $\lambda = 3$ then $\ell = 0$, and setting $u_{0,1}(\rho) = (7+5\rho^2)^{-3}$ in the expansion (4-10) yields (5-26). If $\lambda = 1$, then either $\ell = 0$ and $u_{0,m} = f_0(\cdot\,;1)$, or $\ell = 1$ and $u_{1,m} = f_1(\cdot\,;1)$, for $m = 1, \dots, 9$. Since we can choose $Y_{1,m}(\omega) = \tilde{c}_m \omega_m$ for $m = 1, \dots, 9$, these yield (5-27). For $\lambda = 0$, we have $\ell = 1$ with $u_{1,m} = f_1(\cdot\,,0)$, which similarly leads to (5-28). □

In what follows, we prove that for each unstable eigenvalue the geometric and the algebraic eigenspaces are the same. To this end, we define the associated Riesz projections. Namely, we set

$$\boldsymbol{H}_0 := \frac{1}{2\pi i}\int_{\gamma_2} \boldsymbol{R}_{\boldsymbol{L}_0}(\lambda)\,d\lambda, \quad \boldsymbol{P}_0 := \frac{1}{2\pi i}\int_{\gamma_1} \boldsymbol{R}_{\boldsymbol{L}_0}(\lambda)\,d\lambda, \quad \text{and} \quad \boldsymbol{Q}_0 := \frac{1}{2\pi i}\int_{\gamma_0} \boldsymbol{R}_{\boldsymbol{L}_0}(\lambda)\,d\lambda,$$

where $\gamma_j(s) = \lambda_j + \frac{1}{2}\omega_0 e^{2i\pi s}$ for $s \in [0,1]$ and $j = 0, 1, 2$.

Lemma 5.7. *We have*

$$\dim \operatorname{ran} \boldsymbol{H}_0 = 1, \quad \dim \operatorname{ran} \boldsymbol{P}_0 = 10, \quad \textit{and} \quad \dim \operatorname{ran} \boldsymbol{Q}_0 = 9.$$

Proof. We start with the observation that the ranges of the projections are finite-dimensional. Indeed, λ_j would otherwise belong to the essential spectrum of $\boldsymbol{L}_0$ (see [Kato 1976, Theorems 5.28 and 5.33]) which coincides with the essential spectrum of $\boldsymbol{L}$ (since $\boldsymbol{L}_0$ is a compact perturbation of $\boldsymbol{L}$), but this is in contradiction with (3-2). Now we show that $\dim \operatorname{ran} \boldsymbol{P}_0 = 10$. We know from properties of the Riesz integral that $\ker(\boldsymbol{L}_0 - \lambda_1) \subset \operatorname{ran} \boldsymbol{P}_0$. We therefore only need to prove the reverse inclusion. First, note that the space $\operatorname{ran} \boldsymbol{P}_0$ reduces the operator $\boldsymbol{L}_0$, and we have

$$\sigma(\boldsymbol{L}_0|_{\operatorname{ran} \boldsymbol{P}_0}) = \{1\};$$

see, e.g., [Hislop and Sigal 1996, Proposition 6.9]. Consequently, since $\boldsymbol{P}_0$ is finite-rank, the operator $1-\boldsymbol{L}_0|_{\operatorname{ran}\boldsymbol{P}_0}$ is nilpotent; i.e., there is $m\in\mathbb{N}$ such that $(1-\boldsymbol{L}_0|_{\operatorname{ran}\boldsymbol{P}_0})^m=0$. Note that it suffices to show that $m=1$. We argue by contradiction, and hence assume that $m\geq 2$. Then there is $\boldsymbol{u}\in\mathcal{D}(\boldsymbol{L}_0)$ such that

$$(1-\boldsymbol{L}_0)\boldsymbol{u}=\boldsymbol{v}$$

for a nontrivial $\boldsymbol{v}\in\ker(1-\boldsymbol{L}_0)$. This yields for u_1 the elliptic equation

$$-(\delta^{ij}-\xi^i\xi^j)\partial_i\partial_j u_1(\xi)+2(\lambda+3)\xi^j\partial_j u_1(\xi)+(\lambda+3)(\lambda+2)u_1(\xi)-V_0(\xi)u_1(\xi)=F(\xi), \tag{5-29}$$

where $\lambda=1$ and

$$F(\xi)=\xi^i\partial_i v_1(\xi)+(\lambda+3)v_1(\xi)+v_2(\xi).$$

Since $\boldsymbol{v}\in\ker(1-\boldsymbol{L}_0)=\operatorname{span}(\boldsymbol{g}_0^{(0)},\dots,\boldsymbol{g}_0^{(9)})$, we have that $\boldsymbol{v}=\sum_{k=0}^9\alpha_k\boldsymbol{g}_0^{(k)}$ for some $\alpha_0,\dots,\alpha_9\in\mathbb{C}$, not all of which are zero. To avoid cumbersome notation we let $g_k=g_{0,1}^{(k)}$. In the new notation, based on (5-27), we have

$$F(\xi)=\sum_{k=0}^{9}\alpha_k(2\xi^i\partial_{\xi^i}g_k+7g_k).$$

Furthermore, according to Proposition 5.1 we can rewrite F in polar coordinates as

$$F(\rho\omega)=\alpha_0(2\rho f_0'(\rho)+7f_0(\rho))Y_{0,1}(\omega)+\sum_{i=1}^{9}\alpha_i(2\rho f_1'(\rho)+7f_1(\rho))Y_{1,i}(\omega),$$

where we write $f_0=f_0(\cdot;1)$ and $f_1=f_1(\cdot;1)$. By taking the decomposition of u_1 into spherical harmonics as in (4-10), (5-29) can be written as a system of ODEs:

$$\mathcal{T}_0(1)u_{0,1}=-\alpha_0G_0,\qquad \mathcal{T}_1(1)u_{1,j}=-\alpha_jG_1,\quad j=1,\dots,9, \tag{5-30}$$

posed on the interval $(0,1)$, where $G_i(\rho)=2\rho f_i'(\rho)+7f_i(\rho)$ for $i=0,1$. Moreover, from the properties of u_1, we infer that $u_{\ell,m}\in C^\infty[0,1)\cap H^5\big(\frac{1}{2},1\big)$, and by the Sobolev embedding we have $u_{\ell,m}\in C^2[0,1]$. To obtain a contradiction, we show that if some α_k is nonzero then the corresponding ODE in (5-30) does not admit a $C^2[0,1]$ solution. To start, we assume that $\alpha_0\neq 0$. For convenience, we can without loss of generality assume that $\alpha_0=-1$. Then $u_{0,1}$ solves the ODE

$$(1-\rho^2)u''(\rho)+\Big(\frac{8}{\rho}-8\rho\Big)u'(\rho)-(12-V(\rho))u(\rho)=G_0(\rho), \tag{5-31}$$

where

$$G_0(\rho)=\frac{5\rho^4-102\rho^2+49}{(7+5\rho^2)^4}.$$

Note that

$$u_1(\rho)=f_0(\rho)=\frac{1-\rho^2}{(7+5\rho^2)^3}$$

is a solution to the homogeneous version of (5-31), and by reduction of order we find a second solution:

$$u_2(\rho)=u_1(\rho)\int_{1/2}^{\rho}\frac{ds}{s^8u_1(s)^2}=\frac{1-\rho^2}{(7+5\rho^2)^3}\int_{1/2}^{\rho}\frac{(7+5s^2)^6}{s^8(1-s^2)^2}\,ds.$$

Furthermore, a simple calculation yields

$$u_2(\rho) \simeq \rho^{-7} \quad \text{as } \rho \to 0^+$$

and

$$u_2(\rho) = 864 - 3456(1-\rho)\ln(1-\rho) + O(1-\rho) \quad \text{as } \rho \to 1^-. \tag{5-32}$$

With the fundamental system $\{u_1, u_2\}$ at hand, we can solve (5-31) by the variation of parameters formula. Namely, we have

$$u(\rho) = c_1 u_1(\rho) + c_2 u_2(\rho) - u_1(\rho)\int_0^\rho \frac{u_2(s)G_0(s)s^8}{1-s^2}\,ds + u_2(\rho)\int_0^\rho \frac{u_1(s)G_0(s)s^8}{1-s^2}\,ds$$

for some constants $c_1, c_2 \in \mathbb{C}$. If $u \in C^2[0,1]$, then c_2 must be equal to zero in the above expression, owing to the singular behavior of $u_2(\rho)$ near $\rho = 0$. Then by differentiation we obtain, for $\rho \in (0,1)$,

$$u'(\rho) = c_1 u_1'(\rho) - u_1'(\rho)\int_0^\rho \frac{u_2(s)G_0(s)s^8}{1-s^2}\,ds + u_2'(\rho)\int_0^\rho \frac{u_1(s)G_0(s)s^8}{1-s^2}\,ds.$$

Now we inspect the asymptotic behavior of u' as $\rho \to 1^-$. We first note that u_1' is bounded near $\rho = 1$. Furthermore, note that

$$\int_0^1 \frac{u_1(s)G_0(s)s^8}{1-s^2}\,ds = \int_0^1 \frac{s^2}{1-s^2}\frac{d}{ds}\left[\frac{s^7(1-s^2)^2}{(7+5s^2)^6}\right]ds = -2\int_0^1 \frac{s^8(1-s^2)}{(7+5s^2)^6} =: -C$$

for some $C > 0$, which can be calculated explicitly, and $C < 4\times 10^{-8}$. Hence, based on (5-32), we have

$$u_2'(\rho)\int_0^\rho \frac{u_1(s)G_0(s)s^8}{1-s^2}\,ds \sim -3456\,C\ln(1-\rho) \quad \text{as } \rho \to 1^-.$$

Moreover,

$$-u_1'(\rho)\int_0^\rho \frac{u_2(s)G_0(s)s^8}{1-s^2}\,ds \sim \frac{1}{864}\ln(1-\rho) \quad \text{as } \rho \to 1^-.$$

Finally, we infer that the two integral terms cannot cancel, and thus

$$u'(\rho) \simeq \ln(1-\rho) \quad \text{as } \rho \to 1^-.$$

In conclusion, there is no choice of c_1 and c_2 for which u belongs to $C^2[0,1]$.

We similarly treat α_j for $j \in \{1,\dots,9\}$. It is enough to consider just α_1, and without loss of generality assume that $\alpha_1 = -1$. Then (5-30) yields the ODE

$$(1-\rho^2)u''(\rho) + \left(\frac{8}{\rho} - 8\rho\right)u'(\rho) - \left(12 + \frac{8}{\rho^2} - V(\rho)\right)u(\rho) = G_1(\rho), \tag{5-33}$$

where

$$G_1(\rho) = \frac{\rho(4851 - 1610\rho^2 - 25\rho^4)}{(7+5\rho^2)^4}.$$

Note that

$$u_1(\rho) = f_1(\rho) = \frac{\rho(77-5\rho^2)}{(7+5\rho^2)^3}$$

is a solution for the homogeneous problem. Similarly as above, we obtain another solution by the reduction formula

$$u_2(\rho) = u_1(\rho)\int_1^\rho \frac{ds}{s^8 u_1(s)^2} = \frac{\rho(77-5\rho)}{(7+5\rho^2)^3}\int_1^\rho \frac{(7+5s^2)^6}{s^{10}(77-5s)^2}\,ds,$$

and by inspection of the integral we get $u_2(\rho) \simeq \rho^{-8}$ near the origin and $u_2(\rho) \simeq 1-\rho$ near $\rho = 1$. Now, the general solution of (5-33) on $(0, 1)$ is given by

$$u(\rho) = c_1 u_1(\rho) + c_2 u_2(\rho) - u_1(\rho)\int_0^\rho \frac{u_2(s)G_1(s)s^8}{1-s^2}\,ds + u_2(\rho)\int_0^\rho \frac{u_1(s)G_1(s)s^8}{1-s^2}\,ds. \tag{5-34}$$

Assumption that u belongs to $C^2[0, 1]$ forces $c_2 = 0$ above, due to the singular behavior of u_2 at $\rho = 0$. Furthermore, from the last term in (5-34) we see that $u'(\rho) \simeq \ln(1-\rho)$ as $\rho \to 1^-$. In conclusion, (5-33) admits no $C^2[0, 1]$ solutions, and this finishes the proof for $\boldsymbol{P}_0$.

The remaining two projections are treated similarly, so we omit some details. For $\boldsymbol{H}_0$ we obtain the analogue of (5-29) with

$$F(\xi) = 2\xi^i \partial_i h_{0,1}(\xi) + 11 h_{0,1}(\xi).$$

This leads to the ODE

$$(1-\rho^2)u''(\rho) + \Big(\frac{8}{\rho} - 12\rho\Big)u'(\rho) - (30 - V(\rho))u(\rho) = H(\rho) \tag{5-35}$$

for

$$H(\rho) = \frac{77-5\rho^2}{(7+5\rho^2)^4}.$$

The argument, similarly as above, reduces to showing that (5-35) does not admit $C^2[0, 1]$ solutions. By Proposition 5.1, we have that $u_1(\rho) = (7+5\rho^2)^{-3}$ solves the homogeneous variant of (5-35), with the reduction formula yielding another solution

$$u_2(\rho) = u_1(\rho)\int_{1/2}^\rho \frac{ds}{s^8(1-s^2)^2 u_1(s)^2} = \frac{1}{(7+5\rho^2)^3}\int_{1/2}^\rho \frac{(7+5s^2)^6}{s^8(1-s^2)^2}\,ds. \tag{5-36}$$

Note that u_2 is singular at both $\rho = 0$ and $\rho = 1$; more precisely

$$u_2(\rho) \simeq \rho^{-7} \quad \text{as } \rho \to 0^+ \qquad \text{and} \qquad u_2(\rho) \simeq (1-\rho)^{-1} \quad \text{as } \rho \to 1^-.$$

With u_1 and u_2 at hand, the general solution of (5-35) on the interval $(0, 1)$ can be written as

$$u(\rho) = c_1 u_1(\rho) + c_2 u_2(\rho) - u_1(\rho)\int_0^\rho (1-s^2)s^8 H(s)u_2(s)\,ds + u_2(\rho)\int_0^\rho (1-s^2)s^8 H(s)u_1(s)\,ds,$$

where the parameters $c_1, c_2 \in \mathbb{C}$ are free. The assumption that u is bounded near $\rho = 0$ forces c_2 to equal 0. Note that the first and the third term in (5-36) are bounded near $\rho = 1$. However, due to the singular behavior of u_2, the last term is unbounded near $\rho = 1$, owing to the integrand being strictly positive on $(0, 1)$. In conclusion, the general solution u in (5-36) is unbounded on $(0, 1)$.

Finally, for $\boldsymbol{Q}_0$, we have

$$F(\xi) = \sum_{j=1}^{9} \alpha_j (2\xi^i \partial_{\xi^i} q^j_{0,1}(\xi) + 5q^j_{0,1}(\xi)),$$

and the accompanying analogue of (5-31) is

$$(1-\rho^2)u''(\rho)+\left(\frac{8}{\rho}-6\rho\right)u'(\rho)-\left(6+\frac{8}{\rho^2}-V(\rho)\right)=Q(\rho),$$

where

$$Q(\rho)=\frac{15\rho^5-406\rho^3+343\rho}{(7+5\rho^2)^4}.$$

A fundamental solution set to the homogeneous version of the above ODE is given by

$$u_1(\rho)=\frac{\rho(7-3\rho^2)}{(7+5\rho^2)^3}\quad\text{and}\quad u_2(\rho)=u_1(\rho)\int_1^\rho\frac{1-s^2}{s^8u_1(s)^2}\,ds,$$

and therefore any solution to it on $(0,1)$ can be written as

$$u(\rho)=c_1u_1(\rho)+c_2u_2(\rho)-u_1(\rho)\int_0^\rho\frac{u_2(s)Q(s)s^8}{(1-s^2)^2}\,ds+u_2(\rho)\int_0^\rho\frac{u_1(s)Q(s)s^8}{(1-s^2)^2}\,ds$$

for a choice of $c_1,c_2\in\mathbb{C}$. Again, by similar asymptotic considerations as above, we infer that u'' is necessarily unbounded on $(0,1)$, and this concludes the proof. □

5C. *The spectrum of $\boldsymbol{L_a}$ for $a\neq 0$.* We now investigate the spectrum of $\boldsymbol{L}_a$. In particular, by a perturbative argument we show that, for small a, an analogue of Proposition 5.5 holds for $\boldsymbol{L}_a$ as well.

Lemma 5.8. *There exists $\delta^*>0$ such that, for all $a\in\overline{\mathbb{B}}^9_{\delta^*}$, the following holds:*

$$\sigma(\boldsymbol{L}_a)\cap\left\{\lambda\in\mathbb{C}:\operatorname{Re}\lambda\geq-\tfrac{1}{2}\omega_0\right\}=\{\lambda_0,\lambda_1,\lambda_2\},$$

where ω_0 is the constant from Proposition 5.5 and $\lambda_0=0$, $\lambda_1=1$, and $\lambda_2=3$ are eigenvalues. The geometric eigenspace of λ_2 is spanned by $\boldsymbol{h}_a=(h_{a,1},h_{a,2})$, where

$$h_{a,1}(\xi)=\frac{\gamma(\xi,a)}{(12\gamma(\xi,a)^2+5|\xi|^2-5)^3}\quad\text{and}\quad h_{a,2}(\xi)=\xi^j\partial_jh_{a,1}(\xi)+5h_{a,1}(\xi).$$

Moreover, the geometric eigenspaces of λ_0 and λ_1 are spanned by $\{\boldsymbol{g}_a^{(k)}\}_{k=0}^9=\{(g_{a,1}^{(k)},g_{a,2}^{(k)})\}_{k=0}^9$ and $\{\boldsymbol{q}_a^{(j)}\}_{j=1}^9=\{(q_{a,1}^{(j)},q_{a,2}^{(j)})\}_{j=1}^9$, respectively, where

$$g_{a,1}^{(0)}(\xi)=\frac{(|\xi|^2-1)\gamma(\xi,a)}{(12\gamma(\xi,a)^2+5|\xi|^2-5)^3},\qquad g_{a,2}^{(0)}(\xi)=\xi^j\partial_{\xi^j}g_{a,1}^{(0)}(\xi)+3g_{a,1}^{(0)}(\xi),$$

$$g_{a,1}^{(k)}(\xi)=\frac{(72\gamma(\xi,a)^2+5-5|\xi|^2)\partial_{a_j}\gamma(\xi,a)}{(12\gamma(\xi,a)^2+5|\xi|^2-5)^3},\quad g_{a,2}^{(k)}(\xi)=\xi^j\partial_{\xi^j}g_{a,1}^{(k)}(\xi)+3g_{a,1}^{(k)}(\xi),$$

and

$$q_{a,1}^{(j)}(\xi)=\partial_{a_j}U_a(\xi)\quad\text{and}\quad q_{a,2}^{(j)}(\xi)=\xi^j\partial_jq_{a,1}^{(j)}(\xi)+2q_{a,1}^{(j)}(\xi).$$

Additionally, the eigenfunctions depend Lipschitz continuously on the parameter a, i.e.,

$$\|\boldsymbol{h}_a-\boldsymbol{h}_b\|+\|\boldsymbol{g}_a^{(k)}-\boldsymbol{g}_b^{(k)}\|+\|\boldsymbol{q}_a^{(j)}-\boldsymbol{q}_b^{(j)}\|\lesssim|a-b|$$

for all $a,b\in\overline{\mathbb{B}}^9_{\delta^*}$.

Proof. Let $\varepsilon = -\frac{1}{2}\omega_0 + \frac{1}{2}$ and $\delta > 0$. Then take κ defined by Proposition 4.3, and introduce the sets

$$\Omega = \{z \in \mathbb{C} : \operatorname{Re} z \geq -\tfrac{1}{2}\omega_0 \text{ and } |z| \leq \kappa\}$$

and

$$\widetilde{\Omega} = \{z \in \mathbb{C} : \operatorname{Re} z \geq -\tfrac{1}{2}\omega_0\} \setminus \Omega.$$

Note that Proposition 4.3 implies that $\widetilde{\Omega} \subset \rho(\boldsymbol{L}_a)$ for all $a \in \overline{\mathbb{B}}^9_\delta$. Hence we only need to investigate the spectrum in the compact set Ω. First, note that by Proposition 4.3, the set Ω contains a finite number of eigenvalues. By a direct calculation it can be checked that $\boldsymbol{q}_a^{(j)}$, $\boldsymbol{g}_a^{(k)}$, and $\boldsymbol{h}_a$ are eigenfunctions that correspond to $\lambda_0 = 0$, $\lambda_1 = 1$, and $\lambda_2 = 3$, respectively. Note that we get the explicit expression above simply by Lorentz transforming the corresponding eigenfunctions for $a = 0$. We now show that there are no other eigenvalues in Ω. For this, we utilize the Riesz projection onto the spectrum contained in Ω; see (5-39). This, however, necessitates that $\partial\Omega \subset \rho(\boldsymbol{L}_a)$, and we now show that this holds for small enough a. First, note that for $\lambda \in \partial\Omega$ we have the identity

$$\lambda - \boldsymbol{L}_a = [1 - (\boldsymbol{L}_a' - \boldsymbol{L}_0')\boldsymbol{R}_{\boldsymbol{L}_0}(\lambda)](\lambda - \boldsymbol{L}_0). \tag{5-37}$$

Then, from Proposition 4.2, we have

$$\|\boldsymbol{L}_a' - \boldsymbol{L}_0'\| \|\boldsymbol{R}_{\boldsymbol{L}_0}(\lambda)\| \lesssim |a| \max_{\lambda \in \partial\Omega} \|\boldsymbol{R}_{\boldsymbol{L}_0}(\lambda)\|$$

for all $a \in \overline{\mathbb{B}}^9_\delta$. Therefore, there is small enough $\delta^* > 0$ such that

$$\|\boldsymbol{L}_a' - \boldsymbol{L}_0'\| \|\boldsymbol{R}_{\boldsymbol{L}_0}(\lambda)\| < 1 \tag{5-38}$$

for all $\lambda \in \partial\Omega$ and all $a \in \overline{\mathbb{B}}^9_{\delta^*}$. Now from (5-38) and (5-37) we infer that $\partial\Omega \subset \rho(\boldsymbol{L}_a)$ for all $a \in \overline{\mathbb{B}}^9_{\delta^*}$. Thereupon we define the projection

$$\widetilde{\boldsymbol{T}}_a = \frac{1}{2\pi i} \int_{\partial\Omega} \boldsymbol{R}_{\boldsymbol{L}_a}(\lambda)\, d\lambda. \tag{5-39}$$

For $a = 0$, by Lemma 5.7 the rank of the operator $\widetilde{\boldsymbol{T}}_a$ is 20. Furthermore, continuity of $a \mapsto \boldsymbol{R}_{\boldsymbol{L}_a}(\lambda)$ (which follows from (5-37)) implies continuity of $a \mapsto \widetilde{\boldsymbol{T}}_a$ on $\overline{\mathbb{B}}^9_{\delta^*}$. Thus, we conclude that $\dim \operatorname{ran} \widetilde{\boldsymbol{T}}_a = 20$ for all $a \in \overline{\mathbb{B}}^9_{\delta^*}$; see, e.g., [Kato 1976, p. 34, Lemma 4.10]. By this, we exclude any further eigenvalues. Lipschitz continuity for the eigenfunctions follows from the fact that they depend smoothly on a; see (4-2) and (4-3). □

6. Perturbations around U_a: bounds for the linearized time-evolution

We fix $\delta^* > 0$ as in Lemma 5.8 for the rest of this paper. In this section we propagate Lemma 5.7 to $\boldsymbol{L}_a$. For that, given $a \in \overline{\mathbb{B}}^9_{\delta^*}$, we define the Riesz projections

$$\boldsymbol{H}_a := \frac{1}{2\pi i} \int_{\gamma_2} \boldsymbol{R}_{\boldsymbol{L}_a}(\lambda)\, d\lambda, \quad \boldsymbol{P}_a := \frac{1}{2\pi i} \int_{\gamma_1} \boldsymbol{R}_{\boldsymbol{L}_a}(\lambda)\, d\lambda, \quad \text{and} \quad \boldsymbol{Q}_a := \frac{1}{2\pi i} \int_{\gamma_0} \boldsymbol{R}_{\boldsymbol{L}_a}(\lambda)\, d\lambda,$$

where $\gamma_j(s) = \lambda_j + \frac{1}{4}\omega_0 e^{2\pi i s}$ for $s \in [0, 1]$.

Lemma 6.1. *We have*

$$\operatorname{ran} \boldsymbol{H}_a = \operatorname{span}(\boldsymbol{h}_a), \quad \operatorname{ran} \boldsymbol{P}_a = \operatorname{span}(\boldsymbol{g}_a^{(0)}, \dots, \boldsymbol{g}_a^{(9)}), \quad \textit{and} \quad \operatorname{ran} \boldsymbol{Q}_a = \operatorname{span}(\boldsymbol{q}_a^{(1)}, \dots, \boldsymbol{q}_a^{(9)})$$

for all $a \in \overline{\mathbb{B}}^9_{\delta *}$. *Moreover, the projections are mutually transversal,*

$$\boldsymbol{H}_a \boldsymbol{P}_a = \boldsymbol{P}_a \boldsymbol{H}_a = \boldsymbol{H}_a \boldsymbol{Q}_a = \boldsymbol{Q}_a \boldsymbol{H}_a = \boldsymbol{Q}_a \boldsymbol{P}_a = \boldsymbol{P}_a \boldsymbol{Q}_a = 0,$$

and depend Lipschitz continuously on the parameter a, *i.e.*,

$$\|\boldsymbol{H}_a - \boldsymbol{H}_b\| + \|\boldsymbol{P}_a - \boldsymbol{P}_b\| + \|\boldsymbol{Q}_a - \boldsymbol{Q}_b\| \lesssim |a-b|$$

for all $a, b \in \overline{\mathbb{B}}^9_{\delta *}$.

Proof. The Riesz projections depend continuously on a, hence the dimensions of the ranges remain the same. Transversality follows from the definition of Riesz projections. The Lipschitz bounds follow from the second resolvent identity and Proposition 4.2. □

Since $\boldsymbol{P}_a$ and $\boldsymbol{Q}_a$ are finite-rank, for every $\boldsymbol{f} \in \mathcal{H}$, there are $\alpha^k \in \mathbb{C}$ and $\beta^j \in \mathbb{C}$ such that

$$\boldsymbol{P}_a \boldsymbol{f} = \sum_{k=0}^{9} \alpha^k \boldsymbol{g}_a^{(k)} \quad \text{and} \quad \boldsymbol{Q}_a \boldsymbol{f} = \sum_{j=1}^{9} \beta^j \boldsymbol{q}_a^{(j)}.$$

We thereby define the projections

$$\boldsymbol{P}_a^{(k)} \boldsymbol{f} := \alpha^k \boldsymbol{g}_a^{(k)} \quad \text{and} \quad \boldsymbol{Q}_a^{(j)} \boldsymbol{f} := \beta^j \boldsymbol{q}_a^{(j)}.$$

Clearly, the projections satisfy the identities

$$\boldsymbol{P}_a = \sum_{k=0}^{9} \boldsymbol{P}_a^{(k)}, \quad \boldsymbol{Q}_a = \sum_{j=1}^{9} \boldsymbol{Q}_a^{(j)} \qquad \text{and} \qquad \boldsymbol{P}_a^{(i)} \boldsymbol{P}_a^{(j)} = \delta^{ij} \boldsymbol{P}_a^{(i)}, \quad \boldsymbol{Q}_a^{(k)} \boldsymbol{Q}_a^{(l)} = \delta^{kl} \boldsymbol{Q}_a^{(k)}.$$

We also define

$$\boldsymbol{T}_a := \boldsymbol{I} - \boldsymbol{H}_a - \boldsymbol{P}_a - \boldsymbol{Q}_a.$$

By Lemma 6.1, we have that $\boldsymbol{T}_a$ is Lipschitz continuous with respect to a and the projections $\boldsymbol{T}_a$, $\boldsymbol{H}_a$, $\boldsymbol{P}_a^{(k)}$, and $\boldsymbol{Q}_a^{(j)}$ are mutually transversal. Moreover, the Lipschitz continuity of $\boldsymbol{Q}_a$ and $\boldsymbol{P}_a$ with respect to a implies

$$\|\boldsymbol{Q}_a^{(j)} - \boldsymbol{Q}_b^{(j)}\| \lesssim |a-b|, \quad j = 1, \dots, 9, \quad \text{and} \quad \|\boldsymbol{P}_a^{(k)} - \boldsymbol{P}_b^{(k)}\| \lesssim |a-b|, \quad k = 0, \dots, 9,$$

for all $a, b \in \overline{\mathbb{B}}^9_{\delta *}$. We now describe the interaction of the semigroup $(\boldsymbol{S}_a(\tau))_{\tau \geq 0}$ with these projections.

Proposition 6.2. *The projection operators* $\boldsymbol{H}_a$, $\boldsymbol{P}_a^{(k)}$, *and* $\boldsymbol{Q}_a^{(j)}$ *commute with the semigroup* $\boldsymbol{S}_a(\tau)$, *i.e.*,

$$[\boldsymbol{S}_a(\tau), \boldsymbol{H}_a] = [\boldsymbol{S}_a(\tau), \boldsymbol{P}_a^{(k)}] = [\boldsymbol{S}_a(\tau), \boldsymbol{Q}_a^{(j)}] = 0 \tag{6-1}$$

for $j = 1, \dots, 9$, $k = 0, \dots, 9$, *and* $\tau \geq 0$. *Furthermore,*

$$\boldsymbol{S}_a(\tau) \boldsymbol{H}_a = e^{3\tau} \boldsymbol{H}_a, \quad \boldsymbol{S}_a(\tau) \boldsymbol{P}_a^{(k)} = e^{\tau} \boldsymbol{P}_a^{(k)}, \quad \textit{and} \quad \boldsymbol{S}_a(\tau) \boldsymbol{Q}_a^{(j)} = \boldsymbol{Q}_a^{(j)}, \tag{6-2}$$

and there exists $\omega > 0$ such that

$$\|\boldsymbol{S}_a(\tau)\boldsymbol{T}_a\boldsymbol{u}\| \lesssim e^{-\omega\tau}\|\boldsymbol{T}_a\boldsymbol{u}\| \tag{6-3}$$

for all $\boldsymbol{u} \in \mathcal{H}$, $a \in \overline{\mathbb{B}}^9_{\delta*}$, and $\tau \geq 0$. Moreover, we have

$$\|\boldsymbol{S}_a(\tau)\boldsymbol{T}_a - \boldsymbol{S}_b(\tau)\boldsymbol{T}_b\| \lesssim e^{-\omega\tau}|a-b| \tag{6-4}$$

for all $a, b \in \overline{\mathbb{B}}^9_{\delta*}$ and $\tau \geq 0$.

Proof. Equation (6-1) follows from the properties of the Riesz projections $\boldsymbol{H}_a$, $\boldsymbol{P}_a$, and $\boldsymbol{Q}_a$. In particular, they commute with $\boldsymbol{S}_a(\tau)$, and this yields, for example, that

$$\boldsymbol{P}_a^{(k)}\boldsymbol{S}_a(\tau)\boldsymbol{u} = \boldsymbol{P}_a\boldsymbol{P}_a^{(k)}\boldsymbol{S}_a(\tau)\boldsymbol{u} = \boldsymbol{P}_a^{(k)}\boldsymbol{S}_a\boldsymbol{P}_a(\tau)\boldsymbol{u} = e^{\tau}\boldsymbol{P}_a^{(k)}\boldsymbol{P}_a\boldsymbol{u} = \boldsymbol{S}_a(\tau)\boldsymbol{P}_a^{(k)}\boldsymbol{u}.$$

Equation (6-2) follows from the correspondence between point spectra of a semigroup and its generator. Equation (6-3) follows from Gearhart–Prüss theorem. More precisely, we have that $\operatorname{ran}\boldsymbol{T}_a$ reduces both $\boldsymbol{L}_a$ and $\boldsymbol{S}_a(\tau)$, and furthermore

$$\boldsymbol{R}_{\boldsymbol{L}_a|_{\operatorname{ran}\boldsymbol{T}_a}}(\lambda) \text{ exists in } \left\{z \in \mathbb{C} : \operatorname{Re} z \geq -\tfrac{1}{2}\omega_0\right\}$$

and is uniformly bounded there, i.e., according to Proposition 4.3 there exists $c > 0$ such that

$$\|\boldsymbol{R}_{\boldsymbol{L}_a|_{\operatorname{ran}\boldsymbol{T}_a}}(\lambda)\| \leq c$$

for all $\operatorname{Re}\lambda \geq -\frac{1}{2}\omega_0$ and all $a \in \overline{\mathbb{B}}^9_{\delta*}$. Hence, by the Gearhart–Prüss theorem (see [Engel and Nagel 2000, p. 302, Theorem 1.11]), for every $\varepsilon > 0$, we have

$$\|\boldsymbol{S}_a(\tau)|_{\operatorname{ran}\boldsymbol{T}_a}\| \lesssim_\varepsilon e^{-\left(\frac{\omega_0}{2}-\varepsilon\right)\tau} \tag{6-5}$$

for all $a \in \overline{\mathbb{B}}^9_{\delta*}$ and $\tau \geq 0$. From here (6-3) holds for any $\omega < \frac{1}{2}\omega_0$. We remark in passing that (6-3) also follows from purely abstract considerations; see [Glogić 2022, Theorem B.1]. Finally, to obtain (6-4) we do the following. First, for $\boldsymbol{u} \in \mathcal{D}(\boldsymbol{L}_a)$ we define the function

$$\Phi_{a,b}(\tau) = \frac{\boldsymbol{S}_a(\tau)\boldsymbol{T}_a\boldsymbol{u} - \boldsymbol{S}_b(\tau)\boldsymbol{T}_b\boldsymbol{u}}{|a-b|}.$$

Note that this function satisfies the evolution equation

$$\partial_\tau\Phi_{a,b}(\tau) = \boldsymbol{L}_a\boldsymbol{T}_a\Phi_{a,b}(\tau) + \frac{\boldsymbol{L}_a\boldsymbol{T}_a - \boldsymbol{L}_b\boldsymbol{T}_b}{|a-b|}\boldsymbol{S}_b(\tau)\boldsymbol{T}_b\boldsymbol{u}$$

with the initial condition

$$\Phi_{a,b}(0) = \frac{\boldsymbol{T}_a\boldsymbol{u} - \boldsymbol{T}_b\boldsymbol{u}}{|a-b|},$$

and therefore by Duhamel's principle we have

$$\Phi_{a,b}(\tau) = \boldsymbol{S}_a(\tau)\boldsymbol{T}_a\frac{\boldsymbol{T}_a\boldsymbol{u} - \boldsymbol{T}_b\boldsymbol{u}}{|a-b|} + \int_0^\tau \boldsymbol{S}_a(\tau-\tau')\boldsymbol{T}_a\frac{\boldsymbol{L}_a\boldsymbol{T}_a - \boldsymbol{L}_b\boldsymbol{T}_b}{|a-b|}\boldsymbol{S}_b(\tau')\boldsymbol{T}_b\boldsymbol{u}\,d\tau'.$$

Now, from Proposition 4.2 and Lemma 6.1, we get

$$\|\boldsymbol{L}_a\boldsymbol{T}_a - \boldsymbol{L}_b\boldsymbol{T}_b\| \lesssim |a-b|,$$

and from this and (6-5) we obtain

$$\|\Phi_{a,b}(\tau)\| \lesssim e^{-(\frac{\omega_0}{2}-\varepsilon)\tau}(1+\tau)\|\boldsymbol{u}\| \lesssim e^{-(\frac{\omega_0}{2}-2\varepsilon)}\|\boldsymbol{u}\|.$$

By choosing $\varepsilon > 0$ such that $\omega = \frac{1}{2}\omega_0 - 2\varepsilon > 0$, we conclude the proof. □

7. Nonlinear theory

With the linear theory at hand, we turn to studying the Cauchy problem for the nonlinear equation (2-9). Following the usual approach of first constructing strong solutions, we recast (2-9) in an integral form à la Duhamel,

$$\Phi(\tau) = \boldsymbol{S}_{a_\infty}(\tau)\Phi(0) + \int_0^\tau \boldsymbol{S}_{a_\infty}(\tau-\sigma)(\boldsymbol{G}_{a(\sigma)}(\Phi(\sigma)) - \partial_\sigma \boldsymbol{U}_{a(\sigma)})\,d\sigma \tag{7-1}$$

(where $(\boldsymbol{S}_{a_\infty}(\tau))_{\tau\geq 0}$ is the semigroup generated by $\boldsymbol{L}_{a_\infty}$), and resort to fixed-point arguments. Our aim is to construct global and decaying solutions to (7-1). An obvious obstruction to that is the presence of growing modes of $\boldsymbol{S}_{a_\infty}(\tau)$, see (6-2), and we deal with them in the following way. First, we note that the instabilities coming from $\boldsymbol{Q}_{a_\infty}$ and $\boldsymbol{P}_{a_\infty}$ are not genuine, as they arise from the Lorentz and space-time translation symmetries of (1-1).

We take care of the Lorentz instability by modulation. Namely, the presence of the unstable space ran $\boldsymbol{Q}_{a_\infty}$ is related to the freedom of choice of the function $a : [0,\infty) \mapsto \mathbb{R}^9$ in the ansatz (2-8), and, roughly speaking, we prove that given small enough initial data $\Phi(0)$, there is a way to choose a such that it leads to a solution Φ of (7-1) which eventually (in τ) gets stripped of any remnant of the unstable space ran $\boldsymbol{Q}_{a_\infty}$ brought about by initial data.

With the rest of the instabilities, which cause exponential growth, we deal differently. Namely, we introduce to the initial data suitable correction terms which serve to suppress the growth. Also, as mentioned, the unstable space ran $\boldsymbol{P}_{a_\infty}$ is another apparent instability as it is an artifact of the spacetime translation symmetries, and we use it to prove that the corrections corresponding to $\boldsymbol{P}_{a_\infty}$ can be annihilated by a proper choice of the parameters x_0 and T, which appear in the initial data $\Phi(0)$; see (2-6). The remaining instability, coming from $\boldsymbol{H}_{a_\infty}$, is the only genuine one, and the correction corresponding to it is reflected in the modification of the initial data in the main result; see (1-16).

To formalize the process described above, we first make some technical preparations. For the rest of this paper, we fix $\omega > 0$ from Proposition 6.2. Then, we introduce the function spaces

$$\mathcal{X} := \{\Phi \in C([0,\infty),\mathcal{H}) : \|\Phi\|_{\mathcal{X}} < \infty\}, \quad \text{where } \|\Phi\|_{\mathcal{X}} := \sup_{\tau>0} e^{\omega\tau}\|\Phi(\tau)\|,$$

and

$$X := \{a \in C^1([0,\infty),\mathbb{R}^9) : a(0) = 0,\ \|a\|_X < \infty\}, \quad \text{where } \|a\|_X := \sup_{\tau>0}[e^{\omega\tau}|\dot{a}(\tau)| + |a(\tau)|].$$

For $a \in X$, we define

$$a_\infty := \lim_{\tau\to\infty} a(\tau).$$

Furthermore, for $\delta > 0$, we set

$$\mathcal{X}_\delta := \{\Phi \in \mathcal{X} : \|\Phi\|_{\mathcal{X}} \le \delta\} \quad \text{and} \quad X_\delta := \Big\{a \in X : \sup_{\tau>0}[e^{\omega\tau}|\dot a(\tau)|] \le \delta\Big\}.$$

To ensure that all terms in (7-1) are defined, we must impose some size restriction on the function a. Note that it is enough to consider $a \in X_\delta$ for $\delta < \delta^*\omega$, as then $|a(\tau)| \le \delta/\omega < \delta^*$ for all $\tau \ge 0$. We will also frequently make use of the inequality

$$|a_\infty - a(\tau)| \le \int_\tau^\infty |\dot a(\sigma)|\, d\sigma \le \frac{\delta}{\omega} e^{-\omega\tau}. \tag{7-2}$$

Furthermore, note that, for $a, b \in X_\delta$ and $\tau \ge 0$, we have $|a(\tau) - b(\tau)| \le \|a - b\|_X$; in particular, we have $|a_\infty - b_\infty| \le \|a - b\|_X$.

7A. ***Estimates of the nonlinear terms.*** With an eye toward setting up a fixed-point scheme for (7-1), we now establish necessary bounds for the nonlinear terms. Namely, we treat

$$\boldsymbol{G}_{a(\tau)}(\Phi(\tau)) = [\boldsymbol{L}'_{a(\tau)} - \boldsymbol{L}'_{a_\infty}]\Phi(\tau) + \boldsymbol{F}(\Phi(\tau)).$$

Lemma 7.1. *Given $\delta \in (0, \delta^*\omega)$, we have*

$$\begin{aligned} \|\boldsymbol{G}_{a(\tau)}(\Phi(\tau))\| &\lesssim \delta^2 e^{-2\omega\tau}, \\ \|\boldsymbol{G}_{a(\tau)}(\Phi(\tau)) - \boldsymbol{G}_{b(\tau)}(\Psi(\tau))\| &\lesssim \delta e^{-2\omega\tau}(\|\Phi - \Psi\|_{\mathcal{X}} + \|a - b\|_X) \end{aligned} \tag{7-3}$$

for all $\Phi, \Psi \in \mathcal{X}_\delta$, $a, b \in X_\delta$, and $\tau \ge 0$, where the implicit constants in the above estimates are absolute.

Proof. First, since $H^5(\mathbb{B}^9)$ is a Banach algebra, we have that

$$\|u_1^2 - v_1^2\|_{H^4(\mathbb{B}^9)} \lesssim \|u_1 + v_1\|_{H^5}\|u_1 - v_1\|_{H^5},$$

and hence

$$\|\boldsymbol{F}(\boldsymbol{u}) - \boldsymbol{F}(\boldsymbol{v})\| \lesssim (\|\boldsymbol{u}\| + \|\boldsymbol{v}\|)\|\boldsymbol{u} - \boldsymbol{v}\| \tag{7-4}$$

for all $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{H}$. Next, we prove the second estimate in Lemma 7.1, as the first one follows from it. From (7-4), Proposition 4.2, and inequality (7-2), we obtain

$$\begin{aligned} \|\boldsymbol{F}(\Phi(\tau)) - \boldsymbol{F}(\Psi(\tau))\| &\lesssim \delta e^{-2\omega\tau}\|\Phi - \Psi\|_{\mathcal{X}}, \\ \|[\boldsymbol{L}'_{a(\tau)} - \boldsymbol{L}'_{a_\infty}](\Phi(\tau) - \Psi(\tau))\| &\lesssim \delta e^{-2\omega\tau}\|\Phi - \Psi\|_{\mathcal{X}} \end{aligned} \tag{7-5}$$

for $\Phi, \Psi \in \mathcal{X}_\delta$ and $a \in X_\delta$. Furthermore, using the fact that

$$V_{a_\infty}(\xi) - V_{a(\tau)}(\xi) = \int_\tau^\infty \partial_s V_{a(s)}(\xi)\, ds = \int_\tau^\infty \dot a^k(s)\varphi_{a(s),k}(\xi)\, ds, \tag{7-6}$$

with $\varphi_{a,k}(\xi) = \partial_{a^k} V_a(\xi)$, together with the smoothness of $\varphi_{a,k}$, we infer

$$\begin{aligned}\|([\boldsymbol{L}'_{a(\tau)}-\boldsymbol{L}'_{a_\infty}]-[\boldsymbol{L}'_{b(\tau)}-\boldsymbol{L}'_{b_\infty}])\boldsymbol{u}\| &\lesssim \|u_1\|_{H^4(\mathbb{B}^9)}\int_\tau^\infty \|\dot{a}^k(s)\varphi_{a(s),k}(\xi)-\dot{b}^k(s)\varphi_{b(s),k}(\xi)\|_{W^{4,\infty}(\mathbb{B}^9)}\,ds\\ &\lesssim \|\boldsymbol{u}\|\int_\tau^\infty |\dot{a}(s)-\dot{b}(s)|\,ds+\|\boldsymbol{u}\|\int_\tau^\infty |\dot{a}(s)|\,|a(s)-b(s)|\,ds\\ &\lesssim \|\boldsymbol{u}\|\int_\tau^\infty e^{-\omega s}\|a-b\|_X\,ds.\end{aligned}$$

Hence

$$\|([\boldsymbol{L}'_{a(\tau)}-\boldsymbol{L}'_{a_\infty}]-[\boldsymbol{L}'_{b(\tau)}-\boldsymbol{L}'_{b_\infty}])\Psi(\tau)\| \lesssim \delta e^{-2\omega\tau}\|a-b\|_X$$

for $a, b \in X_\delta$ and $\Psi \in \mathcal{X}_\delta$, and this together with (7-5) concludes the proof. □

7B. ***Suppressing the instabilities.*** In this section we formalize the process of taming the instabilities. In particular, by introducing correction terms to the initial data we arrive at a modified equation, to which we prove existence of global and decaying solutions.

We first derive the so-called modulation equation for the parameter a. Recall that

$$\partial_\tau \boldsymbol{U}_{a(\tau)} = \dot{a}_j(\tau)\boldsymbol{q}^{(j)}_{a(\tau)} = \sum_{j=1}^{9}\dot{a}^j(\tau)\boldsymbol{q}^{(j)}_{a(\tau)};$$

see Remark 5.6. We introduce a smooth cut-off function $\chi : [0,\infty) \to [0,1]$ satisfying $\chi(\tau) = 1$ for $\tau \in [0,1]$, $\chi(\tau) = 0$ for $\tau \geq 4$, and $|\chi'(\tau)| \leq 1$ for all $\tau \in (0,\infty)$. The aim is to construct a function $a : [0,\infty) \mapsto \mathbb{R}^9$ such that it yields a solution Φ to (7-1) for which

$$\boldsymbol{Q}^{(j)}_{a_\infty}\Phi(\tau) = \chi(\tau)\boldsymbol{Q}^{(j)}_{a_\infty}\Phi(0) \tag{7-7}$$

for all $\tau \geq 0$. In that case, although $\boldsymbol{Q}^{(j)}_{a_\infty}\Phi(0) \neq 0$ in general, we have that $\boldsymbol{Q}^{(j)}_{a_\infty}\Phi(\tau) = 0$ eventually in τ. According to (7-1) and Proposition 6.2, (7-7) adopts the form

$$(1-\chi(\tau))\boldsymbol{Q}^{(j)}_{a_\infty}\boldsymbol{u} + \int_0^\tau (\boldsymbol{Q}^{(j)}_{a_\infty}\boldsymbol{G}_{a(\sigma)}(\Phi(\sigma)) - \boldsymbol{Q}^{(j)}_{a_\infty}\dot{a}_i(\sigma)\boldsymbol{q}^{i}_{a(\sigma)})\,d\sigma = 0,$$

where for convenience we write $\boldsymbol{u}$ instead of $\Phi(0)$. Using $\boldsymbol{Q}^{(j)}_{a_\infty}\boldsymbol{q}^{(i)}_{a_\infty} = \delta^{ij}\boldsymbol{q}^{(j)}_{a_\infty}$, we get the modulation equation

$$a^j(\tau)\boldsymbol{q}^{(j)}_{a_\infty} = -\int_0^\tau \chi'(\sigma)\boldsymbol{Q}^{(j)}_{a_\infty}\boldsymbol{u}\,d\sigma + \int_0^\tau (\boldsymbol{Q}^{(j)}_{a_\infty}\boldsymbol{G}_{a(\sigma)}(\Phi(\sigma)) - \boldsymbol{Q}^{(j)}_{a_\infty}\dot{a}_i(\sigma)(\boldsymbol{q}^{(i)}_{a(\sigma)} - \boldsymbol{q}^{(i)}_{a_\infty}))\,d\sigma$$

for $j = 1,\dots,9$. By introducing the notation

$$\boldsymbol{A}_j(a,\Phi,\boldsymbol{u})(\sigma) := \chi'(\sigma)\boldsymbol{Q}^{(j)}_{a_\infty}\boldsymbol{u} + (\boldsymbol{Q}^{(j)}_{a_\infty}\boldsymbol{G}_{a(\sigma)}(\Phi(\sigma)) - \boldsymbol{Q}^{(j)}_{a_\infty}\dot{a}_i(\sigma)(\boldsymbol{q}^{(i)}_{a(\sigma)} - \boldsymbol{q}^{(i)}_{a_\infty})),$$

the modulation equation can be written succinctly as

$$a_j(\tau) = A_j(\cdot,\Phi,\boldsymbol{u}) := \|\boldsymbol{q}^{(j)}_{a_\infty}\|^{-2}\int_0^\tau (\boldsymbol{A}_j(a,\Phi,\boldsymbol{u})(\sigma)\,|\,\boldsymbol{q}^{(j)}_{a_\infty})\,d\sigma, \quad j = 1,\dots,9. \tag{7-8}$$

In the following we prove that, for small enough Φ and $\boldsymbol{u}$, the system (7-8) admits a global (in τ) solution.

Lemma 7.2. *For all sufficiently small $\delta > 0$ and all sufficiently large $C > 0$, the following holds: For every $\boldsymbol{u} \in \mathcal{H}$ satisfying $\|\boldsymbol{u}\| \le \delta/C$ and every $\Phi \in \mathcal{X}_\delta$, there exists a unique $a = a(\Phi, \boldsymbol{u}) \in X_\delta$ such that (7-8) holds for $\tau \ge 0$. Moreover,*

$$\|a(\Phi, \boldsymbol{u}) - a(\Psi, \boldsymbol{v})\|_X \lesssim \|\Phi - \Psi\|_{\mathcal{X}} + \|\boldsymbol{u} - \boldsymbol{v}\| \tag{7-9}$$

for all $\Phi, \Psi \in \mathcal{X}_\delta$ and $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{B}_{\delta/C}$.

Proof. We use a fixed-point argument. Using the bounds from Lemma 7.1, one can show that, given $\boldsymbol{u}$ and Φ that satisfy the above assumptions, the following estimates hold:

$$\|\boldsymbol{A}_j(a, \Phi, \boldsymbol{u})(\tau)\| \lesssim \Big(\frac{\delta}{C} + \delta^2\Big)e^{-2\omega\tau} \quad \text{and} \quad \|\boldsymbol{A}_j(a, \Phi, \boldsymbol{u})(\tau) - \boldsymbol{A}_j(b, \Phi, \boldsymbol{u})(\tau)\| \lesssim \delta e^{-\omega\tau}\|a - b\|_X$$

for all $a, b \in X_\delta$. From here, according to the definition in (7-8), we have that, for all small enough $\delta > 0$ and all large enough $C > 0$, given $\Phi \in \mathcal{X}_\delta$ and $\boldsymbol{u} \in \mathcal{B}_{\delta/C}$, the ball X_δ is invariant under the action of the operator $A(\cdot, \Phi, \boldsymbol{u})$, which is furthermore a contraction on X_δ. Hence (7-8) has a unique solution in X_δ. The Lipschitz continuity of the solution map follows from the estimate

$$\begin{aligned}\|a - b\|_X &\le \|A(a, \Phi, \boldsymbol{u}) - A(b, \Phi, \boldsymbol{u})\|_X + \|A(b, \Phi, \boldsymbol{u}) - A(b, \Phi, \boldsymbol{v})\|_X + \|A(b, \Phi, \boldsymbol{v}) - A(b, \Psi, \boldsymbol{v})\|_X \\ &\lesssim \delta\|a - b\|_X + \|\boldsymbol{u} - \boldsymbol{v}\| + \|\Phi - \Psi\|_{\mathcal{X}}\end{aligned}$$

by taking small enough $\delta > 0$. □

For the remaining instabilities, we introduce the correction terms

$$\begin{aligned}\boldsymbol{C}_1(\Phi, a, \boldsymbol{u}) &:= \boldsymbol{P}_{a_\infty}\Big(\boldsymbol{u} + \int_0^\infty e^{-\sigma}(\boldsymbol{G}_{a(\sigma)}(\Phi(\sigma)) - \partial_\sigma \boldsymbol{U}_{a(\sigma)})\, d\sigma\Big), \\ \boldsymbol{C}_2(\Phi, a, \boldsymbol{u}) &:= \boldsymbol{H}_{a_\infty}\Big(\boldsymbol{u} + \int_0^\infty e^{-3\sigma}(\boldsymbol{G}_{a(\sigma)}(\Phi(\sigma)) - \partial_\sigma \boldsymbol{U}_{a(\sigma)})\, d\sigma\Big),\end{aligned}$$

and set $\boldsymbol{C} := \boldsymbol{C}_1 + \boldsymbol{C}_2$. Consequently, we investigate the modified integral equation

$$\begin{aligned}\Phi(\tau) &= \boldsymbol{S}_{a_\infty}(\tau)(\boldsymbol{u} - \boldsymbol{C}(\Phi, a, \boldsymbol{u})) + \int_0^\tau \boldsymbol{S}_{a_\infty}(\tau - \sigma)(\boldsymbol{G}_{a(\sigma)}(\Phi(\sigma)) - \partial_\sigma \boldsymbol{U}_{a(\sigma)})\, d\sigma \\ &=: \boldsymbol{K}(\Phi, a, \boldsymbol{u})(\tau).\end{aligned} \tag{7-10}$$

Proposition 7.3. *For all sufficiently small $\delta > 0$ and all sufficiently large $C > 0$, the following holds: For every $\boldsymbol{u} \in \mathcal{H}$ with $\|\boldsymbol{u}\| \le \delta/C$ there exist functions $\Phi \in \mathcal{X}_\delta$ and $a \in X_\delta$ such that (7-10) holds for $\tau \ge 0$. Furthermore, the solution map $\boldsymbol{u} \mapsto (\Phi(\boldsymbol{u}), a(\boldsymbol{u}))$ is Lipschitz continuous, i.e.,*

$$\|\Phi(\boldsymbol{u}) - \Phi(\boldsymbol{v})\|_{\mathcal{X}} + \|a(\boldsymbol{u}) - a(\boldsymbol{v})\|_X \lesssim \|\boldsymbol{u} - \boldsymbol{v}\| \tag{7-11}$$

for all $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{B}_{\delta/C}$.

Proof. We choose $C > 0$ and $\delta > 0$ such that Lemma 7.2 holds. Then, for fixed $\boldsymbol{u} \in \mathcal{B}_{\delta/C}$, there is a unique $a = a(\Phi, \boldsymbol{u}) \in X_\delta$ associated to every $\Phi \in \mathcal{X}_\delta$ such that the modulation equation (7-8) is satisfied. Hence we can define $\boldsymbol{K}_{\boldsymbol{u}}(\Phi) := \boldsymbol{K}(\Phi, a, \boldsymbol{u})$. We intend to show that for small enough $\delta > 0$ the operator $\boldsymbol{K}_{\boldsymbol{u}}$ is a contraction on $\mathcal{X}_\delta$. To show the necessary bounds, we first split $\boldsymbol{K}_{\boldsymbol{u}}(\Phi)$ according to projections $\boldsymbol{P}_{a_\infty}$, $\boldsymbol{Q}_{a_\infty}$, $\boldsymbol{H}_{a_\infty}$, and $\boldsymbol{T}_{a_\infty}$, and then estimate each part separately.

First, note that the transversality of the projections implies

$$\boldsymbol{P}_{a_\infty}\boldsymbol{K}_{\boldsymbol{u}}(\Phi)(\tau) = -\int_\tau^\infty e^{\tau-\sigma}\boldsymbol{P}_{a_\infty}(\boldsymbol{G}_{a(\sigma)}(\Phi(\sigma)) - \partial_\sigma \boldsymbol{U}_{a(\sigma)})\,d\sigma$$

and

$$\boldsymbol{H}_{a_\infty}\boldsymbol{K}_{\boldsymbol{u}}(\Phi)(\tau) = -\int_\tau^\infty e^{3(\tau-\sigma)}\boldsymbol{H}_{a_\infty}(\boldsymbol{G}_{a(\sigma)}(\Phi(\sigma)) - \partial_\sigma \boldsymbol{U}_{a(\sigma)})\,d\sigma.$$

Now, since

$$\partial_\tau \boldsymbol{U}_{a(\tau)} = \dot{a}_j(\tau)\boldsymbol{q}^{(j)}_{a_\infty} + \dot{a}_j(\tau)[\boldsymbol{q}^{(j)}_{a(\tau)} - \boldsymbol{q}^{(j)}_{a_\infty}]$$

and

$$\|\boldsymbol{q}^{(j)}_{a(\tau)} - \boldsymbol{q}^{(j)}_{a_\infty}\| \lesssim \delta e^{-\omega\tau},$$

we have

$$\|\boldsymbol{H}_{a_\infty}\partial_\tau \boldsymbol{U}_{a(\tau)}\| + \|\boldsymbol{P}_{a_\infty}\partial_\tau \boldsymbol{U}_{a(\tau)}\| + \|(1-\boldsymbol{Q}_{a_\infty})\partial_\tau \boldsymbol{U}_{a(\tau)}\| \lesssim \delta^2 e^{-2\omega\tau} \tag{7-12}$$

for all $a \in X_\delta$. This, together with Lemma 7.1 and the fact that

$$\boldsymbol{Q}_{a_\infty}\boldsymbol{K}_{\boldsymbol{u}}(\Phi)(\tau) = \chi(\tau)\boldsymbol{Q}_{a_\infty}\boldsymbol{u} \tag{7-13}$$

(see (7-7)), yields the bounds

$$\begin{aligned} \|\boldsymbol{H}_{a_\infty}\boldsymbol{K}_{\boldsymbol{u}}(\Phi)(\tau)\| + \|\boldsymbol{P}_{a_\infty}\boldsymbol{K}_{\boldsymbol{u}}(\Phi)(\tau)\| &\lesssim \delta^2 e^{-2\omega\tau}, \\ \|\boldsymbol{Q}_{a_\infty}\boldsymbol{K}_{\boldsymbol{u}}(\Phi)(\tau)\| &\lesssim \frac{\delta}{C} e^{-2\omega\tau} \end{aligned} \tag{7-14}$$

for all $\Phi \in \mathcal{X}_\delta$. On the other hand, for the stable subspace we have

$$\boldsymbol{T}_{a_\infty}\boldsymbol{K}_{\boldsymbol{u}}(\Phi)(\tau) = \boldsymbol{S}_{a_\infty}(\tau)\boldsymbol{T}_{a_\infty}\boldsymbol{u} + \int_0^\tau \boldsymbol{S}_{a_\infty}(\tau-\sigma)\boldsymbol{T}_{a_\infty}(\boldsymbol{G}_{a(\sigma)}(\Phi(\sigma)) - \partial_\sigma \boldsymbol{U}_{a(\sigma)})\,d\sigma,$$

and by Lemma 7.1, Proposition 6.2, and (7-12), we get

$$\|\boldsymbol{T}_{a_\infty}\boldsymbol{K}_{\boldsymbol{u}}(\Phi)(\tau)\| \lesssim \Big(\frac{\delta}{C} + \delta^2\Big)e^{-\omega\tau} \tag{7-15}$$

for all $\Phi \in \mathcal{X}_\delta$. Now, from (7-14) and (7-15) we see that $\boldsymbol{K}_{\boldsymbol{u}}$ maps $\mathcal{X}_\delta$ into itself for all $\delta > 0$ sufficiently small and all $C > 0$ sufficiently large. The contraction property of $\boldsymbol{K}_{\boldsymbol{u}}$ is established similarly. Namely, there is the analogue of (7-12):

$$\begin{aligned} &\|\boldsymbol{H}_{a_\infty}\partial_\tau \boldsymbol{U}_{a(\tau)} - \boldsymbol{H}_{b_\infty}\partial_\tau \boldsymbol{U}_{b(\tau)}\| + \|\boldsymbol{P}_{a_\infty}\partial_\tau \boldsymbol{U}_{a(\tau)} - \boldsymbol{P}_{b_\infty}\partial_\tau \boldsymbol{U}_{b(\tau)}\| \\ &\qquad + \|(1-\boldsymbol{Q}_{a_\infty})\partial_\tau \boldsymbol{U}_{a(\tau)} - (1-\boldsymbol{Q}_{b_\infty})\partial_\tau \boldsymbol{U}_{b(\tau)}\| \lesssim \delta^2 e^{-2\omega\tau} \end{aligned}$$

for all $a, b \in X_\delta$. Furthermore, by Lemma 7.1, (7-13), and Lemma 7.2, we get the analogous estimates to (7-14); namely, we have

$$\begin{aligned} &\|\boldsymbol{H}_{a_\infty}\boldsymbol{K}_{\boldsymbol{u}}(\Phi)(\tau) - \boldsymbol{H}_{b_\infty}\boldsymbol{K}_{\boldsymbol{u}}(\Psi)(\tau)\| + \|\boldsymbol{P}_{a_\infty}\boldsymbol{K}_{\boldsymbol{u}}(\Phi)(\tau) - \boldsymbol{P}_{b_\infty}\boldsymbol{K}_{\boldsymbol{u}}(\Psi)(\tau)\| \\ &\qquad + \|\boldsymbol{Q}_{a_\infty}\boldsymbol{K}_{\boldsymbol{u}}(\Phi)(\tau) - \boldsymbol{Q}_{b_\infty}\boldsymbol{K}_{\boldsymbol{u}}(\Psi)(\tau)\| \lesssim \delta e^{-2\omega\tau}\|\Phi - \Psi\|_{\mathcal{X}} \end{aligned}$$

for all $\Phi, \Psi \in \mathcal{X}_\delta$, where $a = a(\Phi, \boldsymbol{u})$ and $b = a(\Psi, \boldsymbol{u})$. Also, in line with (7-15), we have

$$\|\boldsymbol{T}_{a_\infty}\boldsymbol{K}_{\boldsymbol{u}}(\Phi)(\tau) - \boldsymbol{T}_{b_\infty}\boldsymbol{K}_{\boldsymbol{u}}(\Psi)(\tau)\| \lesssim \delta e^{-\omega\tau}\|\Phi - \Psi\|_{\mathcal{X}}$$

for all $\Phi, \Psi \in \mathcal{X}_\delta$. By combining these estimates we get

$$\|\boldsymbol{K}_{\boldsymbol{u}}(\Phi) - \boldsymbol{K}_{\boldsymbol{u}}(\Psi)\|_{\mathcal{X}} \lesssim \delta\|\Phi - \Psi\|_{\mathcal{X}} \tag{7-16}$$

for all $\Phi, \Psi \in \mathcal{X}_\delta$, and contractivity follows by taking small enough $\delta > 0$.

For the Lipschitz continuity, similarly to proving (7-9), we use the integral equation (7-10) to show that, given sufficiently small $\delta > 0$,

$$\|\Phi(\boldsymbol{u}) - \Phi(\boldsymbol{v})\|_{\mathcal{X}} \lesssim \|\boldsymbol{u} - \boldsymbol{v}\|$$

for all $\boldsymbol{u} \in \mathcal{B}_{\delta/C}$, and then (7-9) implies (7-11). □

7C. *Conditional stability in similarity variables.* According to Proposition 7.3 and (2-8) there exists a family of initial data close to $\boldsymbol{U}_0$ which lead to global (strong) solutions to (2-3), which furthermore converge to $\boldsymbol{U}_{a_\infty}$ for some a_∞ close to $a = 0$; with minimal modifications, the same argument can be carried out for $\boldsymbol{U}_a$ for any $a \neq 0$. In conclusion, we have conditional asymptotic orbital stability of the family $\{\boldsymbol{U}_a : a \in \mathbb{R}^9\}$, the condition being that the initial data belong to the set which ensures global existence and convergence. In this section we show that this set represents a Lipschitz manifold of codimension 11.

Let $\delta > 0$ and $C > 0$ be as in Proposition 7.3, and let $\boldsymbol{u} \in \mathcal{B}_{\delta/C}$. Also, let us write

$$\boldsymbol{C}(\boldsymbol{u}) := \boldsymbol{C}(\Phi(\boldsymbol{u}), a(\boldsymbol{u}), \boldsymbol{u}),$$

where the mapping $\boldsymbol{u} \mapsto (\Phi(\boldsymbol{u}), a(\boldsymbol{u}))$ is defined in Proposition 7.3. Moreover, we denote the projection corresponding to all unstable directions by

$$\boldsymbol{J}_a := \boldsymbol{P}_a + \boldsymbol{H}_a.$$

Note that by definition $\boldsymbol{J}_{a_\infty}\boldsymbol{C}(\boldsymbol{u}) = \boldsymbol{C}(\boldsymbol{u})$, and we have the Lipschitz estimate

$$\|\boldsymbol{J}_a - \boldsymbol{J}_b\| \lesssim |a - b|$$

for all $a, b \in X_\delta$.

Proposition 7.4. *There exists $C > 0$ such that, for all sufficiently small $\delta > 0$, there exists a codimension-11 Lipschitz manifold $\mathcal{M} = \mathcal{M}_{\delta,C} \subset \mathcal{H}$ with $\boldsymbol{0} \in \mathcal{M}$, defined as the graph of a Lipschitz continuous function $\boldsymbol{M} : \ker \boldsymbol{J}_0 \cap \mathcal{B}_{\delta/2C} \to \operatorname{ran} \boldsymbol{J}_0$,*

$$\mathcal{M} := \left\{\boldsymbol{v} + \boldsymbol{M}(\boldsymbol{v}) : \boldsymbol{v} \in \ker \boldsymbol{J}_0, \|\boldsymbol{v}\| \leq \frac{\delta}{2C}\right\} \subset \{\boldsymbol{u} \in \mathcal{B}_{\delta/C} : \boldsymbol{C}(\boldsymbol{u}) = 0\}.$$

Furthermore, for every $\boldsymbol{u} \in \mathcal{M}$, there exists $(\Phi, a) = (\Phi_{\boldsymbol{u}}, a_{\boldsymbol{u}}) \in \mathcal{X}_\delta \times X_\delta$ satisfying

$$\Phi(\tau) = \boldsymbol{S}_{a_\infty}(\tau)\boldsymbol{u} + \int_0^\tau \boldsymbol{S}_{a_\infty}(\tau - \sigma)(\boldsymbol{G}_{a(\sigma)}(\Phi(\sigma)) - \partial_\sigma \boldsymbol{U}_{a(\sigma)})\, d\sigma \tag{7-17}$$

for all $\tau \geq 0$. Moreover, there exists $K > C$ such that $\{\boldsymbol{u} \in \mathcal{B}_{\delta/K} : \boldsymbol{C}(\boldsymbol{u}) = 0\} \subset \mathcal{M}_{\delta,C}$.

Proof. First, we show that, for small enough $\delta > 0$, we have $\boldsymbol{C}(\boldsymbol{u}) = 0$ if and only if $\boldsymbol{J}_0\boldsymbol{C}(\boldsymbol{u}) = 0$. Assume that $\boldsymbol{J}_0\boldsymbol{C}(\boldsymbol{u}) = 0$. Then we obtain the estimate

$$\|\boldsymbol{C}(\boldsymbol{u})\| \le \|\boldsymbol{J}_0\boldsymbol{C}(\boldsymbol{u}) + (\boldsymbol{J}_{a_{\boldsymbol{u},\infty}} - \boldsymbol{J}_0)\boldsymbol{C}(\boldsymbol{u})\| \lesssim |a_{\boldsymbol{u},\infty}|\|\boldsymbol{C}(\boldsymbol{u})\|.$$

Since $a_{\boldsymbol{u},\infty} = O(\delta)$, we get $\boldsymbol{C}(\boldsymbol{u}) = 0$. The other direction is obvious. Now we construct the mapping $\boldsymbol{M}$. Let $\boldsymbol{u} \in \mathcal{H}$ and take the decomposition $\boldsymbol{u} = \boldsymbol{v} + \boldsymbol{w} \in \ker \boldsymbol{J}_0 \oplus \operatorname{ran} \boldsymbol{J}_0$. Fix $\boldsymbol{v} \in \ker \boldsymbol{J}_0$ and define

$$\widetilde{\boldsymbol{C}}_{\boldsymbol{v}} : \operatorname{ran} \boldsymbol{J}_0 \to \operatorname{ran} \boldsymbol{J}_0, \quad \widetilde{\boldsymbol{C}}_{\boldsymbol{v}}(\boldsymbol{w}) = \boldsymbol{J}_0\boldsymbol{C}(\boldsymbol{v} + \boldsymbol{w}).$$

We establish that this mapping is invertible at zero, provided $\boldsymbol{v}$ is small enough, and we obtain $\boldsymbol{w} = \widetilde{\boldsymbol{C}}_{\boldsymbol{v}}^{-1}(\boldsymbol{0})$. This defines a mapping

$$\boldsymbol{M} : \ker \boldsymbol{J}_0 \to \operatorname{ran} \boldsymbol{J}_0, \quad \boldsymbol{M}(\boldsymbol{v}) := \widetilde{\boldsymbol{C}}_{\boldsymbol{v}}^{-1}(\boldsymbol{0}).$$

To show this, we use a fixed-point argument. Recall the definition of the correction terms $\boldsymbol{C} = \boldsymbol{C}_1 + \boldsymbol{C}_2$, $\boldsymbol{C}_1 = \sum_{k=0}^{9} \boldsymbol{C}_1^k$ with

$$\boldsymbol{C}_1^k(\Phi, a, \boldsymbol{u}) = \boldsymbol{P}_{a_\infty}^{(k)}\boldsymbol{u} + \boldsymbol{P}_{a_\infty}^{(k)}\boldsymbol{I}_1(\Phi, a)$$

and

$$\boldsymbol{C}_2(\Phi, a, \boldsymbol{u}) = \boldsymbol{H}_{a_\infty}\boldsymbol{u} + \boldsymbol{H}_{a_\infty}\boldsymbol{I}_2(\Phi, a),$$

where

$$\boldsymbol{I}_1(\Phi, a) := \int_0^\infty e^{-\sigma}[\boldsymbol{G}_{a(\sigma)}(\Psi(\sigma)) - \partial_\sigma \boldsymbol{U}_{a(\sigma)}]\, d\sigma$$

and

$$\boldsymbol{I}_2(\Phi, a) := \int_0^\infty e^{-3\sigma}[\boldsymbol{G}_{a(\sigma)}(\Psi(\sigma) - \partial_\sigma \boldsymbol{U}_{s(\sigma)})]\, d\sigma.$$

We write

$$\boldsymbol{F}_1(\boldsymbol{u}) := \sum_{k=0}^{9} \boldsymbol{F}_1^k(\boldsymbol{u}) = \sum_{k=0}^{9} \boldsymbol{P}_{a_\infty}^{(k)}\boldsymbol{I}_1(\Phi_{\boldsymbol{u}}, a_{\boldsymbol{u}})$$

and

$$\boldsymbol{F}_2(\boldsymbol{u}) := \boldsymbol{H}_{a_\infty}\boldsymbol{I}_2(\Phi_{\boldsymbol{u}}, a_{\boldsymbol{u}}).$$

By Lemma 7.1 and (7-12) we infer

$$\|\boldsymbol{F}_1^k(\boldsymbol{u})\| \lesssim \delta^2 \quad \text{and} \quad \|\boldsymbol{F}_2(\boldsymbol{u})\| \lesssim \delta^2. \tag{7-18}$$

Now, for $\boldsymbol{v} \in \ker \boldsymbol{J}_0$, we get

$$\begin{aligned}\widetilde{\boldsymbol{C}}_{\boldsymbol{v}}(\boldsymbol{w}) &= \boldsymbol{J}_0\boldsymbol{C}(\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{J}_0\boldsymbol{J}_{a_\infty}(\boldsymbol{v} + \boldsymbol{w}) + \boldsymbol{J}_0(\boldsymbol{F}_1(\boldsymbol{v} + \boldsymbol{w}) + \boldsymbol{F}_2(\boldsymbol{v} + \boldsymbol{w}))\\ &= \boldsymbol{J}_0^2\boldsymbol{w} + \boldsymbol{J}_0(\boldsymbol{J}_{a_\infty} - \boldsymbol{J}_0)\boldsymbol{w} + \boldsymbol{J}_0\boldsymbol{J}_{a_\infty}\boldsymbol{v} + \boldsymbol{J}_0(\boldsymbol{F}_1(\boldsymbol{v} + \boldsymbol{w}) + \boldsymbol{F}_2(\boldsymbol{v} + \boldsymbol{w}))\\ &= \boldsymbol{w} + \boldsymbol{J}_0(\boldsymbol{J}_{a_\infty} - \boldsymbol{J}_0)(\boldsymbol{v} + \boldsymbol{w}) + \boldsymbol{J}_0(\boldsymbol{F}_1(\boldsymbol{v} + \boldsymbol{w}) + \boldsymbol{F}_2(\boldsymbol{v} + \boldsymbol{w})).\end{aligned}$$

Introducing the notation

$$\Omega_{\boldsymbol{v}}(\boldsymbol{w}) := \boldsymbol{J}_0(\boldsymbol{J}_0 - \boldsymbol{J}_{a_\infty})(\boldsymbol{v} + \boldsymbol{w}) - \boldsymbol{J}_0(\boldsymbol{F}_1(\boldsymbol{v} + \boldsymbol{w}) + \boldsymbol{F}_2(\boldsymbol{v} + \boldsymbol{w})),$$

we rewrite equation $\widetilde{\boldsymbol{C}}_{\boldsymbol{v}}(\boldsymbol{w}) = 0$ as

$$\boldsymbol{w} = \Omega_{\boldsymbol{v}}(\boldsymbol{w}). \tag{7-19}$$

Now, for $\delta > 0$ and $C > 0$ from Proposition 7.3, we set

$$\widetilde{B}_{\delta/C}(\boldsymbol{v}) := \left\{\boldsymbol{w} \in \operatorname{ran} \boldsymbol{J}_0 : \|\boldsymbol{v}+\boldsymbol{w}\| \le \frac{\delta}{C}\right\}.$$

We show that $\Omega_{\boldsymbol{v}} : \widetilde{B}_{\delta/C}(\boldsymbol{v}) \to \widetilde{B}_{\delta/C}(\boldsymbol{v})$ is a contraction mapping for sufficiently small $\boldsymbol{v}$. Let $\boldsymbol{v} \in \mathcal{H}$ with $\|\boldsymbol{v}\| \le \delta/(2C)$, and let $\boldsymbol{w} \in \widetilde{B}_{\delta/C}(\boldsymbol{v})$. Using (7-18), we estimate

$$\|\Omega_{\boldsymbol{v}}(\boldsymbol{w})\| \le \|\boldsymbol{J}_0 - \boldsymbol{J}_{a_\infty}\|\|\boldsymbol{v}+\boldsymbol{w}\| + \|\boldsymbol{F}_1(\boldsymbol{v}+\boldsymbol{w})\| + \|\boldsymbol{F}_1(\boldsymbol{v}+\boldsymbol{w})\| \lesssim \frac{\delta^2}{C} + \delta^2.$$

Hence, by fixing $C > 0$, we have $\|\boldsymbol{v} + \Omega_{\boldsymbol{v}}(\boldsymbol{w})\| \le \delta/C$ for all small enough $\delta > 0$. So the ball $\widetilde{B}_{\delta/C}(\boldsymbol{v})$ is invariant under the action of $\Omega_{\boldsymbol{v}}$. To prove contractivity, first, for $\boldsymbol{w}, \tilde{\boldsymbol{w}} \in \widetilde{B}_{\delta/C}(\boldsymbol{v})$, we associate to $\boldsymbol{v}+\boldsymbol{w}$ and $\boldsymbol{v}+\tilde{\boldsymbol{w}}$ the functions (Φ, a) and (Ψ, b) in $\mathcal{X}_\delta \times X_\delta$ by Proposition 7.3, respectively. Then we obtain

$$\begin{aligned}\|\Omega_{\boldsymbol{v}}(\boldsymbol{w}) - \Omega_{\boldsymbol{v}}(\tilde{\boldsymbol{w}})\| \le{}& \|\boldsymbol{J}_0(\boldsymbol{J}_0 - \boldsymbol{J}_{a_\infty})(\boldsymbol{v}+\boldsymbol{w}) - \boldsymbol{J}_0(\boldsymbol{J}_0 - \boldsymbol{J}_{b_\infty})(\boldsymbol{v}+\tilde{\boldsymbol{w}})\| \\ &+ \|\boldsymbol{F}_1(\boldsymbol{v}+\boldsymbol{w}) - \boldsymbol{F}_1(\boldsymbol{v}+\tilde{\boldsymbol{w}})\| + \|\boldsymbol{F}_2(\boldsymbol{v}+\boldsymbol{w}) - \boldsymbol{F}_2(\boldsymbol{v}+\tilde{\boldsymbol{w}})\|,\end{aligned}$$

and writing

$$\boldsymbol{J}_0(\boldsymbol{J}_0 - \boldsymbol{J}_{a_\infty})(\boldsymbol{v}+\boldsymbol{w}) - \boldsymbol{J}_0(\boldsymbol{J}_0 - \boldsymbol{J}_{b_\infty})(\boldsymbol{v}+\tilde{\boldsymbol{w}}) = \boldsymbol{J}_0(\boldsymbol{J}_0 - \boldsymbol{J}_{a_\infty})(\boldsymbol{w}-\tilde{\boldsymbol{w}}) - \boldsymbol{J}_0(\boldsymbol{J}_{a_\infty} - \boldsymbol{J}_{b_\infty})(\boldsymbol{v}+\tilde{\boldsymbol{w}})$$

we get by Proposition 7.3 the estimate

$$\begin{aligned}\|\boldsymbol{J}_0(\boldsymbol{J}_0 - \boldsymbol{J}_{a_\infty})(\boldsymbol{w}-\tilde{\boldsymbol{w}})\| + \|\boldsymbol{J}_0(\boldsymbol{J}_{a_\infty} - \boldsymbol{J}_{b_\infty})(\boldsymbol{v}+\tilde{\boldsymbol{w}})\| &\lesssim |a_\infty|\|\boldsymbol{w}-\tilde{\boldsymbol{w}}\| + |a_\infty - b_\infty|\|\boldsymbol{v}+\boldsymbol{w}\| \\ &\lesssim \delta\|\boldsymbol{w}-\tilde{\boldsymbol{w}}\| + \frac{\delta}{C}\|a-b\|_X \\ &\lesssim \delta\|\boldsymbol{w}-\tilde{\boldsymbol{w}}\|.\end{aligned}$$

On the other hand, by Lemma 7.1 and (7-12), we obtain, for $k = 0, \dots, 9$, that

$$\|\boldsymbol{P}^{(k)}_{a_\infty}\boldsymbol{I}_1(\Phi, a) - \boldsymbol{P}^{(k)}_{b_\infty}\boldsymbol{I}_2(\Psi, b)\| \lesssim \delta(\|\Phi - \Psi\|_{\mathcal{X}} + \|a-b\|_X)$$

and

$$\|\boldsymbol{H}_{a_\infty}\boldsymbol{I}_2(\Phi, a) - \boldsymbol{H}_{b_\infty}\boldsymbol{I}_2(\Psi, b)\| \lesssim \delta(\|\Phi - \Psi\|_{\mathcal{X}} + \|a-b\|_X).$$

Thus we get the Lipschitz estimate

$$\|\boldsymbol{F}_1(\boldsymbol{v}+\boldsymbol{w}) - \boldsymbol{F}_1(\boldsymbol{v}+\tilde{\boldsymbol{w}})\| + \|\boldsymbol{F}_2(\boldsymbol{v}+\boldsymbol{w}) - \boldsymbol{F}_2(\boldsymbol{v}+\tilde{\boldsymbol{w}})\| \lesssim \delta\|\boldsymbol{w}-\tilde{\boldsymbol{w}}\|,$$

and we conclude that, for all small enough $\delta > 0$, the operator $\Omega_{\boldsymbol{v}} : \widetilde{B}_{\delta/C}(\boldsymbol{v}) \to \widetilde{B}_{\delta/C}(\boldsymbol{v})$ is contractive, with the contraction constant $\frac{1}{2}$. Consequently, by the contraction map principle we get that, for every $\boldsymbol{v} \in \ker \boldsymbol{J}_0 \cap \mathcal{B}_{\delta/2C}$, there exists a unique $\boldsymbol{w} \in \widetilde{B}_{\delta/C}(\boldsymbol{v})$ that solves (7-19); hence $\boldsymbol{C}(\boldsymbol{v}+\boldsymbol{w}) = \widetilde{\boldsymbol{C}}_{\boldsymbol{v}}(\boldsymbol{w}) = 0$.

Next, we establish the Lipschitz-continuity of the mapping $\boldsymbol{v} \mapsto \boldsymbol{M}(\boldsymbol{v})$. Let $\boldsymbol{v}, \tilde{\boldsymbol{v}} \in \ker \boldsymbol{J}_0 \cap \mathcal{B}_{\delta/2C}$ and $\boldsymbol{w}, \tilde{\boldsymbol{w}} \in \widetilde{B}_{\delta/C}$ be the corresponding solutions to (7-19). We get

$$\begin{aligned}\|\boldsymbol{M}(\boldsymbol{v}) - \boldsymbol{M}(\tilde{\boldsymbol{v}})\| = \|\boldsymbol{w}-\tilde{\boldsymbol{w}}\| &\le \|\Omega_{\boldsymbol{v}}(\boldsymbol{w}) - \Omega_{\boldsymbol{v}}(\tilde{\boldsymbol{w}})\| + \|\Omega_{\boldsymbol{v}}(\tilde{\boldsymbol{w}}) - \Omega_{\tilde{\boldsymbol{v}}}(\tilde{\boldsymbol{w}})\| \\ &\le \tfrac{1}{2}\|\boldsymbol{w}-\tilde{\boldsymbol{w}}\| + \|\Omega_{\boldsymbol{v}}(\tilde{\boldsymbol{w}}) - \Omega_{\tilde{\boldsymbol{v}}}(\tilde{\boldsymbol{w}})\|.\end{aligned}$$

The second term we estimate with

$$\begin{aligned}\|\Omega_{\boldsymbol{v}}(\tilde{\boldsymbol{w}})-\Omega_{\tilde{\boldsymbol{v}}}(\tilde{\boldsymbol{w}})\| &= \big\| \boldsymbol{J}_0(\boldsymbol{J}_0-\boldsymbol{J}_{a_{\boldsymbol{v}+\tilde{\boldsymbol{w}},\infty}})(\boldsymbol{v}+\tilde{\boldsymbol{w}})-\boldsymbol{J}_0(\boldsymbol{F}_1(\boldsymbol{v}+\tilde{\boldsymbol{w}})+\boldsymbol{F}_2(\boldsymbol{v}+\tilde{\boldsymbol{w}})) \\ &\qquad -\boldsymbol{J}_0(\boldsymbol{J}_0-\boldsymbol{J}_{a_{\tilde{\boldsymbol{v}}+\tilde{\boldsymbol{w}},\infty}})+\boldsymbol{J}_0(\boldsymbol{F}_1(\tilde{\boldsymbol{v}}+\tilde{\boldsymbol{w}})+\boldsymbol{F}_2(\tilde{\boldsymbol{v}}+\tilde{\boldsymbol{w}}))\big\| \\ &\lesssim \|\boldsymbol{J}_0(\boldsymbol{J}_{a_{\tilde{\boldsymbol{v}}+\tilde{\boldsymbol{w}},\infty}}-\boldsymbol{J}_{a_{\boldsymbol{v}+\tilde{\boldsymbol{w}},\infty}})\tilde{\boldsymbol{w}}\|+\|\boldsymbol{J}_0(\boldsymbol{J}_{a_{\boldsymbol{v}+\tilde{\boldsymbol{w}},\infty}}\boldsymbol{v}-\boldsymbol{J}_{a_{\tilde{\boldsymbol{v}}+\tilde{\boldsymbol{w}},\infty}}\boldsymbol{w})\| \\ &\qquad +\|\boldsymbol{J}_0(\boldsymbol{F}_1(\tilde{\boldsymbol{v}}+\tilde{\boldsymbol{w}})+\boldsymbol{F}_2(\tilde{\boldsymbol{v}})+\tilde{\boldsymbol{w}})\| \\ &\lesssim |a_{\tilde{\boldsymbol{v}}+\tilde{\boldsymbol{w}},\infty}-a_{\boldsymbol{v}+\tilde{\boldsymbol{w}}}|\|\tilde{\boldsymbol{w}}\|+\|\tilde{\boldsymbol{v}}-\boldsymbol{v}\|+|a_{\tilde{\boldsymbol{v}}+\tilde{\boldsymbol{w}},\infty}|\|\tilde{\boldsymbol{v}}\|+\delta\|\tilde{\boldsymbol{v}}-\boldsymbol{v}\| \\ &\lesssim \frac{\delta}{C}\|\tilde{\boldsymbol{v}}-\boldsymbol{v}\|+\frac{\delta}{2C}\|\tilde{\boldsymbol{v}}-\boldsymbol{v}\|+\delta\|\tilde{\boldsymbol{v}}-\boldsymbol{v}\| \\ &\lesssim \|\tilde{\boldsymbol{v}}-\boldsymbol{v}\|.\end{aligned}$$

Thereby we obtain the claimed Lipschitz estimate

$$\|\boldsymbol{M}(\boldsymbol{v})-\boldsymbol{M}(\tilde{\boldsymbol{v}})\| \le 2\|\Omega_{\boldsymbol{v}}(\tilde{\boldsymbol{w}})-\Omega(\tilde{\boldsymbol{w}})\| \lesssim \|\boldsymbol{v}-\tilde{\boldsymbol{v}}\|.$$

We note that, for $\boldsymbol{u}=0$, the associated (Φ,a) is trivial, i.e., $\Phi=0$ and $a=0$. Thus, we have $\boldsymbol{C}(\boldsymbol{0})=\boldsymbol{F}_1(\boldsymbol{0})+\boldsymbol{F}_2(\boldsymbol{0})=0$. Moreover, $\boldsymbol{u}=\boldsymbol{v}+\boldsymbol{w}=\boldsymbol{0}$ if and only if $\boldsymbol{v}=\boldsymbol{w}=\boldsymbol{0}$. Since in this case $\boldsymbol{v}$ satisfies the smallness condition, $\boldsymbol{w}$ solving $\boldsymbol{C}(\boldsymbol{0}+\boldsymbol{w})=\boldsymbol{0}$ is unique; hence $\boldsymbol{M}(\boldsymbol{0})=\boldsymbol{0}$.

Finally, let $\boldsymbol{u}\in\mathcal{H}$ satisfy $\boldsymbol{C}(\boldsymbol{u})=\boldsymbol{0}$. Then, since $1-\boldsymbol{J}_0$ is a bounded operator on $\mathcal{H}$,

$$\|(1-\boldsymbol{J}_0)\boldsymbol{u}\| \lesssim \|\boldsymbol{u}\|.$$

We obtain $\boldsymbol{v}_{\boldsymbol{u}} := (1-\boldsymbol{J}_0)\boldsymbol{u} \in \ker \boldsymbol{J}_0$ and $\|\boldsymbol{v}_{\boldsymbol{u}}\| \le \delta/(2C)$ for $\|\boldsymbol{u}\| \le \delta/K$ for $K>C$ large enough. Uniqueness yields $\boldsymbol{w}_{\boldsymbol{u}} := \boldsymbol{J}_0\boldsymbol{u} = \boldsymbol{M}(\boldsymbol{v}_{\boldsymbol{u}})$, and hence $\boldsymbol{u}\in\mathcal{M}_{\delta,C}$. □

Remark 7.5. For each correction term, the same argument yields the existence of Lipschitz manifolds $\mathcal{M}_1,\mathcal{M}_2\subset\mathcal{H}$ of codimensions 10 and 1, respectively, characterized by the vanishing of $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$. In particular, $\mathcal{M}$ can be characterized as a subset of the intersection $\mathcal{M}_1\cap\mathcal{M}_2$ in a small neighborhood around zero.

7D. *Proofs of Propositions 2.1 and 2.2.*

Proof of Proposition 2.1. Let $\Phi_0\in\mathcal{M}_{\delta,C}$, where $\mathcal{M}_{\delta,C}$ is the manifold defined in Proposition 7.4. In particular, $\|\Phi_0\|\le\delta/C$ and $\boldsymbol{C}(\Phi_0)=0$. By Proposition 7.4 there is a pair $(\Phi,a)\in\mathcal{X}_\delta\times X_\delta$ which solves (7-17) with initial data $\boldsymbol{u}=\Phi_0$. Furthermore, after substituting the variation of constants formula

$$\boldsymbol{S}_{a_\infty}(\tau)=\boldsymbol{S}(\tau)+\int_0^\tau \boldsymbol{S}(\tau-\sigma)\boldsymbol{L}'_{a_\infty}\boldsymbol{S}_{a_\infty}(\sigma)\,d\sigma$$

into (7-17), a straightforward calculation yields that $\Psi(\tau):=\boldsymbol{U}_{a(\tau)}+\Phi(\tau)$ satisfies

$$\Psi(\tau)=\boldsymbol{S}(\tau)(\boldsymbol{U}_0+\Phi_0)+\int_0^\tau \boldsymbol{S}(\tau-\sigma)\boldsymbol{F}(\Psi(\sigma))\,d\sigma \tag{7-20}$$

for all $\tau\ge0$. Then, based on (4-3) and (7-2) we infer that

$$\|\Psi(\tau)-\boldsymbol{U}_{a_\infty}\| \le \|\Phi(\tau)\|+\|\boldsymbol{U}_{a(\tau)}-\boldsymbol{U}_{a_\infty}\| \lesssim \delta e^{-\omega\tau}$$

for all $\tau\ge0$, as claimed. □

Proof of Proposition 2.2. Let $\Phi_0 \in \mathcal{M} \cap (C^\infty(\overline{\mathbb{B}}^9) \times C^\infty(\overline{\mathbb{B}}^9))$, and let $\Psi \in C([0,\infty), \mathcal{H})$ be the solution of (7-20) associated to Φ_0 via Proposition 2.1. To prove smoothness of $\Psi(\tau)$ (for fixed τ) we use the representation (7-20). Recall that we defined $\boldsymbol{S}(\tau) := \boldsymbol{S}_k(\tau)$ for $k = \frac{1}{2}(d+1) = 5$ with $(\boldsymbol{S}_k(\tau))_{\tau \geq 0}$ denoting the free wave evolution of Proposition 3.1. Now, using Lemma 3.6, we infer from (7-20) that

$$\begin{aligned}\|\Psi(\tau)\|_{H^6(\mathbb{B}^9)\times H^5(\mathbb{B}^9)} &\lesssim e^{-\frac{\tau}{2}} \|\boldsymbol{U}_0 + \Phi_0\|_{H^6(\mathbb{B}^9)\times H^5(\mathbb{B}^9)} + \int_0^\tau e^{-\frac{\tau-\sigma}{2}} \|\boldsymbol{F}(\Psi(\sigma))\|_{H^6(\mathbb{B}^9)\times H^5(\mathbb{B}^9)}\, d\sigma \\ &\lesssim e^{-\frac{\tau}{2}} \|\boldsymbol{U}_0 + \Phi_0\|_{H^6(\mathbb{B}^9)\times H^5(\mathbb{B}^9)} + \int_0^\tau e^{-\frac{\tau-\sigma}{2}} \|\Psi(\sigma)\|^2_{H^5(\mathbb{B}^9)\times H^4(\mathbb{B}^9)}\, d\sigma \lesssim 1\end{aligned}$$

for all $\tau \geq 0$. Then inductively, for $k \geq 5$, we get

$$\|\Psi(\tau)\|_{H^k(\mathbb{B}^9)\times H^{k-1}(\mathbb{B}^9)} \lesssim 1$$

for all $\tau \geq 0$. Consequently, by the Sobolev embedding we have $\Psi(\tau) \in C^\infty(\overline{\mathbb{B}}^9) \times C^\infty(\overline{\mathbb{B}}^9)$ for all $\tau \geq 0$.

To get regularity in τ we do the following. First, by local Lipschitz continuity of $\boldsymbol{F} : \mathcal{H}_k \mapsto \mathcal{H}_k$ for every $k \geq 5$ and Gronwall's lemma we get from (7-20) that $\Psi : [0, \mathcal{T}] \mapsto \mathcal{H}_k$ is Lipschitz continuous for every $\mathcal{T} > 0$ and $k \geq 5$. Consequently, $\boldsymbol{F}(\Psi(\cdot)), \boldsymbol{L}\boldsymbol{F}(\Psi(\cdot)) : [0, \mathcal{T}] \mapsto \mathcal{H}_k$ are Lipschitz continuous. The latter is immediate from interpreting $\boldsymbol{L}$ as a map from $\mathcal{H}_k$ to $\mathcal{H}_{k+2}$ and using the Lipschitz continuity of Ψ. Therefore, $\Psi \in C^1([0,\infty), \mathcal{H}_k)$, with

$$\begin{aligned}\partial_\tau \Psi(\tau) &= \boldsymbol{L}\Psi(\tau) + \boldsymbol{F}(\Psi(\tau)) \\ &= \boldsymbol{S}(\tau)\boldsymbol{L}(\boldsymbol{U}_0 + \Phi_0) + \int_0^\tau \boldsymbol{S}(\tau-\sigma)\boldsymbol{L}\boldsymbol{F}(\Psi(\sigma))\, d\sigma + \boldsymbol{F}(\Psi(\tau))\end{aligned} \tag{7-21}$$

for every $\tau \geq 0$; see, e.g., [Pazy 1983, p. 108, Corollary 2.6]. Consequently, by regularity of $\boldsymbol{F}$, $\boldsymbol{F}(\Psi(\cdot)), \boldsymbol{L}^m\boldsymbol{F}(\Psi(\cdot)) \in C^1([0,\infty), \mathcal{H}_k)$ for all $m \geq 0$ and $k \geq 5$. Therefore, from the second equality of (7-21), we get that $\partial_\tau \Psi \in C^1([0,\infty), \mathcal{H}_k)$, with

$$\partial_\tau^2 \Psi(\tau) = \boldsymbol{S}(\tau)\boldsymbol{L}^2(\boldsymbol{U}_0 + \Phi_0) + \int_0^\tau \boldsymbol{S}(\tau-\sigma)\boldsymbol{L}^2\boldsymbol{F}(\Psi(\sigma))\, d\sigma + \boldsymbol{L}\boldsymbol{F}(\Psi(\tau)) + \partial_\tau \boldsymbol{F}(\Psi(\tau)) \tag{7-22}$$

for all $\tau \geq 0$. Inductively, we get that $\Psi \in C^m([0,\infty), \mathcal{H}_k)$ for all $m \geq 0$ and $k \geq 5$. In particular, by the Sobolev embedding, $\partial_\tau^m \Psi(\tau) \in C^\infty(\overline{\mathbb{B}}^9) \times C^\infty(\overline{\mathbb{B}}^9)$. Additionally, by the Sobolev embedding $H^k(\mathbb{B}^9) \hookrightarrow L^\infty(\mathbb{B}^9)$ for $k \geq 5$, we get that the derivatives in τ hold pointwise. As a consequence, by (a strong version of) the Schwarz theorem (see, e.g., [Rudin 1976, p. 235, Theorem 9.41]), we get that mixed derivatives of all orders in τ and ξ exist, so $\Psi \in C^\infty(\mathcal{Z}) \times C^\infty(\mathcal{Z})$, and the first equality of (7-21) holds classically. □

7E. ***Variation of blow-up parameters and proof of Proposition 2.4.*** In this section we prove boundedness and continuity properties of the initial data operator Υ (see (2-6)) which are necessary to establish Proposition 2.4. We assume that $x_0 \in \overline{\mathbb{B}}^9_{1/2}$ and $T \in \left[\frac{1}{2}, \frac{3}{2}\right] =: I$. We also introduce the notation

$$\mathcal{Y} := H^6(\mathbb{B}^9_2) \times H^5(\mathbb{B}^9_2),$$

and denote by $\mathcal{B}_{\mathcal{Y}}$ the unit ball in $\mathcal{Y}$.

Lemma 7.6. *The initial data operator* $\Upsilon : \mathcal{B}_{\mathcal{Y}} \times I \times \overline{\mathbb{B}}^9_{1/2} \to \mathcal{H}$ *is Lipschitz continuous, i.e.,*

$$\|\Upsilon(\boldsymbol{v}, T_1, x_0) - \Upsilon(\boldsymbol{w}, T_2, y_0)\| \lesssim \|\boldsymbol{v} - \boldsymbol{w}\|_{\mathcal{Y}} + |T_1 - T_2| + |x_0 - y_0|$$

for all $\boldsymbol{v}, \boldsymbol{w} \in \mathcal{B}_{\mathcal{Y}}$, *all* $T_1, T_2 \in I$, *and all* $x_0, y_0 \in \overline{\mathbb{B}}^9_{1/2}$. *Furthermore, for* $\delta > 0$ *sufficiently small, we have*

$$\|\Upsilon(\boldsymbol{v}, T, x_0)\| \lesssim \delta$$

for all $\boldsymbol{v} \in \mathcal{Y}$ *with* $\|\boldsymbol{v}\|_{\mathcal{Y}} \le \delta$, *all* $T \in [1-\delta, 1+\delta] \subset I$, *and all* $x_0 \in \overline{\mathbb{B}}^9_{\delta}$.

Proof. Let $v \in C^\infty(\overline{\mathbb{B}}^9_2)$. Let $T \in \left[\frac{1}{2}, \frac{3}{3}\right]$ and $x_0, y_0 \in \overline{\mathbb{B}}^9_{1/2}$. Then we get by the fundamental theorem of calculus that

$$v(T\xi + x_0) - v(T\xi + y_0) = (x_0^i - y_0^i) \int_0^1 \partial_i v(T\xi + y_0 + s(x_0 - y_0))\, ds.$$

This implies that $\|v(T\cdot + x_0) - v(T\cdot + y_0)\|_{L^2(\mathbb{B}^9)} \lesssim \|v\|_{H^1(\mathbb{B}^9_2)} |x_0 - y_0|$. The same argument yields, for all $k \in \mathbb{N}$, that

$$\|v(T\cdot + x_0) - v(T\cdot + y_0)\|_{H^k(\mathbb{B}^9)} \lesssim \|v\|_{H^{k+1}(\mathbb{B}^9_2)} |x_0 - y_0|. \tag{7-23}$$

Similarly, we get, for all $T_1, T_2 \in \left[\frac{1}{2}, \frac{3}{2}\right]$ and all $x_0 \in \overline{\mathbb{B}}^9_{1/2}$, that

$$\|v(T_1\cdot + x_0) - v(T_2\cdot + x_0)\|_{H^k(\mathbb{B}^9)} \lesssim \|v\|_{H^{k+1}(\mathbb{B}^9_2)} |T_1 - T_2|, \tag{7-24}$$

where $k \in \mathbb{N}$. The estimates (7-23) and (7-24) can be extended to $v \in H^{k+1}(\mathbb{B}^9_2)$ by density. Now let $\boldsymbol{v}, \boldsymbol{w} \in \mathcal{Y}$, $T_1, T_2 \in \left[\frac{1}{2}, \frac{3}{2}\right]$, and $x_0, y_0 \in \overline{\mathbb{B}}^9_{1/2}$. Inequalities (7-23) and (7-24) imply

$$\|\mathcal{R}(\boldsymbol{v}, T_1, x_0) - \mathcal{R}(\boldsymbol{w}, T_2, y_0)\| \lesssim \|v\|_{\mathcal{Y}}(|T_1 - T_2| + |x_0 - y_0|) + \|\boldsymbol{v} - \boldsymbol{w}\|_{\mathcal{Y}}. \tag{7-25}$$

Moreover, since $\boldsymbol{U}_0$ is smooth, we have

$$\|\mathcal{R}(\boldsymbol{U}_0, T_1, x_0) - \mathcal{R}(\boldsymbol{U}_0, T_2, y_0)\| \lesssim |T_1 - T_2| + |x_0 - y_0| \tag{7-26}$$

for all $T_1, T_2 \in \left[\frac{1}{2}, \frac{3}{2}\right]$ and $x_0, y_0 \in \overline{\mathbb{B}}^9_{1/2}$. Now the inequalities (7-25) and (7-26) imply the first part of the statement. The same inequalities imply

$$\|\Upsilon(\boldsymbol{v}, T, x_0)\| \lesssim \|\boldsymbol{v}\|_{\mathcal{Y}} + |T - 1| + |x_0|,$$

which proves the second part of the statement. □

We have the following result, which has Proposition 2.4 as a direct consequence. To shorten the notation, we write $\boldsymbol{h} := \boldsymbol{h}_0$.

Lemma 7.7. *There exists* $M > 0$ *such that, for all sufficiently small* $\delta > 0$, *the following holds: For every real-valued* $\boldsymbol{v} \in \mathcal{Y}$ *that satisfies* $\|\boldsymbol{v}\|_{\mathcal{Y}} \le \delta/M^2$, *there exist* $\Phi \in \mathcal{X}_\delta$, $a \in X_\delta$, *and parameters* $\alpha \in [-\delta/M, \delta/M]$, $T \in [1-\delta/M, 1+\delta/M] \subset \left[\frac{1}{2}, \frac{3}{2}\right]$, *and* $x_0 \in \overline{\mathbb{B}}^9_{\delta/M} \subset \overline{\mathbb{B}}^9_{1/2}$ *such that*

$$\boldsymbol{C}(\Phi, a, \Upsilon(\boldsymbol{v} + \alpha \boldsymbol{h}_0, T, x_0)) = 0. \tag{7-27}$$

Moreover, the parameters depend Lipschitz continuously on the data, i.e.,

$$|\alpha(\boldsymbol{v})-\alpha(\boldsymbol{w})|+|T(\boldsymbol{v})-T(\boldsymbol{w})|+|x_0(\boldsymbol{v})-x_0(\boldsymbol{w})|\lesssim\|\boldsymbol{v}-\boldsymbol{w}\|_{\mathcal{Y}}$$

for all $\boldsymbol{v},\boldsymbol{w}\in\mathcal{Y}$ *satisfying the above assumptions. In particular,* $\Upsilon(\boldsymbol{v}+\alpha\boldsymbol{h},T,x_0)\in\mathcal{M}_{\delta,C}$.

Proof. Fix constants $C>0$ and $K>0$ from Proposition 7.4. By Lemma 7.6, we have that, for all $M>0$ large enough and all $\delta>0$ small enough, the inequality

$$\|\Upsilon(\boldsymbol{v}+\alpha\boldsymbol{h},T,x_0)\|\le\frac{\delta}{K} \tag{7-28}$$

holds for every $\|\boldsymbol{v}\|_{\mathcal{Y}}\le\delta/M$, $\alpha\in[-\delta/M,\delta/M]$, $T\in[1-\delta/M,1+\delta/M]$, and $x_0\in\overline{\mathbb{B}}^9_{\delta/M}$. Furthermore, in view of (7-28) and Proposition 7.3, we get that, given $\|\boldsymbol{v}\|_{\mathcal{Y}}\le\delta/M^2$, for every $\alpha\in[-\delta/M,\delta/M]$, $T\in[1-\delta/M,1+\delta/M]$, and $x_0\in\overline{\mathbb{B}}^9_{\delta/M}$, there are functions

$$\Phi=\Phi(\boldsymbol{v}+\alpha\boldsymbol{h},T,x_0)\quad\text{and}\quad a=a(\boldsymbol{v}+\alpha\boldsymbol{h},T,x_0)$$

which solve the modified integral equation

$$\begin{aligned}\Phi(\tau)=\boldsymbol{S}_{a_\infty}(\tau)(\Upsilon(\boldsymbol{v},T,x_0)-\boldsymbol{C}(\Phi,a,\Upsilon(\boldsymbol{v},T,x_0)))\\ +\int_0^\tau\boldsymbol{S}_{a_\infty}(\tau-\sigma)(\boldsymbol{G}_{a(\sigma)}(\Phi(\sigma))-\partial_\sigma\boldsymbol{U}_{a(\sigma)})\,d\sigma\end{aligned} \tag{7-29}$$

for all $\tau\ge0$. For such Φ and a, we show that one can associate to any $\|\boldsymbol{v}\|_{\mathcal{Y}}\le\delta/M^2$ suitable parameters T, x_0, and α such that (7-27) holds. From this, via Proposition 7.4, we conclude that $\Upsilon(\boldsymbol{v}+\alpha\boldsymbol{h},T,x_0)\in\mathcal{M}_{\delta,C}$. Recall that the correction terms can be written as $\boldsymbol{C}=\boldsymbol{C}_1+\boldsymbol{C}_2=\sum_{k=0}^9\boldsymbol{C}_1^k+\boldsymbol{C}_2$, where

$$\boldsymbol{C}_1^k(\Phi,a,\boldsymbol{u})=\boldsymbol{P}^{(k)}_{a_\infty}\boldsymbol{u}+\boldsymbol{P}^{(k)}_{a_\infty}\boldsymbol{I}_1(\Phi,a)\quad\text{and}\quad\boldsymbol{C}_2(\Phi,a,\boldsymbol{u})=\boldsymbol{H}_{a_\infty}\boldsymbol{u}+\boldsymbol{H}_{a_\infty}\boldsymbol{I}_2(\Phi,a),$$

and where the integrals are denoted by

$$\begin{aligned}\boldsymbol{I}_1(\Phi,a)&=\int_0^\infty e^{-\sigma}(\boldsymbol{G}_{a(\sigma)}(\Phi(\sigma))-\partial_\sigma\boldsymbol{U}_{a(\sigma)})\,d\sigma,\\ \boldsymbol{I}_2(\Phi,a)&=\int_0^\infty e^{-3\sigma}(\boldsymbol{G}_{a(\sigma)}(\Phi(\sigma))-\partial_\sigma\boldsymbol{U}_{a(\sigma)})\,d\sigma,\end{aligned}$$

and we have

$$\|\boldsymbol{P}^{(k)}_{a_\infty}\boldsymbol{I}_1(\Phi,a)\|\lesssim\delta^2\quad\text{and}\quad\|\boldsymbol{H}_{a_\infty}\boldsymbol{I}_2(\Phi,a)\|\lesssim\delta^2; \tag{7-30}$$

see (7-18). We will show that there are parameters T, α, and x_0 such that, for $k=0,\dots,9$,

$$(\boldsymbol{C}_1^k(\Phi,a,\Upsilon(\boldsymbol{v}+\alpha\boldsymbol{h},T,x_0))\,|\,\boldsymbol{g}^{(k)}_{a_\infty})=0\quad\text{and}\quad(\boldsymbol{C}_2(\Phi,a,\Upsilon(\boldsymbol{v}+\alpha\boldsymbol{h},T,x_0))\,|\,\boldsymbol{h}_{a_\infty})=0, \tag{7-31}$$

which implies (7-27). To this end we expand the initial data operator. First, by Taylor expansion we get, for $T\in[1-\delta/M,1+\delta/M]$ and $x_0\in\overline{\mathbb{B}}^9_{\delta/M}$,

$$\mathcal{R}(\boldsymbol{U}_0,T,x_0)-\mathcal{R}(\boldsymbol{U}_0,1,0)=c_0(T-1)\boldsymbol{g}_0^{(0)}+\sum_{j=1}^9c_jx_0^j\boldsymbol{g}_0^{(j)}+r(T,x_0),$$

where the remainder satisfies

$$\|r(T,x_0)-r(\tilde{T},\tilde{x}_0)\|\lesssim\delta(|T-\tilde{T}|+|x_0-\tilde{x}_0|).$$

Hence we obtain

$$\Upsilon(\boldsymbol{v}+\alpha\boldsymbol{h},T,x_0)=\mathcal{R}(\boldsymbol{v}+\alpha\boldsymbol{h},T,x_0)+c_0(T-1)\boldsymbol{g}^{(0)}_{a_\infty}+\sum_{j=1}^{9}c_jx_0^j\boldsymbol{g}^{(j)}_{a_\infty}+r_{a_\infty}(T,x_0),$$

where

$$r_{a_\infty}(T,x_0)=c_0(T-1)(\boldsymbol{g}^{(0)}_0-\boldsymbol{g}^{(0)}_{a_\infty})+\sum_{j=1}^{9}c_jx_0^j(\boldsymbol{g}^{(j)}_0-\boldsymbol{g}^{(j)}_{a_\infty})+r(T,x_0).$$

It is straightforward to check that

$$\|r_a(T,x_0)-r_b(\widetilde{T},\tilde{x}_0)\|\lesssim\delta(|a-b|+|T-\widetilde{T}|+|x_0-\tilde{x}_0|) \tag{7-32}$$

for all $a,b\in\mathbb{B}^9_\delta$, $T,\widetilde{T}\in[1-\delta/M,1+\delta/M]$, and $x_0,\tilde{x}_0\in\overline{\mathbb{B}}^9_{\delta/M}$. We now write

$$\mathcal{R}(\boldsymbol{v}+\alpha\boldsymbol{h},T,x_0)=\mathcal{R}(\boldsymbol{v},T,x_0)+\alpha\mathcal{R}(\boldsymbol{h}_{a_\infty},T,x_0)+\alpha\mathcal{R}(\boldsymbol{h}-\boldsymbol{h}_{a_\infty},T,x_0).$$

The last term can be estimated by

$$\|\mathcal{R}(\boldsymbol{h}-\boldsymbol{h}_{a_\infty},T,x_0)\|\lesssim|a_\infty|.$$

By taking the Taylor expansion of $\mathcal{R}(\boldsymbol{h}_{a_\infty},T,x_0)$ at $(T,x_0)=(1,0)$, we obtain

$$\mathcal{R}(\boldsymbol{v}+\alpha\boldsymbol{h},T,x_0)=\mathcal{R}(\boldsymbol{v},T,x_0)+\alpha\boldsymbol{h}_{a_\infty}+\alpha\tilde{r}_a(T,x_0),$$

where the remainder satisfies

$$\|\tilde{r}_a(T,x_0)-\tilde{r}_b(\widetilde{T},\tilde{x}_0)\|\lesssim|a-b|+|T-\widetilde{T}|+|x_0-\tilde{x}_0|. \tag{7-33}$$

Hence we obtain the expansion

$$\Upsilon(\boldsymbol{v}+\alpha\boldsymbol{h},T,x_0)=\mathcal{R}(\boldsymbol{v},T,x_0)+\alpha\boldsymbol{h}_{a_\infty}+c_0(T-1)\boldsymbol{g}^{(0)}_{a_\infty}+\sum_{j=1}^{9}c_jx_0^j\boldsymbol{g}^{(j)}_{a_\infty}+r_{a_\infty}(T,x_0)+\alpha\tilde{r}_{a_\infty}(T,x_0).$$

By applying the projections to the initial data operator we get

$$\begin{aligned}
\boldsymbol{P}^{(0)}_{a_\infty}\Upsilon(\boldsymbol{v}+\alpha\boldsymbol{h},T,x_0)&=\boldsymbol{P}^{(0)}_{a_\infty}\mathcal{R}(\boldsymbol{v},T,x_0)+c_0(T-1)\boldsymbol{g}^{(0)}_{a_\infty}+\boldsymbol{P}^{(0)}_{a_\infty}r_{a_\infty}(T,x_0)+\alpha\boldsymbol{P}^{(0)}_{a_\infty}\tilde{r}_{a_\infty}(T,x_0),\\
\boldsymbol{P}^{(j)}_{a_\infty}\Upsilon(\boldsymbol{v}+\alpha\boldsymbol{h},T,x_0)&=\boldsymbol{P}^{(j)}_{a_\infty}\mathcal{R}(\boldsymbol{v},T,x_0)+c_jx_0^j\boldsymbol{g}^{(j)}_{a_\infty}+\boldsymbol{P}^{(j)}_{a_\infty}r_{a_\infty}(T,x_0)+\alpha\boldsymbol{P}^{(j)}_{a_\infty}\tilde{r}_{a_\infty}(T,x_0),\\
\boldsymbol{H}_{a_\infty}\Upsilon(\boldsymbol{v}+\alpha\boldsymbol{h},T,x_0)&=\boldsymbol{H}_{a_\infty}\mathcal{R}(\boldsymbol{v},T,x_0)+\alpha\boldsymbol{h}_{a_\infty}+\boldsymbol{H}_{a_\infty}r_{a_\infty}(T,x_0)+\alpha\boldsymbol{H}_{a_\infty}\tilde{r}_{a_\infty}(T,x_0).
\end{aligned}$$

Hence, by introducing the notation $\beta=T-1$, we define, for $k=0,\dots,9$,

$$\begin{aligned}
\boldsymbol{\Gamma}^{(k)}_{\boldsymbol{v}}(\alpha,\beta,x_0)&=\boldsymbol{P}^{(k)}_{a_\infty}\mathcal{R}(\boldsymbol{v},\beta+1,x_0)+\boldsymbol{P}^{(k)}_{a_\infty}r_{a_\infty}(\beta,x_0)+\alpha\boldsymbol{P}^{(k)}_{a_\infty}\tilde{r}_{a_\infty}(\beta,x_0)+\boldsymbol{P}^{(k)}_{a_\infty}\boldsymbol{I}_1(\alpha,\beta,x_0),\\
\boldsymbol{\Gamma}^{(10)}_{\boldsymbol{v}}(\alpha,\beta,x_0)&=\boldsymbol{H}_{a_\infty}\mathcal{R}(\boldsymbol{v},\beta+1,x_0)+\boldsymbol{H}_{a_\infty}r_{a_\infty}(\beta,x_0)+\alpha\boldsymbol{H}_{a_\infty}\tilde{r}_{a_\infty}(\beta,x_0)+\boldsymbol{H}_{a_\infty}\boldsymbol{I}_2(\alpha,\beta,x_0).
\end{aligned}$$

Using this notation we can rewrite (7-31) as

$$\begin{aligned}
\beta&=\Gamma^{(0)}_{\boldsymbol{v}}(\alpha,\beta,x_0):=\tilde{c}_0(\boldsymbol{\Gamma}^{(0)}_{\boldsymbol{v}}(\alpha,\beta,x_0)\,|\,\boldsymbol{g}^{(0)}_{a_\infty}),\\
x_0^j&=\Gamma^{(j)}_{\boldsymbol{v}}(\alpha,\beta,x_0):=\tilde{c}_j(\boldsymbol{\Gamma}^{(j)}_{\boldsymbol{v}}(\alpha,\beta,x_0)\,|\,\boldsymbol{g}^{(j)}_{a_\infty}),\\
\alpha&=\Gamma^{(10)}_{\boldsymbol{v}}(\alpha,\beta,x_0):=\tilde{c}_{10}(\boldsymbol{\Gamma}^{(10)}_{\boldsymbol{v}}(\alpha,\beta,x_0)\,|\,\boldsymbol{h}_{a_\infty})
\end{aligned} \tag{7-34}$$

for $j = 1, \dots, 9$ and some constants $\tilde{c}_0, \tilde{c}_j, \tilde{c}_{10} \in \mathbb{R}$. We will show that $\Gamma_{\boldsymbol{v}} = (\Gamma_{\boldsymbol{v}}^{(0)}, \dots, \Gamma_{\boldsymbol{v}}^{(10)})$ is a contraction on $\overline{\mathbb{B}}^{11}_{\delta/M}$ for sufficiently small $\delta > 0$ and for sufficiently large $M > 0$. Thereby the first part of the statement follows by Banach's fixed-point theorem.

First we observe that $\Gamma_{\boldsymbol{v}}$ maps $\overline{\mathbb{B}}^{11}_{\delta/M}$ into itself. Indeed, by the proof of Lemma 7.6, we know that $\|\mathcal{R}(\boldsymbol{v}, 1+\beta, x_0)\| \lesssim \|\boldsymbol{v}\|_{\mathcal{Y}}$. Now estimates (7-32)–(7-33), and the integral estimates (7-30) imply

$$\Gamma_{\boldsymbol{v}}^{(j)}(\alpha, \beta, x_0) = O\left(\frac{\delta}{M^2}\right) + O(\delta^2)$$

for all $j = 0, \dots, 10$. Thus, there is a choice of large enough $M > 0$ such that, for all sufficiently small $\delta > 0$, the inequality

$$|\Gamma_{\boldsymbol{v}}(\alpha, \beta, x_0)| \le \frac{\delta}{M}$$

holds for all $(\alpha, \beta, x_0) \in \overline{\mathbb{B}}^{11}_{\delta/M}$. Next we show that by restricting, if necessary, to even smaller $\delta > 0$, the operator $\Gamma_{\boldsymbol{v}}$ is a contraction on $\overline{\mathbb{B}}^{11}_{\delta/M}$. Let $(\Phi, a) \in \mathcal{X}_\delta \times X_\delta$ be the functions solving (7-29) corresponding to parameters $\boldsymbol{v} + \alpha \boldsymbol{h}$, $T = 1+\beta$, and x_0. Furthermore, let $(\Psi, b) \in \mathcal{X}_\delta \times X_\delta$ be the functions corresponding to $\boldsymbol{v} + \tilde{\alpha}\boldsymbol{h}$, $\widetilde{T} = 1 + \tilde{\beta}$, and $\tilde{x}_0$. Then, we obtain

$$\|\Phi - \Psi\|_{\mathcal{X}} + \|a - b\|_X \lesssim \|\Upsilon(\boldsymbol{v} + \alpha\boldsymbol{h}, T, x_0) - \Upsilon(\boldsymbol{v} + \tilde{\alpha}\boldsymbol{h}, \widetilde{T}, \tilde{x}_0)\| \lesssim |\alpha - \tilde{\alpha}| + |\beta - \tilde{\beta}| + |x_0 - \tilde{x}_0|.$$

Hence, by Lemma 7.1, we get, for $k = 0, \dots, 9$, that

$$\|\boldsymbol{P}_{a_\infty}^{(k)} \boldsymbol{I}_1(\Phi, a) - \boldsymbol{P}_{b_\infty}^{(k)} \boldsymbol{I}_1(\Psi, b)\| \lesssim \delta(\|\Phi - \Psi\|_{\mathcal{X}} + \|a - b\|_X) \lesssim \delta(|\alpha - \tilde{\alpha}| + |\beta - \tilde{\beta}| + |x_0 - \tilde{x}_0|)$$

and

$$\|\boldsymbol{H}_{a_\infty} \boldsymbol{I}_2(\Phi, a) - \boldsymbol{H}_{b_\infty} \boldsymbol{I}_2(\Psi, b)\| \lesssim \delta(|\alpha - \tilde{\alpha}| + |\beta - \tilde{\beta}| + |x_0 - \tilde{x}_0|).$$

Furthermore, by (7-25) and the Lipschitz continuity of the Riesz projections $\boldsymbol{P}_a^{(k)}$ and $\boldsymbol{H}_a$, we obtain

$$\begin{aligned}\|\boldsymbol{P}_{a_\infty}^{(k)} \mathcal{R}(\boldsymbol{v}, T, x_0) - \boldsymbol{P}_{b_\infty}^{(k)} \mathcal{R}(\boldsymbol{v}, \widetilde{T}, \tilde{x}_0)\| + \|\boldsymbol{H}_{a_\infty} \mathcal{R}(\boldsymbol{v}, T, x_0) - \boldsymbol{H}_{b_\infty} \mathcal{R}(\boldsymbol{v}, \widetilde{T}, \tilde{x}_0)\| \\ \lesssim \|\boldsymbol{v}\|_{\mathcal{Y}}(\|a - b\|_X + |T - \widetilde{T}| + |x_0 - \tilde{x}_0|) \lesssim \delta(|\alpha - \tilde{\alpha}| + |\beta - \tilde{\beta}| + |x_0 - \tilde{x}_0|).\end{aligned}$$

Moreover, for $k = 0, \dots, 9$, we have

$$\begin{aligned}\|\boldsymbol{P}_{a_\infty}^{(k)} r_{a_\infty}(T, x_0) - \boldsymbol{P}_{b_\infty}^{(k)} r_{b_\infty}(\widetilde{T}, \tilde{x}_0)\| + \|\alpha \boldsymbol{P}_{a_\infty}^{(k)} \tilde{r}_{a_\infty}(T, x_0) - \tilde{\alpha} \boldsymbol{P}_{b_\infty}^{(k)} \tilde{r}_{b_\infty}(\widetilde{T}, \tilde{x}_0)\| \\ \lesssim \delta(|\alpha - \tilde{\alpha}| + |\beta - \tilde{\beta}| + |x_0 - \tilde{x}_0|)\end{aligned}$$

and

$$\begin{aligned}\|\boldsymbol{H}_{a_\infty} r_{a_\infty}(T, x_0) - \boldsymbol{H}_{b_\infty} r_{b_\infty}(\widetilde{T}, \tilde{x}_0)\| + \|\alpha \boldsymbol{H}_{a_\infty} \tilde{r}_{a_\infty}(T, x_0) - \tilde{\alpha} \boldsymbol{H}_{b_\infty} \tilde{r}_{b_\infty}(\widetilde{T}, \tilde{x}_0)\| \\ \lesssim \delta(|\alpha - \tilde{\alpha}| + |\beta - \tilde{\beta}| + |x_0 - \tilde{x}_0|).\end{aligned}$$

From these estimates we infer that

$$\|\Gamma_{\boldsymbol{v}}^{(j)}(\alpha, \beta, x_0) - \Gamma_{\boldsymbol{v}}^{(j)}(\tilde{\alpha}, \tilde{\beta}, \tilde{x}_0)\| \lesssim \delta(|\alpha - \tilde{\alpha}| + |\beta - \tilde{\beta}| + |x_0 - \tilde{x}_0|) \tag{7-35}$$

for $j = 0, \dots, 10$. Therefore, $\Gamma_{\boldsymbol{v}}$ is a contraction for all small enough $\delta > 0$, and this concludes the proof of the first part of the statement.

It remains to establish the Lipschitz continuity of the parameters with respect to the initial data. Let $\boldsymbol{v}, \boldsymbol{w} \in \mathcal{Y}$ satisfy the smallness condition and let (α, β, x_0) and $(\tilde{\alpha}, \tilde{\beta}, \tilde{x}_0)$ be the corresponding set of parameters. The first line in (7-34) implies

$$\begin{aligned}|\beta - \tilde{\beta}| &= |\Gamma_{\boldsymbol{v}}^{(0)}(\alpha, \beta, x_0) - \Gamma_{\boldsymbol{w}}^{(0)}(\tilde{\alpha}, \tilde{\beta}, \tilde{x}_0)| \\ &\lesssim |\Gamma_{\boldsymbol{v}}^{(0)}(\alpha, \beta, x_0) - \Gamma_{\boldsymbol{w}}^{(0)}(\alpha, \beta, x_0)| + |\Gamma_{\boldsymbol{w}}^{(0)}(\alpha, \beta, x_0) - \Gamma_{\boldsymbol{w}}^{(0)}(\tilde{\alpha}, \tilde{\beta}, \tilde{x}_0)|.\end{aligned}$$

The second term can be estimated with (7-35). To estimate the first term, we use the Lipschitz continuity of the Riesz projections to get

$$\begin{aligned}&\| \boldsymbol{P}^{(0)}_{a_\infty(\boldsymbol{v},\beta,x_0)} \mathcal{R}(\boldsymbol{v}, 1+\beta, x_0) - \boldsymbol{P}^{(0)}_{a_\infty(\boldsymbol{w},\beta,x_0)} \mathcal{R}(\boldsymbol{w}, 1+\beta, x_0)\| \\ &\qquad \lesssim \|\boldsymbol{v}\|_{\mathcal{Y}} \|a_\infty(\boldsymbol{v}, \beta, x_0) - a_\infty(\boldsymbol{w}, \beta, x_0)\|_X + \|\boldsymbol{v} - \boldsymbol{w}\|_{\mathcal{Y}} \lesssim \|\boldsymbol{v} - \boldsymbol{w}\|_{\mathcal{Y}}.\end{aligned}$$

Similar estimates using (7-32)–(7-33) and Lemma 7.1 yield

$$|\Gamma_{\boldsymbol{v}}^{(0)}(\alpha, \beta, x_0) - \Gamma_{\boldsymbol{w}}^{(0)}(\alpha, \beta, x_0)| \lesssim \|\boldsymbol{v} - \boldsymbol{w}\|_{\mathcal{Y}}.$$

In summary, we obtain

$$|\beta - \tilde{\beta}| \lesssim \delta(|\alpha - \tilde{\alpha}| + |\beta - \tilde{\beta}| + |x_0 - \tilde{x}_0|) + \|\boldsymbol{v} - \boldsymbol{w}\|_{\mathcal{Y}},$$

and similar estimates for the remaining components yield

$$|\alpha - \tilde{\alpha}| + |\beta - \tilde{\beta}| + |x_0 - \tilde{x}_0| \lesssim \delta(|\alpha - \tilde{\alpha}| + |\beta - \tilde{\beta}| + |x_0 - \tilde{x}_0|) + \|\boldsymbol{v} - \boldsymbol{w}\|_{\mathcal{Y}},$$

which concludes the proof. □

7F. ***Proof of Theorem 1.1.*** Let $M > 0$ be from Proposition 2.4. For $\delta > 0$ define $\delta' := \delta/M$. Then consider $(f, g) \in C^\infty(\bar{\mathbb{B}}_2^9) \times C^\infty(\bar{\mathbb{B}}_2^9)$ satisfying

$$\|(f, g)\|_{H^6(\mathbb{B}_2^9) \times H^5(\mathbb{B}_2^9)} \le \frac{\delta'}{M} = \frac{\delta}{M^2}.$$

By Propositions 2.4 and 2.1, we have that, for all $\delta > 0$ sufficiently small, there exist $a \in \bar{\mathbb{B}}^9_{M\delta'/\omega}$, $T \in [1 - \delta', 1 + \delta']$, $x_0 \in \bar{\mathbb{B}}^9_{\delta'}$, and $\alpha \in [-\delta', \delta']$, depending Lipschitz continuously on (f, g) with respect to the norm on $\mathcal{Y}$, as well a function $\Psi \in C([0, \infty), \mathcal{H})$ that solves

$$\Psi(\tau) = \boldsymbol{S}(\tau)[\boldsymbol{U}_0 + \Upsilon((f, g) + \alpha \boldsymbol{h}, T, x_0)] + \int_0^\tau \boldsymbol{S}(\tau - \sigma) \boldsymbol{F}(\Psi(\sigma))\, d\sigma, \tag{7-36}$$

and obeys the estimate

$$\|\Psi(\tau) - \boldsymbol{U}_a\| \lesssim \delta e^{-\omega\tau} \tag{7-37}$$

for all $\tau \ge 0$. By standard arguments, Ψ is the unique solution to (7-36) in $C([0, \infty), \mathcal{H})$. Now, from the smoothness of f and g, we have that the initial data $\Psi(0) = \boldsymbol{U}_0 + \Upsilon((f, g) + \alpha \boldsymbol{h}, T, x_0)$ belongs to $C^\infty(\bar{\mathbb{B}}_2^9) \times C^\infty(\bar{\mathbb{B}}_2^9)$, and therefore from Proposition 2.2 we infer that Ψ is smooth and solves (2-3)

classically. More precisely, by writing $\Psi(\tau) = (\psi_1(\tau,\cdot), \psi_2(\tau,\cdot))$, we have that $\psi_j \in C^\infty(\mathcal{Z})$ for $j = 1, 2$ and

$$\partial_\tau \psi_1(\tau,\xi) = -\xi\cdot\nabla\psi_1(\tau,\xi) - 2\psi_1(\tau,\xi) + \psi_2(\tau,\xi),$$
$$\partial_\tau \psi_2(\tau,\xi) = \Delta\psi_1(\tau,\xi) - \xi\cdot\nabla\psi_2(\tau,\xi) - 3\psi_2(\tau,\xi), +\psi_1(\tau,\xi)^2$$

for $(\tau,\xi) \in \mathcal{Z}$, with

$$\psi_1(0,\cdot) = T^2[\boldsymbol{U}_0]_1(T\cdot + x_0) + T^2 f(T\cdot + x_0) + \alpha T^2 h_1(T\cdot + x_0),$$
$$\psi_2(0,\cdot) = T^3[\boldsymbol{U}_0]_2(T\cdot + x_0) + T^3 g(T\cdot + x_0) + \alpha T^3 h_2(T\cdot + x_0).$$

Furthermore, by writing $\Phi(\tau) = \Psi(\tau) - \boldsymbol{U}_a$, where $\Phi(\tau) = (\varphi_1(\tau,\cdot), \varphi_2(\tau,\cdot))$, from (7-37) we have

$$\|\varphi_1(\tau,\cdot)\|_{H^5(\mathbb{B}^9)} \lesssim \delta e^{-\omega\tau} \quad \text{and} \quad \|\varphi_2(\tau,\cdot)\|_{H^4(\mathbb{B}^9)} \lesssim \delta e^{-\omega\tau} \tag{7-38}$$

for all $\tau \geq 0$. Furthermore, by the Sobolev embedding we have, for the first component, that

$$\|\varphi_1(\tau,\cdot)\|_{L^\infty(\mathbb{B}^9)} \lesssim \delta e^{-\omega\tau} \tag{7-39}$$

for all $\tau \geq 0$. Now, we translate these results back to physical coordinates and let

$$u(t,x) = \frac{1}{(T-t)^2}\psi_1\left(-\log(T-t) + \log T, \frac{x-x_0}{T-t}\right).$$

Based on the smoothness properties of ψ_1, we conclude that $u \in C^\infty(\mathcal{C}_{T,x_0})$. Furthermore, u solves

$$(\partial_t^2 - \Delta_x)u(t,x) = u(t,x)^2$$

on $\mathcal{C}_{T,x_0}$ and satisfies

$$u(0,\cdot) = U(|\cdot|) + f + \alpha h_1, \quad \partial_t u(0,\cdot) = 2U(|\cdot|) + |\cdot|U'(|\cdot|) + g + \alpha h_2$$

on $\overline{\mathbb{B}}^9_T(x_0)$. Uniqueness of u follows from uniqueness of Ψ, though it also follows by standard results concerning wave equations in physical coordinates. Furthermore,

$$u(t,x) = \frac{1}{(T-t)^2}\left[U_a\left(\frac{x-x_0}{T-t}\right) + \varphi(t,x)\right],$$

with $\varphi(t,x) := \varphi_1(-\log(T-t) + \log T, (x-x_0)/(T-t))$. The bound (7-39) yields

$$\|\varphi(t,\cdot)\|_{L^\infty(\mathbb{B}^9_{T-t}(x_0))} = \|\varphi_1(-\log(T-t) + \log T,\cdot)\|_{L^\infty(\mathbb{B}^9)} \lesssim \delta(T-t)^\omega \tag{7-40}$$

for all $t \in [0,T)$. Furthermore, by (7-38),

$$(T-t)^{k-\frac{9}{2}}\|\varphi(t,\cdot)\|_{\dot{H}^k(\mathbb{B}^9_{T-t}(x_0))} = \|\varphi_1(-\log(T-t) + \log T,\cdot)\|_{\dot{H}^k(\mathbb{B}^9)} \lesssim (T-t)^\omega$$

for $k = 0, \dots, 5$, which implies the first line in (1-17). The second line follows also from (7-38) and the fact that

$$\partial_t u(t,x) = \frac{1}{(T-t)^3}\psi_2\left(-\log(T-t) + \log T, \frac{x-x_0}{T-t}\right).$$

Relabelling δ' with δ concludes the proof of Theorem 1.1 for (1-1). Now, let c_0 be the constant from (1-13). Recall that the above conclusions hold for all sufficiently small $\delta > 0$. Therefore, from (7-40) we see that we can choose small enough $\delta > 0$ so as to ensure

$$\|\varphi(t,\cdot)\|_{L^\infty(\mathbb{B}^9_{T-t}(x_0))} \le \tfrac{1}{2}c_0$$

for all $t \in [0,T)$. As a consequence, u is strictly positive on $\mathcal{C}_{T,x_0}$ and therefore provides a solution to (1-6) as well. □

8. Proof of Theorem 1.6: stable ODE blowup

The proof of Theorem 1.6 follows mutatis mutandis the proof of Theorem 1.1. However, for the convenience of the reader we outline the most important steps and stick to the notation introduced above. Starting with (2-3), we consider solutions of the form $\Psi(\tau) = \boldsymbol{\kappa}_{a(\tau)} + \Phi(\tau)$, which yields

$$\partial_\tau \Phi(\tau) = [\boldsymbol{L} + \boldsymbol{L}'_{\kappa_{a_\infty}}]\Phi(\tau) + \widetilde{\boldsymbol{G}}_{a(\tau)}(\Phi(\tau)) - \partial_\tau \boldsymbol{\kappa}_{a(\tau)}, \tag{8-1}$$

where

$$\boldsymbol{L}'_{\kappa_a}\boldsymbol{u}(\xi) = \begin{pmatrix} 0 \\ 2\kappa_a(\xi)u_1(\xi) \end{pmatrix}$$

and

$$\widetilde{\boldsymbol{G}}_{a(\tau)}(\Phi(\tau)) = [\boldsymbol{L}'_{\kappa_{a(\tau)}} - \boldsymbol{L}'_{\kappa_{a_\infty}}]\Phi(\tau) + \boldsymbol{F}(\Phi(\tau)).$$

In this equation, $\boldsymbol{L} : \mathcal{D}(\boldsymbol{L}) \subset \mathcal{H} \to \mathcal{H}$ denotes, as usual, the operator describing the free wave evolution. This is fully characterized for both $d = 7$ and $d = 9$ in Section 3; recall that $\mathcal{H} := \mathcal{H}_k$ for $k = \frac{1}{2}(d+1)$. For the perturbation theory, the spectral analysis is crucial. Once this is obtained, most results are purely abstract and the proofs can be adapted from previous sections.

Since $\boldsymbol{L}'_{\kappa_a}$ is compact and depends Lipschitz continuously on a, the results of Section 4 apply. In particular, for small enough a, the spectrum of $\boldsymbol{L} + \boldsymbol{L}'_{\kappa_a}$ in the right half-plane consists of isolated eigenvalues confined to a compact region. Furthermore, an analogous result to Proposition 4.5 holds with V replaced by a constant. This substantially simplifies the spectral analysis and with the above prerequisites it is easy to derive the following statement. For all of the ensuing statements, $d \in \{7, 9\}$.

Proposition 8.1. *There are constants $\delta^* > 0$ and $\omega > 0$ such that the following holds: For any $a \in \overline{\mathbb{B}}^d_{\delta^*}$, the operator $\boldsymbol{L} + \boldsymbol{L}'_{\kappa_a} : \mathcal{D}(\boldsymbol{L}) \subset \mathcal{H} \to \mathcal{H}$ generates a strongly continuous semigroup $(\boldsymbol{S}_{\kappa_a}(\tau))_{\tau\ge 0}$ on $\mathcal{H}$. Furthermore, there exist projections $\widetilde{\boldsymbol{P}}_a, \widetilde{\boldsymbol{Q}}^{(k)}_a \in \mathcal{B}(\mathcal{H})$, $k = 1, \dots, d$, of rank 1 that are mutually transversal and depend Lipschitz continuously on a. Furthermore, they commute with $\boldsymbol{S}_{\kappa_a}(\tau)$ and, for all $\boldsymbol{u} \in \mathcal{H}$ and $\tau \ge 0$,*

$$\boldsymbol{S}_{\kappa_a}(\tau)\widetilde{\boldsymbol{P}}_a\boldsymbol{u} = e^\tau \boldsymbol{u} \quad \text{and} \quad \boldsymbol{S}_{\kappa_a}(\tau)\widetilde{\boldsymbol{Q}}^{(k)}_a\boldsymbol{u} = \boldsymbol{u},$$

as well as

$$\|\boldsymbol{S}_{\kappa_a}(\tau)[1 - \widetilde{\boldsymbol{P}}_a - \widetilde{\boldsymbol{Q}}_a]\boldsymbol{u}\| \lesssim e^{-\omega\tau}\|[1 - \widetilde{\boldsymbol{P}}_a - \widetilde{\boldsymbol{Q}}_a]\boldsymbol{u}\|$$

with $\widetilde{\boldsymbol{Q}}_a = \sum_{k=1}^d \widetilde{\boldsymbol{Q}}^{(k)}_a$. Moreover,

$$\|\boldsymbol{S}_{\kappa_a}(\tau)[1 - \widetilde{\boldsymbol{P}}_a - \widetilde{\boldsymbol{Q}}_a] - \boldsymbol{S}_{\kappa_b}(\tau)[1 - \widetilde{\boldsymbol{P}}_b - \widetilde{\boldsymbol{Q}}_b]\| \lesssim e^{-\omega\tau}|a-b| \tag{8-2}$$

for all $a, b \in \overline{\mathbb{B}}^9_{\delta^*}$ *and* $\tau \geq 0$. *Also,*

$$\operatorname{ran} \widetilde{\boldsymbol{P}}_a = \operatorname{span}(\tilde{\boldsymbol{g}}_a) \quad \textit{and} \quad \operatorname{ran} \widetilde{\boldsymbol{Q}}_a^{(k)} = \operatorname{span}(\tilde{\boldsymbol{q}}_a^{(k)}), \tag{8-3}$$

where

$$\tilde{\boldsymbol{g}}_a(\xi) = \begin{pmatrix} A_0(a)[A_0(a) - A_j\xi^j]^{-3} \\ 3A_0(a)^2[A_0(a) - A_j\xi^j]^{-4} \end{pmatrix} \quad \textit{and} \quad \tilde{\boldsymbol{q}}_a^{(k)} = \partial_{a_k}\boldsymbol{\kappa}_a \tag{8-4}$$

for $k = 1, \dots, d$.

Proof. We only sketch the main steps of the proof, since many parts are abstract operator theory and can be copied verbatim from previous sections.

The results of Section 3 together with the bounded perturbation theorem immediately imply that $\boldsymbol{L} + \boldsymbol{L}'_{\kappa_a}$ generates a strongly continuous semigroup, which we denote by $(\boldsymbol{S}_{\kappa_a}(\tau))_{\tau\geq 0}$. Furthermore, the results of Propositions 4.3 and 4.5 hold in particular for our case at hand, and we infer that, for $\operatorname{Re}\lambda > -\frac{1}{2}$, the spectrum of $\boldsymbol{L} + \boldsymbol{L}'_{\kappa_a}$ consists of isolated eigenvalues confined to a compact region. For $a = 0$, Proposition 4.5 holds mutatis mutandis with V replaced by the constant potential $2\kappa_0 = 12$. In this case, in the spectral ODE the number of regular singular points can be reduced to three, and we can therefore resolve the connection problem by using the standard theory of hypergeometric equations. This is outlined in the following, where we show that there exists $0 < \mu_0 \leq \frac{1}{2}$ such that

$$\sigma(\boldsymbol{L} + \boldsymbol{L}'_{\kappa_0}) \subset \{\lambda \in \mathbb{C} : \operatorname{Re}\lambda \leq -\mu_0\} \cup \{0, 1\}.$$

In fact, we convince ourselves that

$$\{\lambda \in \mathbb{C} : \operatorname{Re}\lambda \geq 0\} \setminus \{0, 1\} \subset \rho(\boldsymbol{L} + \boldsymbol{L}'_{\kappa_0}). \tag{8-5}$$

We argue by contradiction. Let us assume that $\lambda \in \sigma(\boldsymbol{L} + \boldsymbol{L}'_{\kappa_0}) \setminus \{0, 1\}$ and $\operatorname{Re}\lambda \geq 0$. Then, for some $\ell \in \mathbb{N}_0$, (4-7) with potential $2\kappa_0$ must have an analytic solution on $[0, 1]$. We show that this cannot be the case. By changing variables and setting $f(\rho) = \rho^\ell v(\rho^2)$, (4-7) transforms into the standard hypergeometric form

$$z(1-z)v''(z) + (c - (a+b+1)z)v'(z) + abv(z) = 0,$$

with

$$a = \tfrac{1}{2}(\lambda + \ell - 1), \quad b = \tfrac{1}{2}(\lambda + \ell + 6), \quad \text{and} \quad c = \tfrac{1}{2}d + \ell.$$

Fundamental systems around the regular singular points $\rho = 0$ and $\rho = 1$ are given by $\{v_0, \tilde{v}_0\}$ and $\{v_1, \tilde{v}_1\}$, respectively, where

$$\begin{aligned} v_0(z) &= {}_2F_1(a, b; c; z), \\ \tilde{v}_0(z) &= z^{1-c}{}_2F_1(a+1-c, b+1-c; 2-c; z), \\ v_1(z) &= {}_2F_1(a, b; a+b+1-c; 1-z), \\ \tilde{v}_1(z) &= (1-z)^{c-a-b}{}_2F_1(c-a, c-b; 1+c-a-b; 1-z). \end{aligned}$$

In fact, this holds if $a + b - c \neq 0$, and, for $a + b - c = 0$, the function $\tilde{v}_1$ behaves logarithmically. In either case, the solutions $\tilde{v}_1$ and $\tilde{v}_0$ are not admissible since they are not analytic for $\operatorname{Re}\lambda \geq 0$. We

therefore look for λ which connect v_0 and v_1, i.e., for which v_0 and v_1 are constant multiples of each other. For the hypergeometric equation, the connection coefficients are known explicitly and we have

$$v_0(z) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} v_1(z) + \frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)} \tilde{v}_1(z);$$

see [DLMF 2010]. So the condition that quantifies our eigenvalues is

$$\frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)} = 0.$$

This can only be the case if a or b is a pole of the gamma function, i.e., $-a \in \mathbb{N}_0$ or $-b \in \mathbb{N}_0$. In particular, this implies that $\lambda \in \mathbb{R}$. Since $\operatorname{Re}\lambda \geq 0$, we can exclude $-b \in \mathbb{N}_0$. On the other hand, $-a \in \mathbb{N}_0$ is possible only if $\lambda \in \{0, 1\}$, which contradicts our assumption and proves (8-5). For $\lambda = 1$ and $\lambda = 0$, one can easily check that explicit solutions to the eigenvalue equation are given by $\tilde{\boldsymbol{g}}_0$ and $\tilde{\boldsymbol{q}}_0^{(k)}$, respectively, where

$$\tilde{\boldsymbol{g}}_0 = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \quad \text{and} \quad \tilde{\boldsymbol{q}}_0^{(k)} = \partial_{a_k} \boldsymbol{\kappa}_a|_{a=0} = 6 \begin{pmatrix} \xi_k \\ 3\xi_k \end{pmatrix}$$

for $k = 1, \dots, d$. Similar to the above reasoning one shows that these functions indeed span the eigenspaces for the corresponding eigenvalues, i.e., the geometric multiplicities of $\lambda = 1$ and $\lambda = 0$ are 1 and d, respectively. The algebraic multiplicities are determined by the dimension of the ranges of the corresponding Riesz projections

$$\tilde{\boldsymbol{P}}_0 = \frac{1}{2\pi i} \int_{\gamma_1} \boldsymbol{R}_{\boldsymbol{L}+\boldsymbol{L}'_{\kappa_0}}(\lambda)\, d\lambda \quad \text{and} \quad \tilde{\boldsymbol{Q}}_0 = \frac{1}{2\pi i} \int_{\gamma_0} \boldsymbol{R}_{\boldsymbol{L}+\boldsymbol{L}'_{\kappa_0}}(\lambda)\, d\lambda,$$

where, for $j \in \{0, 1\}$, we have $\gamma_j(s) = \lambda_j + \frac{1}{4}\omega_0 e^{2\pi i s}$ for $s \in [0, 1]$. An ODE argument analogous to the proof of Lemma 5.7 yields

$$\operatorname{ran} \tilde{\boldsymbol{P}}_0 = \operatorname{span}(\tilde{\boldsymbol{g}}_0) \quad \text{and} \quad \operatorname{ran} \tilde{\boldsymbol{Q}}_0 = \operatorname{span}(\tilde{\boldsymbol{q}}_0^{(1)}, \dots, \tilde{\boldsymbol{q}}_0^{(d)}).$$

The perturbative characterization of the spectrum of $\boldsymbol{L} + \boldsymbol{L}'_{\kappa_a}$ for $a \in \bar{\mathbb{B}}^d_{\delta*}$ is purely abstract. Along the lines of the proof of Lemma 5.8, one shows that

$$\sigma(\boldsymbol{L} + \boldsymbol{L}'_{\kappa_a}) \subset \left\{\lambda \in \mathbb{C} : \operatorname{Re}\lambda < -\tfrac{1}{2}\omega_0\right\} \cup \{0, 1\},$$

where for $\lambda = 0$ and $\lambda = 1$, the eigenfunctions are Lorentz boosted versions of $\tilde{\boldsymbol{g}}_0$ and $\tilde{\boldsymbol{q}}_0^{(k)}$. In fact, one can check by direct calculations that the functions $\tilde{\boldsymbol{g}}_a$ and $\tilde{\boldsymbol{q}}_a^{(k)}$ stated in (8-4) solve the corresponding eigenvalue equations. Equation (8-3) for the spectral projections

$$\tilde{\boldsymbol{P}}_a = \frac{1}{2\pi i} \int_{\gamma_1} \boldsymbol{R}_{\boldsymbol{L}+\boldsymbol{L}'_{\kappa_a}}(\lambda)\, d\lambda \quad \text{and} \quad \tilde{\boldsymbol{Q}}_a = \frac{1}{2\pi i} \int_{\gamma_0} \boldsymbol{R}_{\boldsymbol{L}+\boldsymbol{L}'_{\kappa_a}}(\lambda)\, d\lambda$$

follows from the abstract arguments provided in the proof of Lemma 6.1. The same holds for the Lipschitz dependence of the projections on the parameter a. The growth bounds for the semigroup follow from the structure of the spectrum, resolvent bounds and the Gearhart–Prüss theorem analogous to the proof of Proposition 6.2. Finally, the proof for the Lipschitz continuity (8-2) can be copied verbatim. □

The analysis of the integral equation

$$\Phi(\tau) = \boldsymbol{S}_{\kappa_{a_\infty}}(\tau)\boldsymbol{u} + \int_0^\tau \boldsymbol{S}_{\kappa_{a_\infty}}(\tau-\sigma)(\widetilde{\boldsymbol{G}}_{a(\sigma)}(\Phi(\sigma)) - \partial_\sigma \boldsymbol{\kappa}_{a(\sigma)})\, d\sigma$$

is completely analogous to Section 7. In particular, to derive the modulation equation for a, one uses the fact that $\partial_\tau \boldsymbol{\kappa}_{a(\tau)} = \dot{a}_k(\tau)\tilde{\boldsymbol{q}}_0^{(k)}$. By introducing the correction

$$\widetilde{\boldsymbol{C}}(\Phi, a, \boldsymbol{u}) = \widetilde{\boldsymbol{P}}_{a_\infty}\boldsymbol{u} + \widetilde{\boldsymbol{P}}_{a_\infty}\int_0^\infty e^{-\sigma}(\widetilde{\boldsymbol{G}}_{a(\sigma)}(\Phi(\sigma)) - \partial_\sigma \boldsymbol{\kappa}_{a(\sigma)})\, d\sigma,$$

it is straightforward to prove the following result.

Proposition 8.2. *There exists $\omega > 0$ such that, for all sufficiently large $C > 0$ and all sufficiently small $\delta > 0$, the following holds: For every $\|\boldsymbol{v}\|_{\mathcal{Y}} \le \delta/C$, every $T \in [1-\delta/C, 1+\delta/C]$, and every $x_0 \in \overline{\mathbb{B}}^d_{\delta/C}$, there exist functions $\Phi \in \mathcal{X}_\delta$ and $a \in X_\delta$ such that the integral equation*

$$\Phi(\tau) = \boldsymbol{S}_{\kappa_{a_\infty}}(\tau)(\Upsilon(\boldsymbol{v}, T, x_0) - \widetilde{\boldsymbol{C}}(\Phi, a, \Upsilon(\boldsymbol{v}, T, x_0))) + \int_0^\tau \boldsymbol{S}_{\kappa_{a_\infty}}(\tau-\sigma)(\widetilde{\boldsymbol{G}}_{a(\sigma)}(\Phi(\sigma)) - \partial_\sigma \beta\kappa_{a(\sigma)})\, d\sigma$$

holds for $\tau \ge 0$, and also $\|\Phi(\tau)\| \lesssim \delta e^{-\omega\tau}$ for all $\tau \ge 0$. Moreover, the solution map is Lipschitz continuous, i.e.,

$$\|\Phi(\boldsymbol{v}, T_1, x_0) - \Phi(\boldsymbol{w}, T_2, y_0)\|_{\mathcal{X}} + \|a(\boldsymbol{v}, T_1, x_0) - a(\boldsymbol{w}, T_2, y_0)\|_X \lesssim \|\boldsymbol{v} - \boldsymbol{w}\|_{\mathcal{Y}} + |T_1 - T_2| + |x_0 - y_0|$$

for all $\boldsymbol{v}, \boldsymbol{w} \in \mathcal{Y}$ satisfying the smallness assumption, all $T_1, T_2 \in [1-\delta/C, 1+\delta/C]$, and all $x_0, y_0 \in \overline{\mathbb{B}}^d_{\delta/C}$.

We note that similarly to the manifold $\mathcal{M}$ one can construct a manifold $\mathcal{N} \subset \ker \widetilde{\boldsymbol{P}}_0 \oplus \operatorname{ran} \widetilde{\boldsymbol{P}}_0$ of codimension $(1+d)$ characterized by the vanishing of the correction term $\widetilde{\boldsymbol{C}}$. However, in the context of stable blowup this is not of much interest, since the existence of this manifold is solely caused by the translation instability. In particular, no correction of the physical initial data is necessary, if blow-up time and point are chosen appropriately, i.e., for suitably small (f, g), there are T and x_0 such that $\Upsilon(\boldsymbol{v}, T, x_0) \in \mathcal{N}$. This is contained in the following result, where $\mathcal{Y} = H^{(d+3)/2}(\mathbb{B}^d) \times H^{(d+1)/2}(\mathbb{B}^d)$.

Lemma 8.3. *There exists $C > 0$ such that, for all sufficiently small $\delta > 0$, the following holds: For every $\boldsymbol{v} \in \mathcal{Y}$ satisfying $\|\boldsymbol{v}\|_{\mathcal{Y}} \le \delta/C^2$, there is a choice of parameters $T \in [1-\delta/C, 1+\delta/C]$ and $x_0 \in \overline{\mathbb{B}}^d_{\delta/C}$ in Proposition 8.2 such that*

$$\widetilde{\boldsymbol{C}}(\Phi, a, \Upsilon(\boldsymbol{v}, T, x_0)) = 0.$$

Moreover, the parameters depend Lipschitz continuously on the data, i.e.,

$$|T(\boldsymbol{v}) - T(\boldsymbol{w})| + |x_0(\boldsymbol{v}) - x_0(\boldsymbol{w})| \lesssim \|\boldsymbol{v} - \boldsymbol{w}\|_{\mathcal{Y}}$$

for all $\boldsymbol{v}, \boldsymbol{w} \in \mathcal{Y}$ satisfying the above smallness assumption.

The proof is along the lines of the proof of Lemma 7.7 on page 668 with obvious simplifications. With these results, in combination with persistence of regularity that is completely analogous to Proposition 2.2, Theorem 1.6 is obtained by the same arguments as in Section 7F.

Appendix: Proof of Lemma 3.4

We will frequently use the identities

$$2\,\mathrm{Re}[\xi^j \partial_j f(\xi)\bar{f}(\xi)] = \partial_j[\xi^j |f(\xi)|^2] - d|f(\xi)|^2 \tag{A-1}$$

and

$$\partial_{i_1}\partial_{i_2}\cdots\partial_{i_k}[\xi^j\partial_j f] = k\partial_{i_1}\partial_{i_2}\cdots\partial_{i_k} f(\xi) + \xi^j\partial_j\partial_{i_1}\partial_{i_2}\cdots\partial_{i_k} f(\xi), \tag{A-2}$$

which hold for all $k \in \mathbb{N}$ and $f \in C^\infty(\overline{\mathbb{B}}^d)$. Furthermore, by the divergence theorem, we have

$$\int_{\mathbb{B}^d} \partial_i \Delta u(\xi)\overline{\partial^i v(\xi)}\,d\xi = -\int_{\mathbb{B}^d} \Delta u(\xi)\overline{\Delta v(\xi)}\,d\xi + \int_{\mathbb{S}^{d-1}} \Delta u(\omega)\overline{\omega_i\partial^i v(\omega)}\,d\sigma(\omega) \tag{A-3}$$

for smooth u and v, and similarly

$$\int_{\mathbb{B}^d} \partial_i \Delta u(\xi)\overline{\partial^i v(\xi)}\,d\xi = -\int_{\mathbb{B}^d} \partial_i\partial_j u(\xi)\overline{\partial^j\partial^i v(\xi)}\,d\xi + \int_{\mathbb{S}^{d-1}} \omega_j\partial^j\partial_i u(\omega)\overline{\partial^i v(\omega)}\,d\sigma(\omega). \tag{A-4}$$

We first prove the result for $d = 9$, starting with those parts of $(\cdot\,|\,\cdot)_{\mathcal{H}_k}$ that correspond to the standard $\dot{H}^k \times \dot{H}^{k-1}$ inner product.

For the sake of concreteness, we consider the case $k = 5$, which corresponds to the space we are going to use later on. Using the above identities, we infer

$$\begin{aligned}
&\mathrm{Re}\int_{\mathbb{B}^9} \partial_i\partial_j\partial_k\partial_l\partial_m[\widetilde{\boldsymbol{L}}\boldsymbol{u}]_1(\xi)\overline{\partial^i\partial^j\partial^k\partial^l\partial^m u_1(\xi)}\,d\xi \\
&\quad = \mathrm{Re}\int_{\mathbb{B}^9} \partial_i\partial_j\partial_k\partial_l\partial_m u_2(\xi)\overline{\partial^i\partial^j\partial^k\partial^l\partial^m u_1(\xi)}\,d\xi - \tfrac{5}{2}\int_{\mathbb{B}^9} \partial_i\partial_j\partial_k\partial_l\partial_m u_1(\xi)\overline{\partial^i\partial^j\partial^k\partial^l\partial^m u_1(\xi)}\,d\xi \\
&\qquad\qquad - \tfrac{1}{2}\int_{\mathbb{S}^8} \partial_i\partial_j\partial_k\partial_l\partial_m u_1(\omega)\overline{\partial^i\partial^j\partial^k\partial^l\partial^m u_1(\omega)}\,d\sigma(\omega).
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\mathrm{Re}\int_{\mathbb{B}^9} \partial_i\partial_j\partial_k\partial_l[\widetilde{\boldsymbol{L}}\boldsymbol{u}]_2(\xi)\overline{\partial^i\partial^j\partial^k\partial^l u_2(\xi)}\,d\xi \\
&\quad = -\tfrac{5}{2}\int_{\mathbb{B}^9} \partial_i\partial_j\partial_k\partial_l u_2(\xi)\overline{\partial^i\partial^j\partial^k\partial^l u_2(\xi)}\,d\xi - \mathrm{Re}\int_{\mathbb{B}^9} \partial_i\partial_j\partial_k\partial_l\partial_m u_1(\xi)\overline{\partial^i\partial^j\partial^k\partial^l\partial^m u_2(\xi)}\,d\xi \\
&\qquad\qquad - \tfrac{1}{2}\int_{\mathbb{S}^8} \partial_i\partial_j\partial_k\partial_l u_2(\omega)\overline{\partial^i\partial^j\partial^k\partial^l u_2(\omega)}\,d\sigma(\omega) \\
&\qquad\qquad\qquad + \mathrm{Re}\int_{\mathbb{S}^8} \omega^m\partial_i\partial_j\partial_k\partial_l\partial_m u_1(\omega)\overline{\partial^i\partial^j\partial^k\partial^l u_2(\omega)}\,d\sigma(\omega).
\end{aligned} \tag{A-5}$$

Hence

$$\begin{aligned}
\mathrm{Re}(\widetilde{\boldsymbol{L}}\boldsymbol{u}\,|\,\boldsymbol{u})_5 &\le -\tfrac{5}{2}\|\boldsymbol{u}\|_5^2 - \tfrac{1}{2}\int_{\mathbb{S}^8} \partial_i\partial_j\partial_k\partial_l\partial_m u_1(\omega)\overline{\partial^i\partial^j\partial^k\partial^l\partial^m u_1(\omega)}\,d\sigma(\omega) \\
&\qquad - \tfrac{1}{2}\int_{\mathbb{S}^8} \partial_i\partial_j\partial_k\partial_l u_2(\omega)\overline{\partial^i\partial^j\partial^k\partial^l u_2(\omega)}\,d\sigma(\omega) \\
&\qquad\qquad + \mathrm{Re}\int_{\mathbb{S}^8} \omega^m\partial_i\partial_j\partial_k\partial_l\partial_m u_1(\omega)\overline{\partial^i\partial^j\partial^k\partial^l u_2(\omega)}\,d\sigma(\omega) \le -\tfrac{5}{2}\|\boldsymbol{u}\|_5^2,
\end{aligned} \tag{A-6}$$

where we used $\mathrm{Re}(a\bar{b}) \le \frac{1}{2}(|a|^2 + |b|^2)$ as well as the bound

$$\Big|\sum_k \omega_k \partial_k u(\omega)\Big|^2 \le \sum_k (\omega_k)^2 \sum_k |\partial_k u(\omega)|^2 = \sum_k |\partial_k u(\omega)|^2.$$

A similar calculation yields

$$\mathrm{Re}(\widetilde{\boldsymbol{L}}\,|\,\boldsymbol{u})_4 \le -\tfrac{3}{2}\|\boldsymbol{u}\|_4^2.$$

In view of the logic of these estimates, it is clear that we cannot use the standard homogeneous inner products for integer regularities lower than $j = 3$, since the bound shifts to the right and will be positive eventually. For this reason, we use tailor-made expressions for the remaining $H^3(\mathbb{B}^9) \times H^2(\mathbb{B}^9)$ part. In the following, we prove that

$$\sum_{j=1}^{3} \mathrm{Re}(\widetilde{\boldsymbol{L}}\boldsymbol{u}\,|\,\boldsymbol{u})_j \le -\tfrac{1}{2}\sum_{j=1}^{3} \|\boldsymbol{u}\|_j^2, \tag{A-7}$$

which in combination with the above bounds implies the first claim in Lemma 3.4 for $d = 9$ and $k = 5$ (and in fact for any $3 \le k \le 5$.) For higher regularities, we add again the corresponding standard homogeneous parts. Analogous to the above calculations, one shows that

$$\begin{aligned}
&\mathrm{Re}\int_{\mathbb{B}^9} \partial_{i_1\cdots i_k}[\widetilde{\boldsymbol{L}}\boldsymbol{u}]_1(\xi)\overline{\partial^{i_1\cdots i_k} u_1(\xi)}\,d\xi \\
&\quad = \big(\tfrac{5}{2} - k\big)\int_{\mathbb{B}^9} \partial_{i_1\cdots i_k} u_1(\xi)\overline{\partial^{i_1\cdots i_k} u_1(\xi)}\,d\xi - \tfrac{1}{2}\int_{\mathbb{S}^8} \partial_{i_1\cdots i_k} u_1(\omega)\overline{\partial^{i_1\cdots i_k} u_1(\omega)}\,d\sigma(\omega) \\
&\qquad\qquad + \mathrm{Re}\int_{\mathbb{B}^9} \partial_{i_1\cdots i_k} u_1(\xi)\overline{\partial^{i_1\cdots i_k} u_2(\xi)}\,d\xi
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{Re}\int_{\mathbb{B}^9} \partial_{i_1\cdots i_{k-1}}[\widetilde{\boldsymbol{L}}\boldsymbol{u}]_2(\xi)\overline{\partial^{i_1\cdots i_{k-1}} u_2(\xi)}\,d\xi \\
&\quad = \big(\tfrac{5}{2} - k\big)\int_{\mathbb{B}^9} \partial_{i_1\cdots i_{k-1}} u_2(\xi)\overline{\partial^{i_1\cdots i_{k-1}} u_2(\xi)}\,d\xi - \mathrm{Re}\int_{\mathbb{B}^9} \partial_{i_1\cdots i_k} u_1(\xi)\overline{\partial^{i_1\cdots i_k} u_2(\xi)}\,d\xi \\
&\qquad + \mathrm{Re}\int_{\mathbb{S}^8} \omega^{i_k}\partial_{i_1\cdots i_k} u_1(\omega)\overline{\partial^{i_1\cdots i_{k-1}} u_2(\omega)}\,d\sigma(\omega) - \tfrac{1}{2}\int_{\mathbb{S}^8} \partial_{i_1\cdots i_{k-1}} u_2(\omega)\overline{\partial^{i_1\cdots i_{k-1}} u_2(\omega)}\,d\sigma(\omega).
\end{aligned}$$

As in (A-6), we thus obtain for $j \ge 6$ the bound

$$\mathrm{Re}(\widetilde{\boldsymbol{L}}\boldsymbol{u}\,|\,\boldsymbol{u})_j \le \big(\tfrac{5}{2} - j\big)\|\boldsymbol{u}\|_j^2.$$

It is left to prove (A-7). We first consider $\mathrm{Re}(\widetilde{\boldsymbol{L}}\boldsymbol{u}\,|\,\boldsymbol{u})_3$. Using (A-2), (A-1), and the divergence theorem, we calculate

$$\begin{aligned}
&\mathrm{Re}\int_{\mathbb{B}^9} \partial_i\partial_j\partial_k[\widetilde{\boldsymbol{L}}\boldsymbol{u}]_1(\xi)\overline{\partial^i\partial^j\partial^k u_1(\xi)}\,d\xi \\
&\quad = -\tfrac{1}{2}\int_{\mathbb{B}^9} \partial_i\partial_j\partial_k u_1(\xi)\overline{\partial^i\partial^j\partial^k u_1(\xi)}\,d\xi - \tfrac{1}{2}\int_{\mathbb{S}^8} \partial_i\partial_j\partial_k u_1(\omega)\overline{\partial^i\partial^j\partial^k u_1(\omega)}\,d\sigma(\omega) \\
&\qquad\qquad + \mathrm{Re}\int_{\mathbb{B}^9} \partial_i\partial_j\partial_k u_2(\xi)\overline{\partial^i\partial^j\partial^k u_1(\xi)}\,d\xi.
\end{aligned}$$

An application of (A-4) shows

$$\begin{aligned}
&\operatorname{Re}\int_{\mathbb{B}^9} \partial_i\partial_j[\widetilde{\boldsymbol{L}}\boldsymbol{u}]_2(\xi)\overline{\partial^i\partial^j u_2(\xi)}\,d\xi\\
&\quad=\operatorname{Re}\int_{\mathbb{B}^9} \partial_i\partial_j\partial^k\partial_k u_1(\xi)\overline{\partial^i\partial^j u_2(\xi)}\,d\xi-\tfrac{1}{2}\int_{\mathbb{B}^9}\partial_i\partial_j u_2(\xi)\overline{\partial^i\partial^j u_2(\xi)}\,d\xi\\
&\qquad\qquad-\tfrac{1}{2}\int_{\mathbb{S}^8}\partial_i\partial_j u_2(\omega)\overline{\partial^i\partial^j u_2(\omega)}\,d\sigma(\omega)\\
&\quad=\operatorname{Re}\int_{\mathbb{S}^8}\omega^k\partial_k\partial_i\partial_j u_1(\omega)\overline{\partial^i\partial^j u_2(\omega)}\,d\sigma(\omega)-\operatorname{Re}\int_{\mathbb{B}^9}\partial_i\partial_j\partial_k u_1(\xi)\overline{\partial^i\partial^j\partial^k u_2(\xi)}\,d\xi\\
&\qquad\qquad-\tfrac{1}{2}\int_{\mathbb{B}^9}\partial_i\partial_j u_2(\xi)\overline{\partial^i\partial^j u_2(\xi)}\,d\xi-\tfrac{1}{2}\int_{\mathbb{S}^8}\partial_i\partial_j u_2(\omega)\overline{\partial^i\partial^j u_2(\omega)}\,d\sigma(\omega),
\end{aligned}$$

and finally

$$\begin{aligned}
&\int_{\mathbb{S}^8}\partial_i\partial_j[\widetilde{\boldsymbol{L}}\boldsymbol{u}]_1(\omega)\overline{\partial^i\partial^j u_1(\omega)}\,d\sigma(\omega)\\
&\quad=-\operatorname{Re}\int_{\mathbb{S}^8}\omega^k\partial_k\partial_i\partial_j u_1(\omega)\overline{\partial^i\partial^j u_1(\omega)}\,d\sigma(\omega)\\
&\qquad\qquad-4\int_{\mathbb{S}^8}\partial_i\partial_j u_1(\omega)\overline{\partial^i\partial^j u_1(\omega)}\,d\sigma(\omega)+\operatorname{Re}\int_{\mathbb{S}^8}\partial_i\partial_j u_2(\omega)\overline{\partial^i\partial^j u_1(\omega)}\,d\sigma(\omega).
\end{aligned}$$

In summary, we infer that

$$\operatorname{Re}(\widetilde{\boldsymbol{L}}\boldsymbol{u}\,|\,\boldsymbol{u})_3=-\tfrac{1}{2}\|\boldsymbol{u}\|_3^2-12\int_{\mathbb{S}^8}\partial_i\partial_j u_1(\omega)\overline{\partial^i\partial^j u_1(\omega)}\,d\sigma(\omega)+4\int_{\mathbb{S}^8}A(\omega)\,d\sigma(\omega),$$

where

$$\begin{aligned}
A(\omega)=&-\tfrac{1}{2}\partial_i\partial_j\partial_k u_1(\omega)\overline{\partial^i\partial^j\partial^k u_1(\omega)}-\tfrac{1}{2}\partial_i\partial_j u_2(\omega)\overline{\partial^i\partial^j u_2(\omega)}\\
&-\tfrac{1}{2}\partial_i\partial_j u_1(\omega)\overline{\partial^i\partial^j u_1(\omega)}+\operatorname{Re}(\omega^k\partial_i\partial_j\partial_k u_1(\omega)\overline{\partial^i\partial^j u_2(\omega)})\\
&-\operatorname{Re}(\omega^k\partial_i\partial_j\partial_k u_1(\omega)\overline{\partial^i\partial^j u_1(\omega)})+\operatorname{Re}(\partial_i\partial_j u_2(\omega)\overline{\partial^i\partial^j u_1(\omega)}).
\end{aligned}$$

By using the inequality

$$\operatorname{Re}(a\bar{b})+\operatorname{Re}(a\bar{c})-\operatorname{Re}(b\bar{c})\le\tfrac{1}{2}(|a|^2+|b|^2+|c|^2),\quad a,b,c\in\mathbb{C},$$

we get $A(\omega)\le 0$. Analogously, to estimate $\operatorname{Re}(\widetilde{\boldsymbol{L}}\boldsymbol{u}\,|\,\boldsymbol{u})_2$, we get

$$\begin{aligned}
&\operatorname{Re}\int_{\mathbb{B}^9}\partial_i\partial^j\partial_j[\widetilde{\boldsymbol{L}}\boldsymbol{u}]_1(\xi)\overline{\partial^i\partial_l\partial^l u_1(\xi)}\,d\xi\\
&\quad=-\tfrac{1}{2}\operatorname{Re}\int_{\mathbb{B}^9}\partial_i\partial^j\partial_j u_1(\xi)\overline{\partial^i\partial_l\partial^l u_1(\xi)}\,d\xi-\tfrac{1}{2}\int_{\mathbb{S}^8}\partial_i\partial^j\partial_j u_1(\omega)\overline{\partial^i\partial_l\partial^l u_1(\omega)}\,d\sigma(\omega)\\
&\qquad\qquad+\operatorname{Re}\int_{\mathbb{B}^9}\partial_i\partial^j\partial_j u_1(\xi)\overline{\partial^i\partial_l\partial^l u_2(\xi)}\,d\xi
\end{aligned}$$

and

$$\begin{aligned}
&\operatorname{Re}\int_{\mathbb{S}^8}\partial_i[\widetilde{\boldsymbol{L}}\boldsymbol{u}]_2(\omega)\overline{\partial^i u_2(\omega)}\,d\sigma(\omega)\\
&\quad=\operatorname{Re}\int_{\mathbb{S}^8}\partial_i\partial^j\partial_j u_1(\omega)\overline{\partial^i u_2(\omega)}\,d\sigma(\omega)\\
&\qquad\qquad-\operatorname{Re}\int_{\mathbb{S}^8}\omega^j\partial_i\partial_j u_2(\omega)\overline{\partial^i u_2(\omega)}\,d\sigma(\omega)-4\operatorname{Re}\int_{\mathbb{S}^8}\partial_i u_2(\omega)\overline{\partial^i u_2(\omega)}\,d\sigma(\omega).
\end{aligned}$$

For the remaining term, we do a similar calculation as in (A-5), but we use instead (A-3) in order to cancel the mixed term. In summary, we obtain

$$\operatorname{Re}(\widetilde{\boldsymbol{L}}\boldsymbol{u}\,|\,\boldsymbol{u})_2 = -\tfrac{1}{2}\|\boldsymbol{u}\|_2^2 - 3\int_{\mathbb{S}^8} \partial_i u_2(\omega)\overline{\partial^i u_2(\omega)}\,d\sigma(\omega) + \int_{\mathbb{S}^8} B(\omega)\,d\sigma(\omega),$$

where

$$\begin{aligned} B(\omega) = {} & -\tfrac{1}{2}\partial_i\partial^j\partial_j u_1(\omega)\overline{\partial^i\partial_l\partial^l u_1(\omega)} - \tfrac{1}{2}\partial_i\partial_j u_2(\omega)\overline{\partial^i\partial^j u_2(\omega)} - \tfrac{1}{2}\partial_i u_2(\omega)\overline{\partial^i u_2(\omega)} \\ & + \operatorname{Re}(\omega^k\partial_i\partial^j\partial_j u_1(\omega)\overline{\partial^i\partial_k u_2(\omega)}) + \operatorname{Re}(\partial_i\partial^j\partial_j u_1(\omega)\overline{\partial^i u_2(\omega)}) - \operatorname{Re}(\omega^j\partial_i\partial_j u_2(\omega)\overline{\partial^i u_2(\omega)}), \end{aligned}$$

and we observe that $B(\omega) \le 0$. Now, we consider $\operatorname{Re}(\widetilde{\boldsymbol{L}}\boldsymbol{u}\,|\,\boldsymbol{u})_1$, which consists only of boundary integrals. For the first term, we get

$$\begin{aligned} & \operatorname{Re}\int_{\mathbb{S}^8} \partial_i[\boldsymbol{L}\boldsymbol{u}]_1(\omega)\overline{\partial^i u_1(\omega)}\,d\sigma(\omega) \\ & \qquad = -3\operatorname{Re}\int_{\mathbb{S}^8} \partial_i u_1(\omega)\overline{\partial^i u_1(\omega)}\,d\sigma(\omega) - \operatorname{Re}\int_{\mathbb{S}^8} \omega^j\partial_i\partial_j u_1(\omega)\overline{\partial^i u_1(\omega)}\,d\sigma(\omega) \\ & \qquad\qquad + \operatorname{Re}\int_{\mathbb{S}^8} \partial_i u_2(\omega)\overline{\partial^i u_1(\omega)}\,d\sigma(\omega). \end{aligned}$$

By the Cauchy–Schwarz inequality,

$$\begin{aligned} & \operatorname{Re}\int_{\mathbb{S}^8} (\partial_i u_2(\omega) - \omega^j\partial_i\partial_j u_1(\omega))\overline{\partial^i u_1(\omega)}\,d\sigma(\omega) \\ & \quad \le \tfrac{1}{2}\int_{\mathbb{S}^8} \partial_i u_1(\omega)\overline{\partial^i u_1(\omega)}\,d\sigma(\omega) + \int_{\mathbb{S}^8} \partial_i u_2(\omega)\overline{\partial^i u_2(\omega)}\,d\sigma(\omega) + \int_{\mathbb{S}^8} \partial_i\partial_j u_1(\omega)\overline{\partial^i\partial^j u_1(\omega)}\,d\sigma(\omega), \end{aligned}$$

which implies

$$\begin{aligned} & \operatorname{Re}\int_{\mathbb{S}^8} \partial_i[\boldsymbol{L}\boldsymbol{u}]_1(\omega)\overline{\partial^i u_1(\omega)}\,d\sigma(\omega) \\ & \le -\tfrac{5}{2}\operatorname{Re}\int_{\mathbb{S}^8} \partial_i u_1(\omega)\overline{\partial^i u_1(\omega)}\,d\sigma(\omega) + \int_{\mathbb{S}^8} \partial_i u_2(\omega)\overline{\partial^i u_2(\omega)}\,d\sigma(\omega) + \int_{\mathbb{S}^8} \partial_i\partial_j u_1(\omega)\overline{\partial^i\partial^j u_1(\omega)}\,d\sigma(\omega). \end{aligned}$$

Analogously,

$$\begin{aligned} & \operatorname{Re}\int_{\mathbb{S}^8} [\widetilde{\boldsymbol{L}}\boldsymbol{u}]_2(\omega)\overline{u_2(\omega)}\,d\sigma(\omega) \\ & \qquad = -3\int_{\mathbb{S}^8} |u_2(\omega)|^2\,d\sigma(\omega) + \operatorname{Re}\int_{\mathbb{S}^8} \partial^i\partial_i u_1(\omega)\overline{u_2(\omega)}\,d\sigma(\omega) - \operatorname{Re}\int_{\mathbb{S}^8} \omega^i\partial_i u_2(\omega)\overline{u_2(\omega)}\,d\sigma(\omega) \\ & \qquad \le -\tfrac{5}{2}\int_{\mathbb{S}^8} |u_2(\omega)|^2\,d\sigma(\omega) + \int_{\mathbb{S}^8} |\Delta u_1(\omega)|^2\,d\sigma(\omega) + \int_{\mathbb{S}^8} \partial_i u_2(\omega)\overline{\partial^i u_2(\omega)}\,d\sigma(\omega), \end{aligned}$$

$$\begin{aligned} & \operatorname{Re}\int_{\mathbb{S}^8} [\widetilde{\boldsymbol{L}}\boldsymbol{u}]_1(\omega)\overline{u_1(\omega)}\,d\sigma(\omega) \\ & \qquad = -2\int_{\mathbb{S}^8} |u_1(\omega)|^2\,d\sigma(\omega) - \operatorname{Re}\int_{\mathbb{S}^8} \omega^i\partial_i u_1(\omega)\overline{u_1(\omega)}\,d\sigma(\omega) + \operatorname{Re}\int_{\mathbb{S}^8} u_2(\omega)\overline{u_1(\omega)}\,d\sigma(\omega) \\ & \qquad \le -\tfrac{3}{2}\int_{\mathbb{S}^8} |u_1(\omega)|^2\,d\sigma(\omega) + \int_{\mathbb{S}^8} \partial_i u_1(\omega)\overline{\partial^i u_1(\omega)}\,d\sigma(\omega) + \int_{\mathbb{S}^8} |u_2(\omega)|^2\,d\sigma(\omega), \end{aligned}$$

and hence

$$\operatorname{Re}(\widetilde{\boldsymbol{L}}\boldsymbol{u}\,|\,\boldsymbol{u})_1 \le -\tfrac{3}{2}\|\boldsymbol{u}\|_1^2 + 2\int_{\mathbb{S}^8} \partial_i u_2(\omega)\overline{\partial^i u_2(\omega)}\,d\sigma(\omega) + \int_{\mathbb{S}^8} \partial_i\partial_j u_1(\omega)\overline{\partial^i\partial^j u_1(\omega)}\,d\sigma(\omega) + \int_{\mathbb{S}^8} |\Delta u_1(\omega)|^2\,d\sigma(\omega).$$

In conclusion,

$$\sum_{j=1}^{3} \operatorname{Re}(\widetilde{\boldsymbol{L}}\boldsymbol{u}\,|\,\boldsymbol{u})_j \le -\tfrac{1}{2}\sum_{j=1}^{3}\|\boldsymbol{u}\|_j^2 - 11\int_{\mathbb{S}^8} \partial_i\partial_j u_1(\omega)\overline{\partial^i\partial^j u_1(\omega)}\,d\sigma(\omega) + \int_{\mathbb{S}^8} |\Delta u_1(\omega)|^2\,d\sigma(\omega).$$

By the Cauchy–Schwarz inequality,

$$|\Delta u(\omega)|^2 \le \Big|\sum_{i=1}^{9} \partial_i^2 u(\omega)\Big|^2 \le 9\sum_{i=1}^{9} |\partial_i^2 u(\omega)|^2 \le 9\sum_{i,j=1}^{9} |\partial_i\partial_j u(\omega)|^2,$$

which proves (A-7).

Analogous calculations for $d = 7$ and $k \ge 3$ yield an even better bound, namely,

$$\operatorname{Re}(\widetilde{\boldsymbol{L}}\boldsymbol{u}\,|\,\boldsymbol{u})_{\mathcal{H}_k} \le -\tfrac{3}{2}\|\boldsymbol{u}\|_{\mathcal{H}_k}^2 \tag{A-8}$$

for all $\boldsymbol{u} \in \mathcal{D}(\widetilde{\boldsymbol{L}})$, from which we obtain in particular the claimed estimate. Another way to see that (A-8) holds is by Lemma 3.2 of [Glogić and Schörkhuber 2021], which is formulated in terms of the above inner product for the specific case $d = 7$ and $k = 3$. The operator considered there corresponds to $\widetilde{\boldsymbol{L}}$ shifted by a constant, which immediately gives the inequality (A-8). □

References

[Atkinson and Han 2012] K. Atkinson and W. Han, *Spherical harmonics and approximations on the unit sphere: an introduction*, Lecture Notes in Math. **2044**, Springer, 2012. MR Zbl

[Biernat et al. 2017] P. Biernat, P. Bizoń, and M. Maliborski, "Threshold for blowup for equivariant wave maps in higher dimensions", *Nonlinearity* **30**:4 (2017), 1513–1522. MR Zbl

[Bizoń 2001] P. Bizoń, "Threshold behavior for nonlinear wave equations", *J. Nonlinear Math. Phys.* **8**:suppl. (2001), 35–41. MR Zbl

[Bizoń 2002] P. Bizoń, "Formation of singularities in Yang–Mills equations", *Acta Phys. Polon. B* **33**:7 (2002), 1893–1922. MR Zbl

[Bizoń and Tabor 2001] P. Bizoń and Z. Tabor, "On blowup of Yang–Mills fields", *Phys. Rev. D* (3) **64**:12 (2001), art. id. 121701. MR

[Bizoń et al. 2000] P. Bizoń, T. Chmaj, and Z. Tabor, "Dispersion and collapse of wave maps", *Nonlinearity* **13**:4 (2000), 1411–1423. MR Zbl

[Bizoń et al. 2004] P. Bizoń, T. Chmaj, and Z. Tabor, "On blowup for semilinear wave equations with a focusing nonlinearity", *Nonlinearity* **17**:6 (2004), 2187–2201. MR Zbl

[Bizoń et al. 2007] P. Bizoń, D. Maison, and A. Wasserman, "Self-similar solutions of semilinear wave equations with a focusing nonlinearity", *Nonlinearity* **20**:9 (2007), 2061–2074. MR Zbl

[Chatzikaleas and Donninger 2019] A. Chatzikaleas and R. Donninger, "Stable blowup for the cubic wave equation in higher dimensions", *J. Differential Equations* **266**:10 (2019), 6809–6865. MR Zbl

[Collot 2018] C. Collot, *Type II blow up manifolds for the energy supercritical semilinear wave equation*, Mem. Amer. Math. Soc. **1205**, Amer. Math. Soc., Providence, RI, 2018. MR Zbl

[Costin et al. 2016] O. Costin, R. Donninger, I. Glogić, and M. Huang, "On the stability of self-similar solutions to nonlinear wave equations", *Comm. Math. Phys.* **343**:1 (2016), 299–310. MR Zbl

[Costin et al. 2017] O. Costin, R. Donninger, and I. Glogić, "Mode stability of self-similar wave maps in higher dimensions", *Comm. Math. Phys.* **351**:3 (2017), 959–972. MR Zbl

[Dai and Duyckaerts 2021] W. Dai and T. Duyckaerts, "Self-similar solutions of focusing semi-linear wave equations in $\mathbb{R}^N$", *J. Evol. Equ.* **21**:4 (2021), 4703–4750. MR Zbl

[DLMF 2010] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark (editors), *NIST handbook of mathematical functions*, US Dept. Commerce, Washington, DC, 2010. MR Zbl

[Donninger and Schörkhuber 2016] R. Donninger and B. Schörkhuber, "On blowup in supercritical wave equations", *Comm. Math. Phys.* **346**:3 (2016), 907–943. MR Zbl

[Donninger and Schörkhuber 2017] R. Donninger and B. Schörkhuber, "Stable blowup for wave equations in odd space dimensions", *Ann. Inst. H. Poincaré C Anal. Non Linéaire* **34**:5 (2017), 1181–1213. MR Zbl

[Elaydi 2005] S. Elaydi, *An introduction to difference equations*, 3rd ed., Springer, 2005. MR Zbl

[Engel and Nagel 2000] K.-J. Engel and R. Nagel, *One-parameter semigroups for linear evolution equations*, Grad. Texts in Math. **194**, Springer, 2000. MR Zbl

[Glogić 2018] I. Glogić, *On the existence and stability of self-similar blowup in nonlinear wave equations*, Ph.D. thesis, Ohio State University, 2018, available at https://www.proquest.com/docview/2150090828.

[Glogić 2022] I. Glogić, "Stable blowup for the supercritical hyperbolic Yang–Mills equations", *Adv. Math.* **408**:B (2022), art. id. 108633. MR Zbl

[Glogić and Schörkhuber 2021] I. Glogić and B. Schörkhuber, "Co-dimension one stable blowup for the supercritical cubic wave equation", *Adv. Math.* **390** (2021), art. id. 107930. MR Zbl

[Glogić et al. 2020] I. Glogić, M. Maliborski, and B. Schörkhuber, "Threshold for blowup for the supercritical cubic wave equation", *Nonlinearity* **33**:5 (2020), 2143–2158. MR Zbl

[Hislop and Sigal 1996] P. D. Hislop and I. M. Sigal, *Introduction to spectral theory: with applications to Schrödinger operators*, Appl. Math. Sci. **113**, Springer, 1996. MR Zbl

[Kato 1976] T. Kato, *Perturbation theory for linear operators*, 2nd ed., Grundlehren der Math. Wissenschaften **132**, Springer, 1976. MR Zbl

[Kavian and Weissler 1990] O. Kavian and F. B. Weissler, "Finite energy self-similar solutions of a nonlinear wave equation", *Comm. Partial Differential Equations* **15**:10 (1990), 1381–1420. MR Zbl

[Kenig and Merle 2008] C. E. Kenig and F. Merle, "Global well-posedness, scattering and blow-up for the energy-critical focusing non-linear wave equation", *Acta Math.* **201**:2 (2008), 147–212. MR Zbl

[Krieger and Schlag 2017] J. Krieger and W. Schlag, "Large global solutions for energy supercritical nonlinear wave equations on $\mathbb{R}^{3+1}$", *J. Anal. Math.* **133** (2017), 91–131. MR Zbl

[Krieger et al. 2015] J. Krieger, K. Nakanishi, and W. Schlag, "Center-stable manifold of the ground state in the energy space for the critical wave equation", *Math. Ann.* **361**:1-2 (2015), 1–50. MR Zbl

[Kycia 2011] R. Kycia, "On self-similar solutions of semilinear wave equations in higher space dimensions", *Appl. Math. Comput.* **217**:22 (2011), 9451–9466. MR Zbl

[Levine 1974] H. A. Levine, "Instability and nonexistence of global solutions to nonlinear wave equations of the form $Pu_{tt} = -Au + \mathcal{F}(u)$", *Trans. Amer. Math. Soc.* **192** (1974), 1–21. MR Zbl

[Lindblad and Sogge 1995] H. Lindblad and C. D. Sogge, "On existence and scattering with minimal regularity for semilinear wave equations", *J. Funct. Anal.* **130**:2 (1995), 357–426. MR Zbl

[Pazy 1983] A. Pazy, *Semigroups of linear operators and applications to partial differential equations*, Appl. Math. Sci. **44**, Springer, 1983. MR Zbl

[Rudin 1976] W. Rudin, *Principles of mathematical analysis*, 3rd ed., McGraw-Hill, New York, 1976. MR Zbl

[Simon 2015] B. Simon, *Operator theory*, Comprehensive Course in Anal. **4**, Amer. Math. Soc., Providence, RI, 2015. MR Zbl

[Titchmarsh 1939] E. C. Titchmarsh, *The theory of functions*, 2nd ed., Oxford Univ. Press, 1939. MR Zbl

Received 8 Oct 2021. Accepted 11 Jul 2022.

ELEK CSOBO: csoboelek@gmail.com
Universität Innsbruck, Institut für Mathematik, Technikerstraße 13, 6020 Innsbruck, Austria

IRFAN GLOGIĆ: irfan.glogic@univie.ac.at
Faculty of Mathematics, University of Vienna, Vienna, Austria

BIRGIT SCHÖRKHUBER: birgit.schoerkhuber@uibk.ac.at
Universität Innsbruck, Institut für Mathematik, Technikerstraße 13, 6020 Innsbruck, Austria

ANALYSIS AND PDE
Vol. 17 (2024), No. 2, pp. 681–722
DOI: 10.2140/apde.2024.17.681

ARNOLD'S VARIATIONAL PRINCIPLE AND ITS APPLICATION TO THE STABILITY OF PLANAR VORTICES

THIERRY GALLAY AND VLADIMÍR ŠVERÁK

We consider variational principles related to V. I. Arnold's stability criteria for steady-state solutions of the two-dimensional incompressible Euler equation. Our goal is to investigate under which conditions the quadratic forms defined by the second variation of the associated functionals can be used in the stability analysis, both for the Euler evolution and for the Navier–Stokes equation at low viscosity. In particular, we revisit the classical example of Oseen's vortex, providing a new stability proof with stronger geometric flavor. Our analysis involves a fairly detailed functional-analytic study of the inviscid case, which may be of independent interest, and a careful investigation of the influence of the viscous term in the particular example of the Gaussian vortex.

1. Introduction

We investigate the applicability of V. I. Arnold's geometric methods to certain stability problems related to Navier–Stokes vortices at high Reynolds number. Our main goal is a "proof of concept" that such applications are possible, at least in simple cases, even though much of the geometric structure behind the inviscid stability analysis does not survive the addition of the viscosity term. In particular, we give a new proof of a known result concerning the stability of Oseen's vortex as a steady state of the Navier–Stokes equation in self-similar variables. We expect that the approach we advertise here will be useful to tackle stability problems involving solutions that are less symmetric and less explicit than the classical Oseen vortex. In such cases one may not have good alternative methods for proving stability in the presence of viscosity. Our investigation leads to a detailed study of the quadratic forms naturally arising in Arnold's approach. Some of their functional-analytic properties, which are established in the course of our analysis, may be of independent interest.

1A. ***A finite-dimensional model.*** Following the seminal paper [Arnold 1965], we first illustrate the issues we want to address in a model situation where the "phase space" is finite-dimensional. We consider the ordinary differential equation

$$\dot{x} = b(x), \quad x \in \mathbb{R}^n, \tag{1-1}$$

where b is a smooth vector field in $\mathbb{R}^n$. Let us assume that $f, g_1, \dots, g_m : \mathbb{R}^n \to \mathbb{R}$ are (sufficiently smooth) conserved quantities for the evolution (1-1), with $m < n$. This means

$$f'(x)b(x) = 0 \quad \text{and} \quad g_j'(x)b(x) = 0, \quad x \in \mathbb{R}^n, \ j = 1, \dots, m, \tag{1-2}$$

MSC2020: 35Q30, 35Q31.
Keywords: two-dimensional flows, vortices, stability, variational principle, constrained optimization.

where we adopt the standard notation $f'(x)$ for the linear form given by the first derivative of f at x. The situation we have ultimately in mind is somewhat more specific: it corresponds to the case where the phase space $\mathbb{R}^n$ is equipped with a Poisson bracket $\{\cdot,\cdot\}$, where system (1-1) is of the form

$$\dot{x} = \{f, x\}, \tag{1-3}$$

and where $g_1, \dots, g_m$ are Casimir functions. The Poisson structure is of course important in many respects, but for our arguments here it does not play a big role. We can therefore proceed in the general context of (1-1) and (1-2).

For any $c = (c_1, \dots, c_m) \in \mathbb{R}^m$, let us define $X_c = \{x \in \mathbb{R}^n : g_1(x) = c_1, \dots, g_m(x) = c_m\}$. We assume that, for some $c \in \mathbb{R}^m$, the function f attains a *nondegenerate local maximum* on X_c at some point $\bar{x} \in X_c$ and that the derivatives $g_1'(\bar{x}), \dots, g_m'(\bar{x})$ are linearly independent. The stationarity condition at $\bar{x}$ gives the linear relation

$$f'(\bar{x}) - \sum_{j=1}^{m} \lambda_j g_j'(\bar{x}) = 0 \tag{1-4}$$

for some Lagrange multipliers $\lambda_1, \dots, \lambda_m \in \mathbb{R}$. Moreover, the second-order differential[1] of the function $f|_{X_c}$ (the restriction of f to X_c) at $\bar{x}$ is given by the restriction to the tangent space $T_{\bar{x}}X_c$ of the quadratic form

$$\mathcal{Q} = f''(\bar{x}) - \sum_{j=1}^{m} \lambda_j g_j''(\bar{x}), \tag{1-5}$$

where we denote by $f''(\bar{x})$ the quadratic form given by the Hessian of f at $\bar{x}$, and similarly for $g_1''(\bar{x}), \dots, g_m''(\bar{x})$. Our nondegeneracy assumption means that the restriction of the form $\mathcal{Q}$ to $T_{\bar{x}}X_c$ is strictly negative definite. Now, let $B = b'(\bar{x})$ be the $n \times n$ matrix corresponding to the linearization of (1-1) at the point $\bar{x}$, which is a steady state by construction [Arnold 1965]. If we differentiate twice the relations (1-2) and use (1-4) together with $b(\bar{x}) = 0$, we see that the evolution defined by the linearized equation $\dot{\xi} = B\xi$ leaves the form $\mathcal{Q}$ invariant. In other words,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{Q}(\xi, \xi) = \mathcal{Q}(B\xi, \xi) + \mathcal{Q}(\xi, B\xi) = 0 \quad \text{for all } \xi \in \mathbb{R}^n. \tag{1-6}$$

The above structure[2] gives various options for the stability analysis of the equilibrium $\bar{x}$ of (1-1), depending on the index of the quadratic form $\mathcal{Q}$ in (1-5). Our assumptions readily imply that $\bar{x}$ is stable in the sense of Lyapunov with respect to perturbations on the invariant submanifold X_c. Moreover, since a neighborhood of $\bar{x}$ in $\mathbb{R}^n$ is foliated by submanifolds of this form for nearby values of the parameter $c = (c_1, \dots, c_m)$, one can show that $\bar{x}$ is in fact Lyapunov stable with respect to small *unconstrained* perturbations [Arnold 1965]. The perspective changes qualitatively if we add to the vector field b in (1-1) a small "dissipative" term, with the effect that the quantities f and $g_1, \dots, g_m$ are no longer exactly

[1] We recall that the second-order differential of a function on a manifold is intrinsically defined at the points where the first-order differential vanishes.

[2] Pointed out in [Arnold 1965] in the form we use here, although in the finite-dimensional case these ideas go back to the founders of the analytical mechanics.

conserved under the modified evolution. This is in the spirit of what we intend to do in the infinite-dimensional case, when we consider the Navier–Stokes equation as a perturbation of the Euler equation. Since the evolution no longer takes place on the manifolds X_c, the argument above leading to unconstrained Lyapunov stability is not applicable anymore. However, in good situations, stability can still be obtained if the quadratic form $\mathcal{Q}$ in (1-5) happens to be negative definite not just on $T_{\bar{x}}X_c$, but on larger subspaces as well, for instance on the whole space $\mathbb{R}^n$. This is, roughly speaking, the idea we shall pursue in the infinite-dimensional case, to study the stability of vortex-like solutions of the Navier–Stokes equation.

To conclude with the (unmodified) evolution (1-1), we emphasize that the problem of determining the index of the form (1-5) is also very natural from the viewpoint of the usual constrained optimization theory. Clearly, the "Lagrange function"

$$\mathcal{L}(x) = f(x) - \sum_{j=1}^{m} \lambda_j g_j(x), \quad x \in \mathbb{R}^n, \tag{1-7}$$

when considered on the whole space $\mathbb{R}^n$, has a critical point at $\bar{x}$ (and a local maximum at $\bar{x}$ when restricted to X_c). The form $\mathcal{Q}$ will be strictly negative definite[3] in the whole space $\mathbb{R}^n$ if and only if $\mathcal{L}$ has a nondegenerate *unconstrained* maximum at $\bar{x}$. As is explained in Section 2D, this is related to the concavity of the function

$$(c_1, \dots, c_m) \longmapsto M(c_1, \dots, c_m) := \sup_{x \in X_c} f(x). \tag{1-8}$$

1B. ***Arnold's geometric view of the two-dimensional incompressible Euler equation.*** V. I. Arnold [1966b; 1966a] (see also [Arnold and Khesin 1998]) carried out the analogue of the above calculations in an infinite-dimensional setting to handle in particular the two-dimensional incompressible Euler equation $\partial_t \omega + u \cdot \nabla \omega = 0$, where u denotes the velocity of the fluid and $\omega = \operatorname{curl} u$ is the associated vorticity. In this case the evolution is generated by the Hamiltonian function, which represents the kinetic energy of the fluid, and the constraints are given by the Casimir functionals

$$\mathcal{C}_\Phi(\omega) = \int_\Omega \Phi(\omega(x))\, \mathrm{d}x, \tag{1-9}$$

where $\Omega \subset \mathbb{R}^2$ is the fluid domain and Φ is an "arbitrary" function on $\mathbb{R}$. The idea of maximizing or minimizing the energy on the set of vorticities satisfying suitable constraints has been widely used since then to study the stability of steady-state solutions of the two-dimensional Euler equations and related fluid models; see [Arnold and Khesin 1998; Burton 2005; Cao et al. 2019].

Let us briefly recall the setup relevant for our goals here, making the similarities with the finite-dimensional case as transparent as possible. Our main objects will be the following:

(1) The *phase space* $\mathcal{P} = \{\omega \colon \mathbb{R}^2 \to (0, \infty) : \omega \text{ is smooth and decays "sufficiently fast" at } \infty\}$. This is our infinite-dimensional replacement for the manifold $\mathbb{R}^n$ in the finite-dimensional model. We restrict ourselves to positive vorticity distributions defined on $\Omega = \mathbb{R}^2$, because this is the appropriate framework to study the stability of radially symmetric vortices in the whole plane. Admittedly, the definition above

[3] Our use of the terms "positive definite" and "negative definite" allows for vanishing along some directions. When this is not the case, we speak of strictly positive definite or strictly negative definite forms.

is somewhat vague, but it serves only as a motivation and our results will be independent of the vague parts of the definitions. There is a natural Poisson structure on $\mathcal{P}$ that is relevant for the Euler equation, see Section A5, but here we only need some of its Casimir functionals (to be specified now).

(2) The *Casimir functionals*, which play the role of the constraints g_j in the finite-dimensional example. These are linear combinations of elementary functionals of the form

$$h(a,\omega) = |\{\omega > a\}| = \int_{\mathbb{R}^2} \chi(\omega(x) - a)\,\mathrm{d}x, \quad a > 0, \tag{1-10}$$

where $\chi = \mathbf{1}_{(0,\infty)}$ is the indicator function of $(0,\infty)$. Here and in what follows, we denote by $|S|$ the Lebesgue measure of any (Borel) set $S \subset \mathbb{R}^2$. Due to our assumptions on the vorticities in $\mathcal{P}$, the functions $a \mapsto h(a,\omega)$ are finite and nonincreasing on $(0,\infty)$. In general, they do not have to be continuous in a but they will have this property in the examples considered later. Similarly, the functionals $\omega \mapsto h(a,\omega)$ may in general not be differentiable in every direction, but they will be in our examples. It is useful to single out the quantity

$$M_0(\omega) = \int_{\mathbb{R}^2} \omega(x)\,\mathrm{d}x = \int_0^\infty h(a,\omega)\,\mathrm{d}a, \tag{1-11}$$

which will be referred to as the "mass" of the vorticity distribution $\omega \in \mathcal{P}$.

(3) The *orbits* defined for any $\bar\omega \in \mathcal{P}$ by

$$\mathcal{O}_{\bar\omega} = \{\omega \in \mathcal{P} : h(a,\omega) = h(a,\bar\omega) \text{ for all } a \in (0,\infty)\}. \tag{1-12}$$

These subsets of the phase space are the analogues of the manifolds X_c defined by the constraints and can be considered as a measure-theoretical replacement for the symplectic leaves

$$\mathcal{O}_{\bar\omega}^{\mathrm{SDiff}} = \{\omega \in \mathcal{P} : \omega = \bar\omega \circ \phi \text{ for some } \phi \in \mathrm{SDiff}\} \subset \mathcal{O}_{\bar\omega},$$

where SDiff denotes the group of area-preserving diffeomorphisms in $\mathbb{R}^2$. In contrast to $\mathcal{O}_{\bar\omega}^{\mathrm{SDiff}}$, the orbit $\mathcal{O}_{\bar\omega}$ does not carry any topological information about $\bar\omega$, since $\omega \in \mathcal{O}_{\bar\omega}$ as soon as ω is a measure-preserving rearrangement of $\bar\omega$.

(4) The *Hamiltonian* (or energy functional) $E : \mathcal{P} \to \mathbb{R}$, given by

$$E(\omega) = -\frac{1}{2}\int_{\mathbb{R}^2} \psi(x)\omega(x)\,\mathrm{d}x = -\frac{1}{4\pi}\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \log|x-y|\,\omega(x)\,\omega(y)\,\mathrm{d}x\,\mathrm{d}y, \tag{1-13}$$

where $\psi = \Delta^{-1}\omega$ is the stream function defined by

$$\psi(x) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \log|x-y|\,\omega(y)\,\mathrm{d}y, \quad x \in \mathbb{R}. \tag{1-14}$$

This is an analogue of the function f in the finite-dimensional example. Note that the usual kinetic energy defined by $\frac{1}{2}\int_{\mathbb{R}^2} |u|^2\,\mathrm{d}x$, where $u = \nabla^\perp\psi$, is infinite for $\omega \in \mathcal{P}$. However, both definitions of the energy coincide when $\int_{\mathbb{R}^2} \omega\,\mathrm{d}x = 0$, which is the case for instance if ω is the difference of two vorticities in $\mathcal{P}$

with the same mass. It is also worth observing that the functional E is not invariant under the scaling transformation $\omega(x) \mapsto \omega^{(\lambda)}(x) := \lambda^2\omega(\lambda x)$ when $M_0 = \int_{\mathbb{R}^2} \omega \, \mathrm{d}x \neq 0$. In fact, one can easily check that

$$E(\omega^{(\lambda)}) = E(\omega) + \frac{M_0^2}{4\pi} \log \lambda \quad \text{for all } \lambda > 0.$$

(5) The *conserved quantities* induced by Euclidean symmetries. These are the first-order moments M_1, M_2 and the symmetric second-order moment I defined by

$$M_j(\omega) = \int_{\mathbb{R}^2} x_j \omega(x) \, \mathrm{d}x, \quad j = 1, 2, \qquad I(\omega) = \int_{\mathbb{R}^2} |x|^2 \omega(x) \, \mathrm{d}x. \tag{1-15}$$

Note that M_1, M_2 are associated to the translational symmetry, via Noether's theorem, and I to the rotational symmetry.

With these definitions, the Euler equation can be written in the form $\partial_t \omega = \{E(\omega), \omega\}$, where $\{\cdot, \cdot\}$ denotes the Poisson bracket on $\mathcal{P}$; see Section A5. Any steady state $\bar{\omega} \in \mathcal{P}$ is a critical point of the Hamiltonian E on the orbit $\mathcal{O}_{\bar{\omega}}$. Stability can be inferred when the restriction of the energy E to $\mathcal{O}_{\bar{\omega}}$ has a strict local extremum at $\bar{\omega}$. In what follows, we focus on the maximizers of the energy, which correspond to radially symmetric vortices.

1C. ***The constrained maximization of the energy in $\mathcal{P}$.*** Under our assumptions, it is easy to determine the maximizers of the Hamiltonian E under the constraints given by the functions $h(a, \omega)$ for $a \in (0, \infty)$. Indeed, for any $\omega \in \mathcal{P}$, the orbit $\mathcal{O}_\omega$ contains a unique element ω^* that is radially symmetric and nonincreasing in the radial direction; this is the *symmetric decreasing rearrangement* of ω [Lieb and Loss 1997]. The Riesz's rearrangement inequality then shows that $E(\omega) \leq E(\omega^*)$ for all $\omega \in \mathcal{O}_{\omega^*}$, with equality if and only if ω is a translate of ω^*; see [Carlen and Loss 1992, Lemma 2]. Of course ω^* is a stationary solution of the Euler equation, which represents a radially symmetric vortex with nonincreasing vorticity profile. Our main focus here will be on the analogue of the quadratic form (1-5) for the steady state $\bar{\omega} = \omega^*$.

First, the analogue of the Lagrange function (1-7) is

$$E(\omega) - \int_0^\infty \Lambda(a) h(a, \omega) \, \mathrm{d}a = E(\omega) - \int_0^\infty \Lambda(a) \left(\int_{\mathbb{R}^2} \chi(\omega(x) - a) \, \mathrm{d}x \right) \mathrm{d}a,$$

where the quantities $\Lambda(a)$ for $a \in (0, \infty)$ can be thought of as the Lagrange multipliers. The role of the discrete index j in (1-7) is now played by the continuous parameter $a > 0$. Defining[4]

$$\Phi(s) = -\int_0^\infty \Lambda(a) \chi(s - a) \, \mathrm{d}a = -\int_0^s \Lambda(a) \, \mathrm{d}a, \quad s > 0, \tag{1-16}$$

we see that the Lagrange function can also be expressed as

$$F(\omega) = E(\omega) + \int_{\mathbb{R}^2} \Phi(\omega(x)) \, \mathrm{d}x, \quad \omega \in \mathcal{P}. \tag{1-17}$$

[4]The reason for the minus sign in (1-16) will become clear later.

This quantity will be referred to later as the "free energy" of the vorticity distribution ω, a terminology that will be discussed in Section 1D below.

Next, the analogue of the stationarity condition (1-4) at $\bar\omega = \omega^*$ is $F'(\bar\omega) = 0$, where the linear form $\eta \mapsto F'(\bar\omega)\eta$ is defined for all $\eta \in T_{\bar\omega}\mathcal{P}$ by

$$F'(\bar\omega)\eta = \int_{\mathbb{R}^2} \big(-\bar\psi(x) + \Phi'(\bar\omega(x))\big)\eta(x)\,\mathrm{d}x, \qquad \bar\psi(x) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \log|x-y|\,\bar\omega(y)\,\mathrm{d}y.$$

Stationarity is thus equivalent to the relation $\bar\psi(x) = \Phi'(\bar\omega(x))$ for all $x \in \mathbb{R}^2$. Finally the analogue of (1-5) is the quadratic form $\eta \mapsto F''(\bar\omega)[\eta,\eta]$, where

$$F''(\bar\omega)[\eta,\eta] = \int_{\mathbb{R}^2} (-\varphi\eta + \Phi''(\bar\omega)\eta^2)\,\mathrm{d}x, \qquad \varphi(x) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \log|x-y|\,\eta(y)\,\mathrm{d}y.$$

Using the relation $\nabla\bar\psi(x) = \Phi''(\bar\omega(x))\nabla\bar\omega(x)$, the second variation can be rewritten in the form

$$F''(\bar\omega)[\eta,\eta] = \int_{\mathbb{R}^2}\left(-\varphi\eta + \frac{\nabla\bar\psi}{\nabla\bar\omega}\eta^2\right)\mathrm{d}x = 2E(\eta) + \int_{\mathbb{R}^2}\frac{\nabla\bar\psi}{\nabla\bar\omega}\,\eta^2\,\mathrm{d}x, \tag{1-18}$$

which is well known from Arnold's work. Note that the ratio $\nabla\bar\psi/\nabla\bar\omega$ is meaningful only when the vector $\nabla\bar\omega(x)$ is nonzero and colinear with $\nabla\bar\psi(x)$ for almost all $x \in \mathbb{R}^2$. This condition is obviously satisfied for all radially symmetric vortices with strictly decreasing vorticity profile.

1D. ***Overview of our results.*** We are now able to describe more precisely the results of this paper. We consider a general family of radially symmetric vortices $\bar\omega \in \mathcal{P}$ with vorticity profile satisfying Hypotheses 2.1 below. Typical examples are the "algebraic vortex" $\bar\omega(x) = (1+|x|^2)^{-\kappa}$, where $\kappa > 1$ is a parameter, and the Oseen vortex for which $\bar\omega(x) = e^{-|x|^2/4}$.

1D1. *Arnold's quadratic forms with and without constraints.* In Section 2, we study in detail the quadratic form (1-18) associated with the second variation of the Lagrange function (1-17) at the steady state $\bar\omega \in \mathcal{P}$, paying some attention to the functional-analytic questions. First of all, while we know from the constrained maximization result that the restriction of that form to the tangent space $T_{\bar\omega}\mathcal{O}_{\bar\omega}$ is negative, it is not clear if this restriction is strictly negative definite, and if so in which function space. Our first main result is Theorem 2.5, where we show that, if two neutral directions corresponding to translational symmetry are disregarded, the restriction to $T_{\bar\omega}\mathcal{O}_{\bar\omega}$ of the quadratic form (1-18) is indeed strictly negative in an appropriate weighted L^2 space. The proof ultimately relies on a variant of the Krein–Rutman theorem.

We next investigate the index of the quadratic form (1-18) on a much larger subspace, corresponding to perturbations $\eta \in T_{\bar\omega}\mathcal{P}$ satisfying $\int_{\mathbb{R}^2}\eta(x)\,\mathrm{d}x = 0$. In other words, we relax all constraints given by the Casimir functions (1-10), except for the mass M_0 defined in (1-11), which is still supposed to be constant. A priori there is no reason why the form (1-18) should be negative definite in this larger sense, and indeed Theorem 2.8 shows that this is not always the case. More precisely, we show that negativity holds in the large sense if and only if the optimal constant in some weighted Hardy inequality (where the weight function depends on the vorticity profile $\bar\omega$) is smaller than 1. While that condition is not easy to check in general, we deduce from Corollary 2.11 that it is fulfilled at least for the Oseen vortex, as well as for the algebraic vortex $\bar\omega(x) = (1+|x|^2)^{-\kappa}$ if $\kappa \geq 2$.

Although the mass constraint is rather natural, one may wonder if, for some vorticity profiles, the quadratic form (1-18) can be negative definite for all perturbations $\eta \in T_{\bar{\omega}}\mathcal{P}$; this question is briefly discussed in Section 2C. Finally, in Section 2D, we give a fairly explicit expression of the energy $E(\bar{\omega})$ in terms of the constraints $h(a, \bar{\omega})$ for all $a > 0$; see Proposition 2.18. One obtains in this way an infinite-dimensional analogue of the quantity $M(c_1, \dots, c_n)$ defined in (1-8). Among other things, we justify our claim that the index of the quadratic form (1-5) is related to the concavity of the function (1-8) (which is a well-known fact), and we discuss a similar link in the infinite-dimensional case.

As an aside, we mention here that the stability of radially symmetric vortices for the two-dimensional Euler equations can also be studied using other conserved quantities, such as the second-order symmetric moment I defined in (1-15); see, e.g., [Marchioro and Pulvirenti 1994, Chapter 3].

1D2. *The global maximizers of the free energy.* Let $\bar{\psi}$ be the stream function associated with the radially symmetric vortex $\bar{\omega}$. We have seen that the analogue of the Lagrange function (1-7) is given by the "free energy" (1-17), where the function Φ is defined, up to an additive constant, by the relation $\bar{\psi}(x) = \Phi'(\bar{\omega}(x))$. The appellation "free energy" is partially justified by a (loose) analogy of formula (1-17) with the classical thermodynamical expression for the free energy

$$F = U - TS. \tag{1-19}$$

Here U is the internal energy (of a suitable system), T is the temperature, and S is the entropy. In (1-17), the energy E is analogous to U, the integral $\int_{\mathbb{R}^2} \Phi(\omega(x))\,\mathrm{d}x$ is analogous to S, and one can argue that it is reasonable to take $T = -1$. Of course, T has nothing to do with the real temperature of the fluid, but should roughly be thought of as the statistical mechanics temperature of our system in the sense of [Onsager 1949]. We have not attempted to make this connection rigorous, which would take us in a different direction.

In Section 3, we consider vortices $\bar{\omega}$ which are *global maximizers* of the free energy $F(\omega)$ for all $\omega \in \mathcal{P}$ satisfying $\int_{\mathbb{R}^2} \omega\,\mathrm{d}x = \int_{\mathbb{R}^2} \bar{\omega}\,\mathrm{d}x$. Such equilibria can be expected to have strong stability properties, and may be useful for other purposes too. Using a direct approach, in the sense of the calculus of variations, we prove the existence of global maximizers under fairly general assumptions on the function Φ; see Theorem 3.4. However, we do not have any efficient method to determine if a given vortex $\bar{\omega}$ is a global maximizer or not. A necessary condition is of course that the quadratic form (1-18) be negative on perturbations η with zero mean, see Theorem 2.8, but there is no reason to believe that this is sufficient. Numerical evidence indicates that the Oseen vortex is a global maximizer, and so are the algebraic vortices $\bar{\omega}(x) = (1 + |x|^2)^{-\kappa}$ for $\kappa \geq 2$. In the particular case $\kappa = 2$, maximality can be deduced from the logarithmic Hardy–Littlewood–Sobolev inequality

$$\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \log\frac{1}{|x-y|}\omega(x)\omega(y)\,\mathrm{d}x\,\mathrm{d}y \leq \frac{1}{2}\int_{\mathbb{R}^2} \omega(x)\log(\omega(x)) + \frac{1+\log(\pi)}{2}, \tag{1-20}$$

which holds for all $\omega \in \mathcal{P}$ with $M_0(\omega) = 1$; see [Carlen and Loss 1992]. We mention that (1-20) is related to Onofri's sharp version [1982] of the Moser–Trudinger inequality.

1D3. *The effect of viscosity: application to Oseen vortices.* In Section 4, we consider the stability of the Gaussian vortex under the evolution defined by the Navier–Stokes equation $\partial_t\omega + u \cdot \nabla\omega = \nu\Delta\omega$, where

$\nu > 0$ is the viscosity parameter. More precisely, we show that the quadratic form (1-18) can be used to give an alternative proof of the local stability results established in [Gallay and Wayne 2005]. We believe that a proof relying on the second variation of the energy is of some interest, because the analogue of the form (1-18) can be defined for more complicated vortex structures as well, whereas the simpler approach in [Gallay and Wayne 2005] may be more difficult to adapt.

The addition of the viscous term results in important new issues: the radial vortices are no longer steady states and the orbits (1-12) are no longer invariant under the evolution, so that much of the geometric picture underlying the Euler equation is destroyed. The first problem is settled by introducing self-similar variables and restricting ourselves to Oseen's vortex, which is a stationary solution of the Navier–Stokes equation in these new coordinates. Thanks to Theorem 2.8 and Corollary 2.11, the quadratic form (1-18) is positive definite for all perturbations with zero mean. This form is invariant under the evolution defined by the linearized Euler equation at the vortex, but not under the Navier–Stokes evolution due to the viscous term and the nonlinearity. The effect of viscosity is measured by a second quadratic form, which happens to have a favorable sign; see Theorem 4.2. We do not know if this is just a lucky coincidence, or if there are deeper reasons behind that. In any event, this nice structure allows us to recover the local stability result of [Gallay and Wayne 2005], except for a slight difference in the choice of the function space; see Theorem 4.5. Again, we emphasize that the functional setting used in that work relies in an essential way on the radial symmetry of Oseen's vortex, through the existence of conserved quantities such as the moment I in (1-15), whereas our new approach can, at least in principle, be adapted to more general situations, where other methods do not work.

2. The second variation of the energy

In this section we study the coercivity, on various subspaces, of the quadratic form (1-18) which represents the second variation of the free energy (1-17) at a radially symmetric vortex $\bar\omega \in \mathscr{P}$. We assume that $\bar\omega(x) = \omega_*(|x|)$ for all $x \in \mathbb{R}^2$ and that the vorticity profile $\omega_* : [0, +\infty) \to \mathbb{R}$ is a C^2 function with the following properties:

Hypotheses 2.1. *The vorticity profile $\omega_* \in C^2([0, +\infty))$ satisfies*

(1) $\omega_*(0) > 0$, $\omega_*'(0) = 0$, *and* $\omega_*''(0) < 0$,

(2) $\omega_*'(r) < 0$ *for all* $r > 0$, *and* $\omega_*(r) \to 0$ *as* $r \to +\infty$,

(3) *there exist* $C > 0$ *and* $\beta > 2$ *such that* $|\omega_*'(r)| \le C(1+r)^{-\beta-1}$ *for all* $r > 0$.

It follows in particular from (2) and (3) that $\omega_*(r) = -\int_r^\infty \omega_*'(s)\,\mathrm{d}s$, so that

$$0 < \omega_*(r) \le \frac{C}{(1+r)^\beta} \quad \text{for all } r > 0 \qquad \text{and} \qquad 0 < \int_0^\infty r\omega_*(r)\,\mathrm{d}r < \infty. \tag{2-1}$$

Let $\bar\psi$ be the stream function associated with $\bar\omega$ as in (1-14). We have $\bar\psi(x) = \psi_*(|x|)$, where the stream profile $\psi_* : [0, +\infty) \to \mathbb{R}$ satisfies

$$\psi_*''(r) + \frac{1}{r}\psi_*'(r) = \omega_*(r); \qquad \text{hence} \quad \psi_*'(r) = \frac{1}{r}\int_0^r s\omega_*(s)\,\mathrm{d}s \quad \text{for all } r > 0. \tag{2-2}$$

We introduce the weight function $A : [0, +\infty) \to \mathbb{R}$ defined by $A(0) = -\omega_*(0)/(2\omega_*''(0))$ and

$$A(r) = -\frac{\psi_*'(r)}{\omega_*'(r)} = -\frac{1}{r\omega_*'(r)} \int_0^r s\omega_*(s)\,\mathrm{d}s, \quad r > 0. \tag{2-3}$$

Hypotheses 2.1 ensure that $A \in C^0([0, +\infty)) \cap C^1((0, +\infty))$. Moreover, there exists a constant $C > 0$ such that $A(r) \geq C(1 + r)^\beta$ for all $r \geq 0$.

Let $\mathcal{A} : \mathbb{R}^2 \to (0, \infty)$ be the radially symmetric extension of A to $\mathbb{R}^2$, namely $\mathcal{A}(x) = A(|x|)$ for all $x \in \mathbb{R}^2$. We introduce the weighted L^2 space X defined by

$$X = \left\{\omega \in L^2(\mathbb{R}^2) : \|\omega\|_X^2 := \int_{\mathbb{R}^2} \mathcal{A}(x)|\omega(x)|^2\,\mathrm{d}x < \infty\right\}, \tag{2-4}$$

so that $\omega \in X$ if and only if $\mathcal{A}^{1/2}\omega \in L^2(\mathbb{R}^2)$. Our assumptions ensure that $\mathcal{A}^{-1} \in L^1(\mathbb{R}^2)$, and using Hölder's inequality we easily deduce that $X \hookrightarrow L^1(\mathbb{R}^2)$. We also consider the closed subspaces $X_1 \subset X_0 \subset X$ defined by

$$\begin{aligned} X_0 &= \left\{\omega \in X : \int_{\mathbb{R}^2} \omega(x)\,\mathrm{d}x = 0\right\}, \\ X_1 &= \left\{\omega \in X_0 : \int_{\mathbb{R}^2} \frac{x_j}{|x|}\omega(x)\,\mathrm{d}x = 0 \text{ for } j = 1, 2\right\}. \end{aligned} \tag{2-5}$$

We observe that, for any $\omega \in X$, the energy $E(\omega)$ introduced in (1-13) is well-defined. This a consequence of the following classical estimate, whose proof is reproduced in Section A1 for the reader's convenience.

Proposition 2.2. *Assume that $\omega \in L^1(\mathbb{R}^2)$ satisfies*

$$\int_{\mathbb{R}^2} |\omega(x)|\,\log(1 + |x|)\,\mathrm{d}x < \infty \quad \textit{and} \quad \int_{\mathbb{R}^2} |\omega(x)|\,\log\bigl(1 + |\omega(x)|\bigr)\,\mathrm{d}x < \infty. \tag{2-6}$$

Then the last member in (1-13) is well-defined, and the energy $E(\omega)$ satisfies the bound

$$|E(\omega)| \leq C\|\omega\|_{L^1}\left(\int_{\mathbb{R}^2} |\omega(x)|\,\log(2 + |x|)\,\mathrm{d}x + \int_{\mathbb{R}^2} |\omega(x)|\,\log_+ \frac{|\omega(x)|}{\|\omega\|_{L^1}}\,\mathrm{d}x\right), \tag{2-7}$$

where $\log_+(a) = \max(\log(a), 0)$. *If, moreover,* $\int_{\mathbb{R}^2} \omega(x)\,\mathrm{d}x = 0$, *then* $E(\omega) = \frac{1}{2}\int_{\mathbb{R}^2} |u|^2\,\mathrm{d}x$, *where*

$$u(x) = \nabla^\perp\psi(x) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \frac{(x - y)^\perp}{|x - y|^2}\,\omega(y)\,\mathrm{d}y, \quad x \in \mathbb{R}^2. \tag{2-8}$$

Since any $\omega \in X$ obviously satisfies (2-6), we can consider the quadratic form J on X defined by $J(\omega) = \frac{1}{2}\|\omega\|_X^2 - E(\omega)$, or explicitly

$$J(\omega) = \frac{1}{2}\int_{\mathbb{R}^2} \mathcal{A}(x)\omega(x)^2\,\mathrm{d}x + \frac{1}{4\pi}\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \log|x - y|\,\omega(x)\,\omega(y)\,\mathrm{d}x\,\mathrm{d}y, \quad \omega \in X. \tag{2-9}$$

In the particular case where $\omega \in X_0$, namely when ω has zero average over $\mathbb{R}^2$, Proposition 2.2 gives the alternative expression

$$J(\omega) = \frac{1}{2}\int_{\mathbb{R}^2} (\mathcal{A}(x)\omega(x)^2 - |u(x)|^2)\,\mathrm{d}x, \quad \omega \in X_0, \tag{2-10}$$

where u is the velocity field associated with ω via the Biot–Savart formula (2-8). In view of (1-18) and (2-3), we have $J = -\frac{1}{2}F''(\bar{\omega})$, where $F''(\bar{\omega})$ is the second variation of the free energy (1-17) at the equilibrium $\bar{\omega}$. It is clear that X is the largest function space on which this second variation is well-defined.

Our main goal in this section is to study the positivity and coercivity properties of the quadratic form J on the spaces X, X_0, and X_1 defined in (2-4), (2-5). To formulate our results, it is useful to take the decomposition $X = X_{\mathrm{rs}} \oplus X_{\mathrm{rs}}^{\perp}$, where

$$X_{\mathrm{rs}} = \{\omega \in X : \omega \text{ is radially symmetric}\}, \tag{2-11}$$

and $X_{\mathrm{rs}}^{\perp}$ is the orthogonal complement of X_{rs} in the Hilbert space X. Referring to the geometric picture of Section 1B, we consider $X_{\mathrm{rs}}^{\perp}$ as the *tangent space to the orbit* $\mathcal{O}_{\bar{\omega}}$ at $\bar{\omega}$. This interpretation can be formally justified as follows: if $\bar{\omega} \in X$ is smooth, the tangent space $T_{\bar{\omega}}\mathcal{O}_{\bar{\omega}}$ is spanned by vorticities of the form $v \cdot \nabla\bar{\omega}$, where v is a (smooth and localized) divergence-free vector field, and using polar coordinates as in Section 2A below one verifies that such vorticities are indeed orthogonal in X to all radially symmetric functions. A contrario, since there is a one-to-one correspondence in $\mathcal{P}$ between orbits and symmetric decreasing rearrangements, it is clear that any radially symmetric perturbation of the equilibrium $\bar{\omega}$ is transverse to the orbit $\mathcal{O}_{\bar{\omega}}$.

It is easy to verify that $J(\omega_1+\omega_2) = J(\omega_1)+J(\omega_2)$ when $\omega_1 \in X_{\mathrm{rs}}$ and $\omega_2 \in X_{\mathrm{rs}}^{\perp}$, so that the restrictions of J to X_{rs} and $X_{\mathrm{rs}}^{\perp}$ can be studied separately. We first consider the tangent space $X_{\mathrm{rs}}^{\perp}$ in Section 2A, and postpone the study of radially symmetric perturbations (with zero or nonzero mass) to Sections 2B and 2C.

Remark 2.3. Differentiating the first equality in (2-2), we see that the function $\phi = \psi_*'$ satisfies

$$(L_0\phi)(r) := -\phi''(r) - \frac{1}{r}\phi'(r) + \frac{1}{r^2}\phi(r) = \frac{1}{A(r)}\phi(r), \quad r > 0, \tag{2-12}$$

where $A(r) \geq C(1+r)^{\beta}$. Since $\phi > 0$, Sturm–Liouville theory asserts that $\mu = 1$ is the lowest eigenvalue of the (generalized) eigenvalue problem $L_0\phi = \mu A^{-1}\phi$ on $\mathbb{R}_+$, with boundary conditions $\phi(0) = \phi(+\infty) = 0$; see [Coddington and Levinson 1955; Hartman 1964]. This observation will be used later.

Remark 2.4. Hypotheses 2.1 are sufficient for our results to hold, but can be relaxed in several ways. In particular, we can consider vortices that are not smooth at the origin, but the assumption that $\omega_*'(r) < 0$ for all $r > 0$ seems essential. This excludes vortices with compact support from our considerations, but as our motivation comes from applications to the Navier–Stokes equations, Hypotheses 2.1 are good enough for our purposes here. Of course, extensions of the theory that would include compactly supported vortices might be relevant in other situations and can probably be constructed, although they may require additional work.

2A. *Positivity of the quadratic form J on $X_{\mathrm{rs}}^{\perp}$.*

Theorem 2.5. *Under Hypotheses 2.1, the quadratic form J defined by (2-10) is nonnegative on the space $X_{\mathrm{rs}}^{\perp} \subset X_0$. Moreover, there exists a constant $\gamma > 0$ such that*

$$J(\omega) \geq \frac{\gamma}{2}\int_{\mathbb{R}^2} \mathcal{A}(x)\omega(x)^2\,\mathrm{d}x \quad \textit{for all } \omega \in X_{\mathrm{rs}}^{\perp} \cap X_1. \tag{2-13}$$

Proof. We introduce polar coordinates (r, θ) in $\mathbb{R}^2$, and given any $\omega \in X_{\mathrm{rs}}^{\perp}$ we use the Fourier decomposition

$$\omega(r\cos(\theta), r\sin(\theta)) = \sum_{k\neq 0} \omega_k(r) e^{ik\theta}, \quad r > 0,\ \theta \in \mathbb{R}/(2\pi\mathbb{Z}), \tag{2-14}$$

where the sum runs over all nonzero integers $k \in \mathbb{Z} \setminus \{0\}$. By Parseval's relation we have

$$\begin{aligned} \int_{\mathbb{R}^2} \mathcal{A}(x)\omega(x)^2\,\mathrm{d}x &= 2\pi \sum_{k\neq 0} \int_0^\infty A(r)\,|\omega_k(r)|^2\, r\,\mathrm{d}r, \\ \int_{\mathbb{R}^2} |u(x)|^2\,\mathrm{d}x = \int_{\mathbb{R}^2} (-\Delta^{-1}\omega)(x)\,\omega(x)\,\mathrm{d}x &= 2\pi \sum_{k\neq 0} \int_0^\infty B_k[\omega_k](r)\,\bar{\omega}_k(r)\, r\,\mathrm{d}r, \end{aligned} \tag{2-15}$$

where B_k is the integral operator on the half-line $\mathbb{R}_+$ defined by the formula

$$(B_k[f])(r) = \frac{1}{2|k|} \int_0^\infty \min\Big(\frac{r}{s}, \frac{s}{r}\Big)^{|k|} f(s)\, s\,\mathrm{d}s, \quad r > 0. \tag{2-16}$$

Note that $g = B_k[f]$ is the unique solution of the ODE

$$-g''(r) - \frac{1}{r} g'(r) + \frac{k^2}{r^2} g(r) = f(r), \quad r > 0, \tag{2-17}$$

which is regular at the origin and converges to zero at infinity.

In view of (2-15), the proof of Theorem 2.5 reduces to the study of the one-dimensional inequality

$$\int_0^\infty (B_k[f])(r)\,\bar{f}(r)\, r\,\mathrm{d}r \le C_k \int_0^\infty A(r)|f(r)|^2\, r\,\mathrm{d}r, \tag{2-18}$$

which depends on the angular Fourier parameter $k \in \mathbb{Z} \setminus \{0\}$. More precisely, the quadratic form J is nonnegative on $X_{\mathrm{rs}}^{\perp}$ if and only if, for all $k \neq 0$, inequality (2-18) holds with some constant $C_k \le 1$. In addition, we have the lower bound (2-13) on the subspace $X_{\mathrm{rs}}^{\perp} \cap X_1$ if and only if inequality (2-18) holds with a better constant $C_k \le 1-\gamma$ for all $k \neq 0$, assuming when $|k| = 1$ that f satisfies the additional condition

$$\int_0^\infty f(r)\, r\,\mathrm{d}r = 0. \tag{2-19}$$

It remains to establish inequality (2-18) for all $k \in \mathbb{Z} \setminus \{0\}$. We obviously have the pointwise bound $|(B_k[f])(r)| \le (B_k[|f|])(r)$, so that we can restrict ourselves to nonnegative functions f. Moreover the operator B_k preserves positivity, and an inspection of the formula (2-16) reveals that $0 \le B_k[f] \le |k|^{-1} B_1[f]$ if $f \ge 0$. As a consequence, to show that J is nonnegative on $X_{\mathrm{rs}}^{\perp}$, it is sufficient to prove inequality (2-18) in the particular case where $|k| = 1$ and $f \ge 0$. Setting $h = A^{1/2} f$, we write that inequality in the equivalent form

$$\int_0^\infty (\widetilde{B}_1[h])(r)\, h(r)\, r\,\mathrm{d}r \le C_1 \int_0^\infty h(r)^2\, r\,\mathrm{d}r, \tag{2-20}$$

where $\widetilde{B}_1[h] = A^{-1/2} B_1[A^{-1/2} h]$. The following assertions play a crucial role in our argument:

<u>Claim 1</u>: The operator $\widetilde{B}_1$ is *self-adjoint and compact* in the (real) space $Y = L^2(\mathbb{R}_+, r\,\mathrm{d}r)$.

Indeed, take $h \in Y$ with $\|h\|_Y \le 1$, and define $f = A^{-1/2}h$, $g = B_1[f] = A^{1/2}\widetilde{B}_1[h]$. Applying (2-16) with $|k| = 1$, we see that

$$g(r) = \frac{1}{2r}\int_0^r A(s)^{-1/2}\, h(s)\, s^2\, \mathrm{d}s + \frac{r}{2}\int_r^\infty A(s)^{-1/2}\, h(s)\, \mathrm{d}s, \quad r > 0,$$

and using Hölder's inequality we deduce

$$|g(r)| \le \left\{ \frac{1}{2r}\left(\int_0^r A(s)^{-1}\, s^3\, \mathrm{d}s\right)^{1/2} + \frac{r}{2}\left(\int_r^\infty A(s)^{-1}\, s^{-1}\, \mathrm{d}s\right)^{1/2} \right\} \|h\|_Y. \tag{2-21}$$

As $A(r) \ge C(1+r)^\beta$ with $\beta > 2$, the right-hand side of (2-21) is uniformly bounded, so that $\|g\|_{L^\infty} \le C$ for some universal constant C. It also follows from (2-21) that $g(r) \to 0$ as $r \to 0$ and $r \to +\infty$. On the other hand, since g satisfies the ODE (2-17) with $k = 1$ and $f = A^{-1/2}h$, a standard energy estimate yields the bound

$$\int_0^\infty \left(g'(r)^2 + \frac{g(r)^2}{r^2}\right) r\, \mathrm{d}r = \int_0^\infty g(r)\, A(r)^{-1/2}\, h(r)\, r\, \mathrm{d}r \le \|g\|_{L^\infty} \|A^{-1/2}\|_Y \|h\|_Y \le C. \tag{2-22}$$

In view of (2-21) and (2-22), the Fréchet–Kolmogorov theorem [Reed and Simon 1978, Theorem XIII.66] implies that the function $\widetilde{B}_1[h] = A^{-1/2}g$ lies in a compact set of Y, so that the operator $\widetilde{B}_1$ is compact. To prove that $\widetilde{B}_1$ is self-adjoint, we take $h_1, h_2 \in Y$ and observe that

$$\int_0^\infty (\widetilde{B}_1[h_1])(r)\, h_2(r)\, r\, \mathrm{d}r = \int_0^\infty \left(g_1'(r) g_2'(r) + \frac{g_1(r) g_2(r)}{r^2}\right) r\, \mathrm{d}r,$$

where $g_j = B_1[A^{-1/2}h_j]$ for $j = 1, 2$. This expression is clearly a symmetric function of (h_1, h_2).

<u>Claim 2</u>: The *spectral radius* of $\widetilde{B}_1$ is equal to 1, and $\lambda = 1$ is a *simple eigenvalue* of $\widetilde{B}_1$.

To see that, we first observe that $\lambda = 1$ is an eigenvalue of $\widetilde{B}_1$ with a positive eigenfunction. Indeed, using (2-2), it is straightforward to verify that the function $g = \psi_*'$ satisfies the ODE (2-17) with $k = 1$ and $f = -\omega_*'$. This shows that $B_1[-\omega_*'] = \psi_*'$; hence defining $h = A^{-1/2}\psi_*' = -A^{1/2}\omega_*'$ we conclude that $\widetilde{B}_1[h] = h$. On the other hand, assume that $\lambda > 0$ is an eigenvalue of $\widetilde{B}_1$, with eigenfunction $h \in Y$. Defining $f = A^{-1/2}h$, we see that $B_1[f] = \lambda A f$, so that the function $g = B_1[f]$ satisfies the generalized eigenvalue problem

$$-g''(r) - \frac{1}{r}g'(r) + \frac{1}{r^2}g(r) = \mu\frac{g(r)}{A(r)}, \quad r > 0, \tag{2-23}$$

with the boundary conditions $g(0) = g(+\infty) = 0$, where $\mu = 1/\lambda$, We already observed that $\mu = 1$ is the lowest eigenvalue of (2-23); see Remark 2.3. It follows that $\lambda = 1$ is the largest eigenvalue of the integral operator $\widetilde{B}_1$, whose spectral radius is therefore equal to 1. The argument above also shows that all positive eigenvalues of $\widetilde{B}_1$ are simple, because (2-23) is a second-order differential equation.

It is now a simple task to conclude the proof of Theorem 2.5. Claims 1 and 2 imply the validity of inequality (2-20) with $C_1 = 1$. We deduce that (2-18) holds for $|k| = 1$ with $C_k = 1$, and (since $B_k \le |k|^{-1}B_1$) for $|k| \ge 2$ with $C_k \le 1/|k|$. This shows that the quadratic form J is nonnegative on $X_{\mathrm{rs}}^\perp$. On the other hand, if we assume that $\omega \in X_{\mathrm{rs}}^\perp \cap X_1$, the function $f = \omega_{\pm 1}$ satisfies condition (2-19), which

means that $h = A^{1/2} f$ is orthogonal in Y to the one-dimensional subspace Y_0 spanned by the positive function $\chi = A^{-1/2}$. It is clear that $Y_0^{\perp}$ does not contain any positive function, and in particular does not include the principal eigenfunction $h_0 = -A^{1/2}\omega_*'$ of the operator $\widetilde{B}_1$. So, applying Lemma 4.7 and Remark 4.8 below, we deduce that $\mathbf{1} - \widetilde{B}_1 > 0$ on $Y_0^{\perp}$, which means that inequality (2-20) holds on $Y_0^{\perp}$ with some constant $C_1' < 1$. Taking into account the other values of k, for which $C_k \le 1/|k| \le \frac{1}{2}$, we conclude that estimate (2-13) holds with $\gamma = \min\left(\frac{1}{2}, 1 - C_1'\right)$. □

Remark 2.6. The Krein–Rutman theorem [Deimling 1985, Theorem 19.2] asserts that the spectral radius of the compact and positivity-preserving operator $\widetilde{B}_1$ is an eigenvalue with positive eigenfunction. However, since the cone of positive functions has empty interior in Y, we cannot apply Theorem 19.3 in [Deimling 1985] to conclude that $\widetilde{B}_1$ has a *unique* eigenvalue with positive eigenfunction, which is thus equal to the spectral radius. For this reason, we prefer invoking Sturm–Liouville theory to prove that 1 is the largest eigenvalue of $\widetilde{B}_1$.

Remark 2.7. If $\beta > 4$ in Hypotheses 2.1, the conclusion of Theorem 2.5 remains valid, with the same proof, if the subspace X_1 is replaced by

$$\mathcal{X}_1 = \left\{ \omega \in X_0 : \int_{\mathbb{R}^2} x_j\, \omega(x)\, \mathrm{d}x = 0 \text{ for } j = 1, 2 \right\}. \tag{2-24}$$

This possibility will be used in Section 4.

2B. *Positivity of the quadratic form J on $X_{\mathrm{rs}} \cap X_0$.* The quadratic form J is not necessarily positive when considered on the subspace $X_{\mathrm{rs}} \cap X_0$, which consists of radially symmetric functions with zero mean. This question is related to the optimal constant in the weighted Hardy inequality

$$\int_0^{\infty} f(r)^2 \frac{\mathrm{d}r}{r} \le C_H \int_0^{\infty} A(r) f'(r)^2 \frac{\mathrm{d}r}{r}, \tag{2-25}$$

where $f : [0, +\infty) \to \mathbb{R}$ is an absolutely continuous function with $f(0) = f(+\infty) = 0$. Weighted Hardy inequalities are extensively studied in the literature; see, e.g., [Mazya 2011, Section 1.3.2]. In particular, it is known that (2-25) holds for *some* constant $C_H > 0$ if and only if the positive function A satisfies

$$\limsup_{r \to 0} \left(\log \frac{1}{r}\right) \int_0^r \frac{s}{A(s)}\, \mathrm{d}s < \infty \quad \text{and} \quad \limsup_{r \to +\infty} \log(r) \int_r^{\infty} \frac{s}{A(s)}\, \mathrm{d}s < \infty. \tag{2-26}$$

Both conditions in (2-26) are fulfilled in our case, since $A(r) \ge C(1 + r)^{\beta}$ for some $\beta > 2$.

Theorem 2.8. *Under Hypotheses 2.1, the quadratic form J defined by (2-10) is coercive on $X_{\mathrm{rs}} \cap X_0$ if and only if Hardy's inequality (2-25) holds for some $C_H < 1$. In that case we have*

$$J(\omega) \ge \frac{\gamma}{2} \int_{\mathbb{R}^2} \mathcal{A}(x) \omega(x)^2\, \mathrm{d}x \quad \textit{for all } \omega \in X_{\mathrm{rs}} \cap X_0, \tag{2-27}$$

where $\gamma = 1 - C_H$.

Proof. Given $\omega \in X_{\mathrm{rs}} \cap X_0$, we write $\omega(x) = \omega_0(|x|)$ and we consider the stream function ψ_0 defined (up to an irrelevant additive constant) by

$$\psi_0'(r) = \frac{1}{r} \int_0^r s\omega_0(s)\, \mathrm{d}s = -\frac{1}{r} \int_r^{\infty} s\omega_0(s)\, \mathrm{d}s, \quad r > 0.$$

Defining $f(r)=r\psi_0'(r)$, we see that f is absolutely continuous on $\mathbb{R}_+$ with $f(0)=f(+\infty)=0$. Moreover we have $\omega_0(r)=f'(r)/r$ and $u_0(r):=\psi_0'(r)=f(r)/r$ by construction. Finally the assumption that $\omega_0\in X_{\rm rs}\cap X_0$ ensures that $A^{1/2}\omega_0$ and u_0 belong to the space $Y=L^2(\mathbb{R}_+,r\,\mathrm{d}r)$. We thus have

$$J(\omega)=\pi\int_0^\infty\big(A(r)\omega_0(r)^2-u_0(r)^2\big)\,r\,\mathrm{d}r=\pi\int_0^\infty\big(A(r)f'(r)^2-f(r)^2\big)\,\frac{\mathrm{d}r}{r}, \tag{2-28}$$

and using (2-25) we conclude that (2-27) holds with $\gamma=1-C_H$. This proves that the quadratic form J is coercive on $X_{\rm rs}\cap X_0$ if $C_H<1$. Conversely, if (2-27) holds for some $\gamma>0$, it follows from (2-28) that inequality (2-25) is valid with $C_H=1-\gamma$. □

As is well known, the optimal constant in Hardy's inequality (2-25) is related to the lowest eigenvalue of a self-adjoint operator. A convenient way of seeing this is to apply the change of variables $r=e^x$, $h(x)=f(e^x)$, $B(x)=e^{-2x}A(e^x)$, which transforms (2-25) into the equivalent inequality

$$\int_{\mathbb{R}}h(x)^2\,\mathrm{d}x\le C_H\int_{\mathbb{R}}B(x)h'(x)^2\,\mathrm{d}x. \tag{2-29}$$

The integral in the right-hand side of (2-29) defines a closed quadratic form on the Hilbert space $H=L^2(\mathbb{R})$, with dense domain $D=\{h\in H: B^{1/2}h'\in H\}$. Let

$$\mathbb{B}:D(\mathbb{B})\longrightarrow H,\quad h\longmapsto -\partial_x(B(x)\partial_x h),$$

be the self-adjoint operator in H associated with the quadratic form (2-29) by Friedrich's representation theorem [Kato 1966]. Since $B(x)>0$ for all $x\in\mathbb{R}$ we know that $\mathbb{B}$ is positive, and using the fact that $x^2B(x)^{-1}\to 0$ as $|x|\to\infty$ it is easy to verify that $\mathbb{B}$ has compact resolvent in H, and hence purely discrete spectrum. The optimal constant in C_H in (2-29) is precisely the inverse of the lowest eigenvalue of $\mathbb{B}$:

$$C_H=\max\{\lambda^{-1}:\lambda\in\operatorname{spec}(\mathbb{B})\}. \tag{2-30}$$

By Sturm–Liouville theory, if $\mu=C_H^{-1}$ is the lowest eigenvalue of $\mathbb{B}$, there exists a positive eigenfunction $h\in D(\mathbb{B})$ such that $\mathbb{B}h=\mu h$. Setting $h(x)=f(e^x)$, we see that f is a positive solution of the ODE

$$-\partial_r\left(\frac{A(r)}{r}\partial_r f(r)\right)=\mu\frac{f(r)}{r},\quad r>0, \tag{2-31}$$

satisfying the boundary conditions $f(0)=f(+\infty)=0$. Moreover $\int_0^\infty A(r)f'(r)^2\,\mathrm{d}r/r<\infty$ by construction. It is not easy to guess from (2-31) whether μ is smaller or larger than 1, but under additional assumptions on the vortex profile it is possible to make another change of variables which puts (2-31) into a form that allows for a comparison with (2-12).

Lemma 2.9. *If the function A in (2-3) satisfies*

$$A\in C^2([0,+\infty))\quad\textit{and}\quad\sup_{r\ge1}\left(\frac{A(r)}{r^2}+\frac{A'(r)^2}{r^2A(r)}\right)\int_r^\infty\frac{s}{A(s)}\,\mathrm{d}s<\infty, \tag{2-32}$$

then the function $g:[0,+\infty)\to\mathbb{R}$ defined by $g(r)=A(r)^{1/2}f(r)/r$ is a solution of the ODE

$$-g''(r)-\frac{1}{r}g'(r)+\frac{1}{r^2}g(r)+V(r)g(r)=\frac{\mu}{A(r)}g(r), \tag{2-33}$$

with boundary conditions $g(0) = g(+\infty) = 0$, *where*

$$V(r) = \chi''(r) - \frac{1}{r}\chi'(r) + \chi'(r)^2 \quad \textit{and} \quad \chi(r) = \frac{1}{2}\log(A(r)). \tag{2-34}$$

Proof. Since f satisfies (2-31), a direct calculation shows that $g(r) := A(r)^{1/2} f(r)/r$ is a solution of (2-33), where the potential V is defined by (2-34). As for the boundary conditions, we recall that $\int_0^\infty A(r) f'(r)^2 \,\mathrm{d}r/r < \infty$; hence $\int_0^\infty |f'(r)| \,\mathrm{d}r < \infty$. As $f(r) = \int_0^r f'(s)\,\mathrm{d}s$, we have

$$\frac{|f(r)|}{r} \le \frac{1}{r}\left(\int_0^r \frac{s}{A(s)}\,\mathrm{d}s\right)^{1/2}\left(\int_0^r A(s) f'(s)^2 \frac{\mathrm{d}s}{s}\right)^{1/2} \xrightarrow[r\to 0]{} 0,$$

which shows that $g(r) \to 0$ as $r \to 0$. Similarly, since $f(r) = -\int_r^\infty f'(s)\,\mathrm{d}s$, we have

$$|g(r)| \le \frac{A(r)^{1/2}}{r}\left(\int_r^\infty \frac{s}{A(s)}\,\mathrm{d}s\right)^{1/2}\left(\int_r^\infty A(s) f'(s)^2 \frac{\mathrm{d}s}{s}\right)^{1/2} \xrightarrow[r\to +\infty]{} 0,$$

thanks to (2-32). □

Remark 2.10. The same arguments show that $r^2 g'(r) \to 0$ as $r \to 0$ and $g'(r) \to 0$ as $r \to +\infty$, at least along appropriate sequences.

Let L be the differential operator defined by

$$L = L_0 + V = -\partial_r^2 - \frac{1}{r}\partial_r + \frac{1}{r^2} + V(r), \tag{2-35}$$

where L_0 was introduced in (2-12). We know from (2-33) that $Lg = \mu A^{-1} g$, where $\mu = C_H^{-1}$ and g is the positive function defined in Lemma 2.9. On the other hand, we observed in Remark 2.3 that $L_0\phi = A^{-1}\phi$, where $\phi = \psi_*'$ is also a positive function vanishing at the origin and at infinity. Using Sturm–Liouville theory, we easily deduce the following useful criterion:

Corollary 2.11. *Under assumptions* (2-32), *if the function* V *defined by* (2-34) *does not change sign, the optimal constant in Hardy's inequality* (2-25) *satisfies* $C_H \le 1$ *if* $V \ge 0$, *and* $C_H \ge 1$ *if* $V \le 0$; *moreover,* $C_H = 1$ *only if* V *is identically zero.*

Proof. With the notation above, we have $L_0\phi - A^{-1}\phi = 0$ and

$$L_0 g - A^{-1} g = Lg - (A^{-1} + V)g = \mathcal{R}, \quad \text{where } \mathcal{R} = (\mu - 1)A^{-1} g - Vg. \tag{2-36}$$

Since $r\mathcal{R}\phi = r(\phi(L_0 g) - g(L_0\phi)) = (\mathrm{d}/\mathrm{d}r)(r(\phi' g - g'\phi))$, we have for $r_1 > r_0 > 0$ the identity

$$\int_{r_0}^{r_1} \mathcal{R}(r)\phi(r) r\,\mathrm{d}r = r\big(\phi'(r)g(r) - g'(r)\phi(r)\big)\Big|_{r=r_0}^{r=r_1}. \tag{2-37}$$

Now, we let r_0 tend to 0 and r_1 to $+\infty$ along appropriate sequences, in such a way that the right-hand side of (2-37) converges to zero. This possible, because we know that $\phi(r) = \mathcal{O}(r)$ and $\phi'(r) = \mathcal{O}(1)$ as $r \to 0$, while $\phi(r) = \mathcal{O}(1/r)$ and $\phi'(r) = \mathcal{O}(1/r^2)$ as $r \to +\infty$; moreover, the behavior of g in these limits is given in Lemma 2.9 and Remark 2.10. We thus deduce from (2-37) that $\int_0^\infty \mathcal{R}\phi r\,\mathrm{d}r = 0$, which is impossible if the function $\mathcal{R}$ has a constant sign and is not identically zero. So, if V does not change sign, we must have $\mu \ge 1$ if $V \ge 0$ and $\mu \le 1$ if $V \le 0$; moreover, $\mu = 1$ is possible only if $V \equiv 0$. Since $\mu = C_H^{-1}$, this gives the desired conclusion. □

Remark 2.12. As is easily verified, the optimal constant C_H in Hardy's inequality (2-25) is unchanged if the function $A(r)$ is replaced by $\lambda^{-2}A(\lambda r)$ for some $\lambda > 0$. This corresponds to a rescaling of the vortex profile ω_*.

We now give two important examples where the sign of $C_H - 1$ can be determined.

Example 2.13 (algebraic vortex). Given $\kappa > 1$, we define

$$\omega_*(r) = \frac{1}{(1+r^2)^\kappa}, \quad \psi_*'(r) = \frac{1}{2(\kappa-1)r}\left(1 - \frac{1}{(1+r^2)^{\kappa-1}}\right). \tag{2-38}$$

We have

$$A(r) = -\frac{\psi_*'(r)}{\omega_*'(r)} = \frac{1}{4\kappa(\kappa-1)r^2}((1+r^2)^{\kappa+1} - (1+r^2)^2).$$

When $\kappa = 2$ (Kaufmann–Scully vortex), inequality (2-25) holds with optimal constant $C_H = 1$, and is saturated for $f(r) = r^2/(1+r^2)^2$. Indeed, it is easy to verify that $A(r) = (1+r^2)^2/8$ and $V(r) = 0$ in that particular case. Taking $g(r) = r/(1+r^2)$, a direct calculation shows that $Lg = A^{-1}g$, so that $C_H = 1$.

If $\kappa > 2$, we prove in Section A2 that the potential V is positive, so that $C_H < 1$ by Corollary 2.11. Finally, if $1 < \kappa < 2$, the potential V is negative, implying that $C_H > 1$. Summarizing, for the family of algebraic vortices (2-38), the quadratic form J is coercive on $X_{\mathrm{rs}} \cap X_0$ if and only if $\kappa > 2$.

Example 2.14 (Gaussian vortex). We next consider the Oseen vortex given by

$$\omega_*(r) = e^{-r^2/4}, \quad \psi_*'(r) = \frac{2}{r}(1 - e^{-r^2/4}), \quad A(r) = \frac{4}{r^2}(e^{r^2/4} - 1). \tag{2-39}$$

In that case too, the potential V defined in (2-34) is positive; see Section A2. By Corollary 2.11, we conclude that $C_H < 1$, so that the quadratic form J is coercive on $X_{\mathrm{rs}} \cap X_0$. A numerical calculation gives the approximate value $C_H \approx 0.57$, so that $\gamma \approx 0.43$.

Remark 2.15. In a finite-dimensional situation, one can use statements such as Theorems 2.5 and 2.8 for showing the nonlinear Lyapunov stability of the corresponding steady solution, at least if the smoothness class of the relevant objects is C^2. More precisely, if a flow $\dot{x} = b(x)$ on a finite-dimensional manifold preserves a C^2 function f which attains a nondegenerate local maximum at $\bar{x}$, then the sets $\{f(x) > f(\bar{x}) - \epsilon\}$ are invariant under the flow and for small ϵ are well-approximated by the small balls given by the quadratic form $-\frac{1}{2}f''(\bar{x})[x-\bar{x}, x-\bar{x}]$. A standard way to see this is to write $f(x) > f(\bar{x}) - \epsilon$ as

$$-\frac{1}{2}f''(\bar{x})[x-\bar{x}, x-\bar{x}] - \int_0^1 (1-t)\big(f''((1-t)\bar{x} + tx) - f''(\bar{x})\big)[x-\bar{x}, x-\bar{x}]\,\mathrm{d}t < \epsilon.$$

When f'' is continuous at $\bar{x}$ and x is close to $\bar{x}$, the integral in this inequality is dominated by a small multiple of $-\frac{1}{2}f''(\bar{x})[x-\bar{x}, x-\bar{x}]$ and the usual Lyapunov stability statements follow. In our situation here the set $\mathcal{O}_{\bar{\omega}}$ is not a C^2 submanifold and the free energy functional $\omega \mapsto E(\omega) + \int_{\mathbb{R}^2} \Phi(\omega(x))\,\mathrm{d}x$ is not of class C^2. It is not hard to see directly that the expression

$$-\int_{\mathbb{R}^2}\int_0^1 (1-t)\Phi''\big((1-t)\bar{\omega}(x) + t\omega(x)\big)(\omega(x) - \bar{\omega}(x))^2\,\mathrm{d}t\,\mathrm{d}x$$

cannot be dominated by $-\frac{1}{2}\int_{\mathbb{R}^2}\Phi''(\bar\omega)(\omega(x)-\bar\omega(x))^2\,\mathrm{d}x$ in a suitable way. One may still use the invariance of the sets $\mathcal{U}_{\bar\omega,\epsilon} := \{\omega \in \mathcal{O}_{\bar\omega} \cap X_1 : E(\omega) > E(\bar\omega) - \epsilon\}$ under the Euler evolution, and possibly also the conservation of the second-order moment $I(\omega)$ defined in (1-15), to obtain Lyapunov-type stability statements. For results in this spirit when the domain occupied by the fluid is compact, the reader can consult [Burton 2005] and [Arnold and Khesin 1998, Section II.4]. Our situation here is somewhat complicated by the noncompactness of our flow domain $\mathbb{R}^2$, but under our assumptions one still has $\bigcap_{\epsilon>0}\mathcal{U}_{\bar\omega,\epsilon} = \{\bar\omega\}$ (by using the uniqueness of the maximizers discussed in [Carlen and Loss 1992], for example). This could be turned into Lyapunov-type stability statements, although not quite of the same form as in the C^2 case. The important point is that there are estimates for the proximity of "almost maximizers" to the exact maximizers, an issue that also appears in other problems, such as the stability of the isoperimetric inequality [Fusco et al. 2008], and of the Sobolev inequality [Bianchi and Egnell 1991].

In the present work our focus is on quadratic forms, due to their applicability to the viscous case. Of course, at the level of the linearized inviscid equation $\omega_t + \bar u \cdot \nabla\omega + u \cdot \nabla\bar\omega = 0$, the quadratic form J does provide Lyapunov stability in the space X_1 if inequality (2-25) holds with $C_H < 1$. We note that the linearized analysis in other topologies can be more complicated; see for example [Bedrossian et al. 2019].

2C. ***The quadratic form J without mass constraint.*** In this short section we make a few remarks on the index of the quadratic form (2-9) when considered on the whole space X defined by (2-4), and not only on the subspace X_0 given by (2-5). Our first observation is that, due to lack of scale invariance in this context, the form J cannot be positive on X if the underlying steady state $\bar\omega$ is sharply concentrated near the origin. To see this, we consider the rescaled vortex $\bar\omega_\lambda(x) = \lambda^2\bar\omega(\lambda x)$ and the associated weight function $\mathcal{A}_\lambda(x) = \lambda^{-2}\mathcal{A}(\lambda x)$; see Remark 2.12. We denote by J_λ the quadratic form on X corresponding to the steady state $\bar\omega_\lambda$, namely the form (2-9) where $\mathcal{A}$ is replaced by $\mathcal{A}_\lambda$. If $\omega \in X$ and $\omega_\lambda(x) = \lambda^2\omega(\lambda x)$, a simple calculation shows that

$$J_\lambda(\omega_\lambda) = J(\omega) - \frac{M_0^2}{4\pi}\log(\lambda), \quad \text{where } M_0 = \int_{\mathbb{R}^2}\omega(x)\,\mathrm{d}x.$$

If $M_0 \neq 0$, it is clear that $J_\lambda(\omega_\lambda) < 0$ when $\lambda > 0$ is sufficiently large, so that the quadratic form J_λ cannot be positive in this regime.

Remark 2.16. The negative direction arising by such a rescaling is related to a particular choice of the unit of length implicitly involved in the kernel $\frac{1}{2\pi}\log|x|$. In writing $\log|x|$, we imply that x is dimensionless. When x is measured in some units of length, we should write the kernel as $\frac{1}{2\pi}\log(|x|/r_0)$, where r_0 is a reference length. The choice of r_0 does not affect the behavior of the system, and in the stability analysis based on J it can be compensated for by adding to the quadratic form J a suitable multiple of the quantity $\left(\int_{\mathbb{R}^2}\omega(x,t)\,\mathrm{d}x\right)^2$, which is preserved by the evolution. Hence, as one can expect, the stability analysis is independent of the choice of the reference length r_0, or, equivalently, of the scaling parameter λ above.

We next argue that, for any vortex $\bar\omega$ satisfying Hypotheses 2.1, the index of the quadratic form is well-defined in the sense that J has (at most) a finite number of negative directions. In view of Theorem 2.5, it is sufficient to evaluate J on radially symmetric functions $\omega \in X_{\mathrm{rs}}$. The following expression will be useful:

Lemma 2.17. *For any $\omega \in X_{\rm rs}$, we have*

$$J(\omega) = \pi \int_0^\infty A(r)\omega(r)^2 r\,\mathrm{d}r + \pi \int_0^\infty \int_0^\infty \log(\max(r,s)) r\,\omega(r)\,s\,\omega(s)\,\mathrm{d}r\,\mathrm{d}s. \tag{2-40}$$

Proof. Here and below, with a slight abuse of notation, we consider any $\omega \in X_{\rm rs}$ as a function of the one-dimensional variable $r = |x|$. For such vorticities, the first integral in (2-9) obviously gives the first term in (2-40), so it remains to establish the following expression of the energy:

$$E(\omega) = -\pi \int_0^\infty \int_0^\infty \log(\max(r,s)) r\,\omega(r)\,s\,\omega(s)\,\mathrm{d}r\,\mathrm{d}s, \quad \omega \in X_{\rm rs}. \tag{2-41}$$

To this end, we introduce polar coordinates $x = re^{i\theta}$, $y = se^{i\zeta}$ to compute the right-hand side of (1-13), and we use the identity

$$\int_0^{2\pi}\int_0^{2\pi} \log|re^{i\theta} - se^{i\zeta}|\,\mathrm{d}\theta\,\mathrm{d}\zeta = 2\pi \int_0^{2\pi} \log|re^{i\theta} - s|\,\mathrm{d}\theta = 4\pi^2 \log(\max(r,s)). \tag{2-42}$$

The formula (2-42) is well known and can be derived in many ways. For example, assuming that r is a fixed positive number, we interpret the last integral as a function of $s \in \mathbb{C}$. This expression obviously depends only on $|s|$, is continuous everywhere, and is analytic both inside and outside of the circle $|s| = r$. Inside the circle it has to be constant and outside the circle it coincides with the potential of a point particle of mass 2π located at the origin, which is $2\pi \log|s|$. This gives (2-42), and (2-41) follows. □

Applying the change of variables $w(r) = \omega(r)A(r)^{1/2}$, so that $w \in Y = L^2(\mathbb{R}_+, r\,\mathrm{d}r)$ when $\omega \in X_{\rm rs}$, the formula (2-40) becomes

$$\frac{1}{\pi} J(\omega) = \int_0^\infty w(r)^2 r\,\mathrm{d}r - \int_0^\infty \int_0^\infty \mathscr{k}(r,s) w(r) w(s) rs\,\mathrm{d}r\,\mathrm{d}s, \tag{2-43}$$

where $\mathscr{k}(r,s) = -\log(\max(r,s))A(r)^{-1/2}A(s)^{-1/2}$. Under Hypotheses 2.1, we have the lower bound $A(r) \geq C(1+r)^\beta$ for some $\beta > 2$, which implies that

$$\int_0^\infty \int_0^\infty \mathscr{k}(r,s)^2\, rs\,\mathrm{d}r\,\mathrm{d}s < \infty.$$

This means that the right-hand side of (2-43) is the quadratic form in Y associated with a self-adjoint operator of the form $\mathbf{1} - \mathscr{K}$, where $\mathbf{1}$ is the identity and $\mathscr{K}$ is a Hilbert–Schmidt perturbation. By compactness, this operator has (at most) a finite number of negative eigenvalues, which means that the index of the quadratic form J on X is well-defined.

The eigenvalues of $\mathscr{K}$ can also be thought of as eigenvalues of the quadratic form (2-41) with respect to the reference form $\omega \mapsto \pi \int_0^\infty A(r)\omega(r)^2 r\,\mathrm{d}r$. As is easily verified, if λ is such an eigenvalue, the corresponding eigenfunction ω satisfies

$$-\psi(r) = \lambda A(r)\omega(r), \quad \text{where } \psi(r) = \int_0^\infty \log(\max(r,s))\,s\,\omega(s)\,\mathrm{d}s. \tag{2-44}$$

Since $\omega(r) = \psi''(r) + (1/r)\psi'(r)$, the first relation in (2-44) is an ordinary differential equation for the stream function $\psi : \mathbb{R}_+ \to \mathbb{R}$, to be solved with the boundary conditions

$$\psi'(0) = 0 \quad \text{and} \quad \lim_{r\to+\infty} \big(\psi(r)\log(2r) - \psi(2r)\log(r)\big) = 0,$$

which can be deduced from the expression of ψ in (2-44). For the Lamb–Oseen vortex (2-39) a numerical computation gives the largest eigenvalue $\lambda \approx 0.7127$, thus suggesting that the form J is strictly positive definite on the whole space X_{rs} in that case. In contrast, the largest eigenvalue for the algebraic vortices (2-38) seems to exceed the threshold value 1, indicating that for those vortices the form J is not positive definite without additional constraints on ω.

2D. ***The maximal energy as a function of the constraints.*** In Section 1A we considered the classical problem of maximizing a function $f : \mathbb{R}^n \to \mathbb{R}$ under a family of constraints of the form $g_1 = c_1, \dots, g_m = c_m$, where $g_1, \dots, g_m : \mathbb{R}^n \to \mathbb{R}$. Given $c = (c_1, \dots, c_m) \in \mathbb{R}^m$, we recall the notation $X_c = \{x \in \mathbb{R}^n : g_1(x) = c_1, \dots, g_m(x) = c_m\}$. Assuming that f reaches a nondegenerate maximum on X_c at some point $\bar{x} \in X_c$ where the first-order derivatives $g_1'(\bar{x}), \dots, g_m'(\bar{x})$ are linearly independent, we introduced the quadratic form $\mathcal{Q}$ defined by (1-5), which is the second-order differential of the Lagrange function (1-7) at $\bar{x}$. In the present section, we are interested in the index of the form $\mathcal{Q}$ on larger subspaces than $T_{\bar{x}}X_c$. As was already mentioned, this question is closely related to concavity properties of the function M defined by (1-8) or, almost equivalently, to convexity properties of the set $S = \{(g_1(x), \dots, g_m(x), f(x)) : x \in \mathbb{R}^n\} \subset \mathbb{R}^{m+1}$ near its "upper boundary".

The situation becomes particularly transparent if we use adapted coordinates which, as it turns out, have a fairly complete analogy in the two-dimensional Euler case. Let us assume that we can introduce new coordinates $(c_1, \dots, c_m, y_1, \dots, y_{n-m})$ in $\mathbb{R}^n$ such that, as before, $c_1, \dots, c_m$ are the values of the constraints $g_1, \dots, g_m$, and the additional coordinates $y_1, \dots, y_{n-m}$ are chosen so that the points having coordinates $(c_1, \dots, c_m, 0, \dots, 0)$ are those where f attains its maximum on X_c.[5] Writing $M(c_1, \dots, c_m) = f(c_1, \dots, c_m, 0, \dots, 0)$ as in (1-8), one verifies that

$$\frac{\partial M}{\partial c_j}(c_1, \dots, c_m) = \lambda_j, \quad j = 1, \dots, m, \tag{2-45}$$

where $\lambda_1, \dots, \lambda_m$ are the Lagrange multipliers introduced in (1-4). Moreover the extremality condition on X_c implies that

$$\frac{\partial f}{\partial y_k}(c_1, \dots, c_m, 0, \dots, 0) = 0, \quad k = 1, \dots, n-m.$$

We infer that

$$D^2 f(c_1, \dots, c_m, 0, \dots, 0) = \begin{pmatrix} (\partial^2 f/(\partial c_i \partial c_j))_{i,j=1}^m & 0 \\ 0 & (\partial^2 f/(\partial y_k \partial y_\ell))_{k,\ell=1}^{n-m} \end{pmatrix}, \tag{2-46}$$

where all derivatives are evaluated at the point $(c_1, \dots, c_m, 0, \dots, 0)$. The first submatrix in the right-hand side of (2-46) is precisely the Hessian of M, and the second submatrix is always negative definite, due to our assumption that f reaches a maximum at $(y_1, \dots, y_{n-m}) = (0, \dots, 0)$ for any fixed value of $c_1, \dots, c_m$. So we conclude that the quadratic form $\mathcal{Q}$ defined in (1-5) is negative definite at $\bar{x}$ if and only if the Hessian of M is negative definite at $(c_1, \dots, c_m)$, where $c_j = g_j(\bar{x})$ for $j = 1, \dots, m$.

[5] In a nondegenerate situation, the local existence of such a coordinate system is clear by standard arguments, but globally the situation can, of course, be more complicated.

Another interesting object is the function

$$\begin{aligned} N(\lambda_1, \dots, \lambda_m) &= \sup_{x \in \mathbb{R}^n} (f(x) - \lambda_1 g_1(x) - \cdots - \lambda_m g_m(x)) \\ &= \sup_{c \in \mathbb{R}^m} (M(c_1, \dots, c_m) - \lambda_1 c_1 - \cdots - \lambda_m c_m), \end{aligned} \tag{2-47}$$

which is the *Legendre transform* of M. Under appropriate assumptions, the main one being the concavity of M, this quantity is well-defined and the relation (2-45) can be inverted (at least locally) via the formula

$$c_j = -\frac{\partial N}{\partial \lambda_j}(\lambda_1, \dots, \lambda_m), \quad j = 1, \dots, m. \tag{2-48}$$

We now return to the infinite-dimensional framework of the two-dimensional Euler equation, with the manifold $\mathbb{R}^n$ replaced by the phase space $\mathcal{P}$ introduced in Section 1B, the function f replaced by the energy E in (1-13), the constraints g_j replaced by the Casimir functionals $h(a, \omega)$ in (1-10), and the submanifolds X_c replaced by the orbits $\mathcal{O}_\omega$ in (1-12). In that case we have

$$\max_{\omega \in \mathcal{O}_{\bar\omega}} E(\omega) = E(\bar\omega^*), \tag{2-49}$$

where, as before, $\bar\omega^*$ denotes the symmetric decreasing rearrangement of an element $\bar\omega \in \mathcal{P}$. As $\mathcal{O}_{\bar\omega}$ is characterized in terms of the functionals $h(a, \omega)$ defined in (1-10), the energy of the maximizer $\bar\omega^*$ in $\mathcal{O}_{\bar\omega}$ can also be expressed in terms of the constraint function $a \to h(a, \bar\omega)$. It turns out that the representation formula is quite explicit.

Proposition 2.18. *Given $\bar\omega \in \mathcal{P}$, we define $\hbar(a) = \pi^{-1} h(a, \bar\omega) = \pi^{-1} |\{\bar\omega > a\}|$ for any $a > 0$. Then*

$$\mathcal{E}(\hbar) := \max_{\substack{\omega \in \mathcal{P} \\ h(\cdot, \omega) = \pi\hbar}} E(\omega) = \frac{\pi}{8} \int_0^m \int_0^m L(\hbar(a), \hbar(b))\, \mathrm{d}a\, \mathrm{d}b + \frac{1}{8\pi} M_0^2, \tag{2-50}$$

where $m = \max \bar\omega$, $M_0 = \int_{\mathbb{R}^2} \bar\omega \, \mathrm{d}x = \pi \int_0^m \hbar(a)\, \mathrm{d}a$, and

$$L(R, S) = -RS \log \max(R, S) - \tfrac{1}{2} \min(R, S)^2. \tag{2-51}$$

Proof. Replacing $\bar\omega$ with $\bar\omega^*$ (an operation that does not affect the function $\hbar$), we can assume that $\bar\omega$ is radially symmetric and nonincreasing in the radial direction. In view of (2-49), we then have $\mathcal{E}(\hbar) = E(\bar\omega)$, and if we consider $\bar\omega$ as a function of the radius $r = |x|$, we observe that $\hbar(a) = (\bar\omega^{-1}(a))^2$ wherever $\bar\omega$ is strictly decreasing. To compute $E(\bar\omega)$, we start from the expression (2-41), and we introduce the functions

$$k(r, s) = -rs \log \max(r, s), \quad K(R, S) = L(R, S) + RS.$$

Clearly $K(R, 0) = 0$, $K(0, S) = 0$ for $R, S > 0$, and one can verify by direct calculation that $K(R, S)$ is twice continuously differentiable on $(0, \infty) \times (0, \infty)$, with

$$\frac{\partial^2 K}{\partial R \partial S}(R, S) = -\log \max(R, S), \quad R, S > 0.$$

So the function $(r, s) \mapsto K(r^2, s^2)$ is twice continuously differentiable on $[0, \infty) \times [0, \infty)$ and

$$\frac{1}{8} \frac{\partial^2}{\partial r \partial s} K(r^2, s^2) = k(r, s).$$

Integrating by parts in (2-41) and recalling that $m = \max \bar{\omega}$, we can thus write

$$\begin{aligned} E(\bar{\omega}) &= \frac{\pi}{8}\int_0^\infty\int_0^\infty \frac{\partial^2}{\partial r\,\partial s} K(r^2, s^2)\,\bar{\omega}(r)\,\bar{\omega}(s)\,\mathrm{d}r\,\mathrm{d}s = \frac{\pi}{8}\int_0^\infty\int_0^\infty K(r^2, s^2)\,\mathrm{d}\bar{\omega}(r)\,\mathrm{d}\bar{\omega}(s) \\ &= \frac{\pi}{8}\int_0^m\int_0^m K\big((\bar{\omega}^{-1}(a))^2, (\bar{\omega}^{-1}(b))^2\big)\,\mathrm{d}a\,\mathrm{d}b = \frac{\pi}{8}\int_0^m\int_0^m K(\hbar(a), \hbar(b))\,\mathrm{d}a\,\mathrm{d}b \\ &= \frac{\pi}{8}\int_0^m\int_0^m L(\hbar(a), \hbar(b))\,\mathrm{d}a\,\mathrm{d}b + \frac{1}{8\pi}M_0^2, \end{aligned} \tag{2-52}$$

where we have formally used the substitutions $\bar{\omega}(r) = a,\ \bar{\omega}(s) = b$. This is straightforward when $\bar{\omega}$ is strictly decreasing, and the general case where $\bar{\omega}$ is nonincreasing can be treated by integrating only over the intervals where $\bar{\omega}$ is strictly decreasing. □

We now make a more precise comparison with the finite-dimensional situation above. Let us assume that $\bar{\omega} \in \mathscr{P}$ is radially symmetric with $\partial_r\bar{\omega}(r) < 0$ for all $r > 0$ and $\partial_r^2\bar{\omega}(0) < 0$. To eliminate the translational symmetries, we work with the manifold

$$\widetilde{\mathscr{P}} = \{\omega \in \mathscr{P} : M_0(\omega) = M_0(\bar{\omega}),\ M_j(\omega) = 0,\ j = 1, 2\}, \tag{2-53}$$

where M_0, M_j are as in (1-11), (1-15). If $\eta \in \mathscr{X}_1$ (see (2-24)) is smooth and compactly supported with sufficiently small C^2 norm, then $\bar{\omega} + \eta \in \widetilde{\mathscr{P}}$. Denoting by η_{rs} the projection of η onto the subspace X_{rs} defined in (2-11), we can take the quantities $h(a, \bar{\omega}+\eta_{\mathrm{rs}})$ and $\eta_{\mathrm{rs}}^{\perp} := \eta - \eta_{\mathrm{rs}}$ as the (approximate) analogues of the coordinates c_j and y_k, respectively. The analogy is not perfect, due to the stronger-than-ideal assumptions on η, but it is sufficient for concluding that, when $\bar{\omega} = \bar{\omega}^*$, the negative-definiteness of Arnold's form (1-18) on the tangent space $T_{\bar{\omega}}\widetilde{\mathscr{P}}$ is strongly related to the concavity of the energy E in the variable[6] $\hbar$ at the function $\bar{\hbar}(a) = \pi^{-1}h(a, \bar{\omega})$. In some sense the expression (2-50) is "trying to be concave", although not quite achieving this: the function $L(R, S)$ is separately concave, but not concave. The second variation on the space X_0 is given by the quadratic form which takes a function $\xi(a)$ with $\int_0^m \xi(a)\,\mathrm{d}a = 0$ to

$$\frac{\pi}{8}\int_0^m\int_0^m \big(D_1^2L(\hbar(a), \hbar(b))\xi(a)^2 + 2D_1D_2L(\hbar(a), \hbar(b))\xi(a)\xi(b) + D_2^2L(\hbar(a), \hbar(b))\xi(b)^2\big)\,\mathrm{d}a\,\mathrm{d}b.$$

Due to the separate concavity of L the first term and the third term are negative, but the second one can lead to the form being indefinite. In view of our previous considerations, the negativity of the form is equivalent to the validity of the Hardy inequality (2-25) with $C_H \le 1$, and it is not hard to verify directly that this is indeed the case. As an analogue of (2-45), we also note that the variational derivative of $\mathscr{E}$ with respect to $\hbar$ is

$$\frac{1}{\pi}\frac{\delta\mathscr{E}}{\delta\hbar}(a) = \Lambda(a) = -\Phi'(a). \tag{2-54}$$

We will not go into the details as we will not work with this expression. The reader can also derive the analogue of (2-48) (under appropriate assumptions).

[6] It is perhaps worth recalling that E is convex in ω on the subspace given by $\int_{\mathbb{R}^2} \omega\,\mathrm{d}x = 0$. However, in some regions it may be concave in $\hbar$, at least on the subspace given by $\int_0^\infty \hbar(a)\,\mathrm{d}a = 0$.

3. Global maximization of the free energy

In the previous section we observed that some radially symmetric vortices $\bar\omega$, including the Gaussian vortex (2-39) and the algebraic vortex (2-38) with $\kappa > 2$, are nondegenerate local maxima of the associated free energy functional (1-17) once restricted to the manifold $\widetilde{\mathscr{P}}$ defined in (2-53). This was established by showing that the second-order differential $F''(\bar\omega)$ is strictly negative definite on the tangent space $T_{\bar\omega}\widetilde{\mathscr{P}}$. We now follow a different approach, which relies on the direct method in the calculus of variations: under appropriate assumptions on the function Φ in (1-17), we show that the free energy $F(\omega)$ has a global maximum on the set of all vorticity distributions with a fixed mass M. By construction, if $\bar\omega$ is any maximizer obtained in this way, the conclusion of Theorem 2.8 applies with $\gamma \geq 0$, so that Hardy's inequality (2-25) holds with $C_H \leq 1$. Note also that, according to the discussion in Section 2D, prescribing Φ amounts to fixing the "Lagrange multipliers" in our constrained maximization problem.

We start with a preliminary result, which is probably well known. For the reader's convenience, the proof is reproduced in Section A1.

Proposition 3.1. *Assume that $f \in L^1(\mathbb{R}^n)$ is nonnegative and that $M := \int_{\mathbb{R}^n} f(x)\,\mathrm{d}x > 0$. Then*

$$M + \int_{\mathbb{R}^n} (\log_- |x|)\, f(x)\,\mathrm{d}x \lesssim M + \int_{\mathbb{R}^n} \left(\log_+ \frac{f(x)}{M}\right) f(x)\,\mathrm{d}x, \tag{3-1}$$

$$M + \int_{\mathbb{R}^n} (\log_+ |x|)\, f(x)\,\mathrm{d}x \gtrsim M + \int_{\mathbb{R}^n} \left(\log_- \frac{f(x)}{M}\right) f(x)\,\mathrm{d}x, \tag{3-2}$$

where the implicit constants only depend on the space dimension n. Moreover, if f is radially symmetric and nonincreasing in the radial direction, then the reverse inequalities also hold.

We next specify the function space in which we shall solve our maximization problem.

Definition 3.2. Given any $M > 0$, we denote by X_M the set of all $\omega \in L^1(\mathbb{R}^2)$ such that $\omega(x) \geq 0$ for almost all $x \in \mathbb{R}^2$ and

$$\int_{\mathbb{R}^2} \omega(x)\,\mathrm{d}x = M, \qquad \int_{\mathbb{R}^2} \omega(x)\log(1+|x|)\,\mathrm{d}x < \infty, \qquad \int_{\mathbb{R}^2} \omega(x)\log(1+\omega(x))\,\mathrm{d}x < \infty. \tag{3-3}$$

For later use we observe that, if $\omega \in X_M$ and if ω^* denotes the symmetric nonincreasing rearrangement of ω, then $\int_{\mathbb{R}^2} \omega^*(x)\,\mathrm{d}x = \int_{\mathbb{R}^2} \omega(x)\,\mathrm{d}x = M$ and

$$\int_{\mathbb{R}^2} \omega^*(x)\log(1+|x|)\,\mathrm{d}x \leq \int_{\mathbb{R}^2} \omega(x)\log(1+|x|)\,\mathrm{d}x < \infty,$$

$$\int_{\mathbb{R}^2} \omega^*(x)\log(1+\omega^*(x))\,\mathrm{d}x = \int_{\mathbb{R}^2} \omega(x)\log(1+\omega(x))\,\mathrm{d}x < \infty.$$

This shows that the set $X_M \subset L^1(\mathbb{R}^2)$ is invariant under the action of the symmetric nonincreasing rearrangement.

For $\omega \in X_M$, we consider the free energy defined by $F(\omega) = E(\omega) + S(\omega)$, where

$$E(\omega) = \frac{1}{4\pi}\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \log\frac{1}{|x-y|}\,\omega(x)\,\omega(y)\,\mathrm{d}x\,\mathrm{d}y, \qquad S(\omega) = \int_{\mathbb{R}^2} \Phi(\omega(x))\,\mathrm{d}x.$$

We have shown in Proposition 2.2 that the energy $E(\omega)$ is finite for any $\omega \in X_M$. Unlike in Section 2, the function Φ in the entropy term is not related here to any radially symmetric vortex, but is an arbitrary function satisfying the following properties:

Hypotheses 3.3. *The function $\Phi : [0, +\infty) \to \mathbb{R}$ is continuous with $\Phi(0) = 0$. Moreover, there exist constants $C_1 \in \mathbb{R}$, $C_2 < M/(8\pi)$, and $C_3 > M/(8\pi)$ such that*

$$\begin{aligned} \Phi(\omega) &\le C_1\omega + C_2\omega \log \frac{M}{\omega}, \quad \text{when } \omega \le M, \\ \Phi(\omega) &\le C_1\omega - C_3\omega \log \frac{\omega}{M}, \quad \text{when } \omega \ge M. \end{aligned} \tag{3-4}$$

Under Hypotheses 3.3, the positive part of Φ satisfies $\Phi_+(\omega) \le C\omega(1+|\log(\omega/M)|)$ for some constant $C > 0$, and this implies in particular that the entropy $S(\omega)$ is well-defined in $\mathbb{R} \cup \{-\infty\}$ for any $\omega \in X_M$. We are now in a position to state the main result of this section.

Theorem 3.4. *Fix any $M > 0$. Under Hypotheses 3.3, there exists $\bar{\omega} \in X_M$ such that*

$$F(\bar{\omega}) = E(\bar{\omega}) + S(\bar{\omega}) = \sup_{\omega \in X_M} (E(\omega) + S(\omega)).$$

Moreover $\bar{\omega}$ can be chosen to be radially symmetric and nonincreasing in the radial direction.

The proof of Theorem 3.4 is divided into two parts. The first one consists in showing that the free energy F is bounded from above on X_M, and that there exists a maximizing sequence which is convergent in $L^1(\mathbb{R}^2)$. We formulate this in a separate statement:

Proposition 3.5. *Under Hypotheses 3.3, the free energy $F = E + S$ is bounded from above on the space X_M:*

$$F_M := \sup_{\omega \in X_M} (E(\omega) + S(\omega)) < \infty.$$

Moreover, there exists a maximizing sequence $(\omega_j)_{j\in\mathbb{N}}$ in X_M which converges in $L^1(\mathbb{R}^2)$ to some limiting profile $\bar{\omega} = \bar{\omega}^* \in X_M$ as $j \to +\infty$, and we have $S(\bar{\omega}) > -\infty$.

Proof. Our starting point is the logarithmic Hardy–Littlewood–Sobolev inequality

$$E(\omega) + \frac{M}{8\pi} \int_{\mathbb{R}^2} \omega \log \frac{M}{\omega}\, \mathrm{d}x \le \frac{M^2}{8\pi}(1 + \log \pi), \tag{3-5}$$

which holds for all $\omega \in X_M$; see [Carlen and Loss 1992]. In view of (3-4), we deduce from (3-5) that

$$\begin{aligned} E(\omega) + S(\omega) &+ \left(\frac{M}{8\pi} - C_2\right) \int_{\omega<M} \omega \log \frac{M}{\omega}\, \mathrm{d}x + \left(C_3 - \frac{M}{8\pi}\right) \int_{\omega>M} \omega \log \frac{\omega}{M}\, \mathrm{d}x \\ &\le E(\omega) + C_1 M + \frac{M}{8\pi} \int_{\mathbb{R}^2} \omega \log \frac{M}{\omega}\, \mathrm{d}x \le C_1 M + \frac{M^2}{8\pi}(1 + \log \pi). \end{aligned} \tag{3-6}$$

Since $C_2 < M/(8\pi)$ and $C_3 > M/(8\pi)$, this proves that $F_M \le C_1 M + M^2(1 + \log \pi)/(8\pi)$.

Now, let $(\omega_j)_{j\in\mathbb{N}}$ be a sequence in X_M such that $E(\omega_j) + S(\omega_j) \to F_M$ as $j \to +\infty$. If we denote by $(\omega_j)^* \in X_M$ the symmetric nonincreasing rearrangement of ω_j, we know that $E((\omega_j)^*) \ge E(\omega_j)$ and $S((\omega_j)^*) = S(\omega_j)$ for all $j \in \mathbb{N}$, so that $((\omega_j)^*)_{j\in\mathbb{N}}$ is a fortiori a maximizing sequence. So we assume

henceforth that $\omega_j = (\omega_j)^*$; i.e., ω_j is radially symmetric and nonincreasing in the radial direction. In that case, there exists a constant $C_0 > 0$ such that

$$\int_{\mathbb{R}^2} \omega_j(x) \left| \log \frac{\omega_j(x)}{M} \right| \mathrm{d}x \le C_0 \quad \text{and} \quad \int_{\mathbb{R}^2} \omega_j(x) |\log |x|| \, \mathrm{d}x \le C_0 \tag{3-7}$$

for all $j \in \mathbb{N}$. Indeed, the first inequality in (3-7) follows directly from (3-6), and the second one is a consequence of the first inequality and of Proposition 3.1, since $\omega_j = (\omega_j)^*$.

It remains to verify that one can extract from $(\omega_j)_{j\in\mathbb{N}}$ a convergent subsequence in $L^1(\mathbb{R}^2)$. We recall that $\omega_j(x)$ is a nonincreasing function of the radial variable $|x|$, which satisfies the uniform pointwise estimate $0 \le \omega_j(x) \le M/(\pi|x|^2)$; see (A-3) below. By Helly's selection theorem [Rudin 1953], there exists a subsequence, still denoted by $(\omega_j)_{j\in\mathbb{N}}$, which converges pointwise to some limit $\bar\omega : \mathbb{R}^2 \to \mathbb{R}_+$ as $j \to +\infty$. It is clear that $\bar\omega$ is radially symmetric and nonincreasing, so that $\bar\omega = \bar\omega^*$, and Fatou's lemma implies that $\int_{\mathbb{R}^2} \bar\omega(x)\,\mathrm{d}x \le M$. Using in addition (3-7), we obtain similarly

$$\int_{\mathbb{R}^2} \bar\omega(x) \left| \log \frac{\bar\omega(x)}{M} \right| \mathrm{d}x \le C_0 \quad \text{and} \quad \int_{\mathbb{R}^2} \bar\omega(x) |\log |x|| \, \mathrm{d}x \le C_0. \tag{3-8}$$

To prove the convergence in $L^1(\mathbb{R}^2)$ we take the decomposition, for any $\epsilon \in (0, 1)$,

$$\int_{\mathbb{R}^2} |\omega_j(x) - \bar\omega(x)| \, \mathrm{d}x = \int_{A_\epsilon} |\omega_j(x) - \bar\omega(x)| \, \mathrm{d}x + \int_{\mathbb{R}^2 \setminus A_\epsilon} |\omega_j(x) - \bar\omega(x)| \, \mathrm{d}x, \tag{3-9}$$

where $A_\epsilon = \{x \in \mathbb{R}^2 : \epsilon \le |x| \le \epsilon^{-1}\}$. The integral over A_ϵ converges to zero as $j \to +\infty$ by the dominated convergence theorem, and in view of (3-7), (3-8) the integral over $\mathbb{R}^2 \setminus A_\epsilon$ is bounded by $2C_0/|\log \epsilon|$ uniformly in j. It thus follows from (3-9) that

$$\limsup_{j\to+\infty} \int_{\mathbb{R}^2} |\omega_j(x) - \bar\omega(x)| \, \mathrm{d}x \le \frac{2C_0}{|\log \epsilon|} \xrightarrow[\epsilon\to 0]{} 0,$$

which shows that $\omega_j \to \bar\omega$ in $L^1(\mathbb{R}^2)$. In particular $\int_{\mathbb{R}^2} \bar\omega(x)\,\mathrm{d}x = M$, so that $\bar\omega \in X_M$.

Finally, if we take the decomposition $\Phi = \Phi_+ - \Phi_-$, where Φ_+, Φ_- denote the positive and negative parts of Φ, we have the lower bound

$$S(\bar\omega) \ge -\int_{\mathbb{R}^2} \Phi_-(\bar\omega(x))\,\mathrm{d}x \ge -\liminf_{j\to+\infty} \int_{\mathbb{R}^2} \Phi_-(\omega_j(x))\,\mathrm{d}x, \tag{3-10}$$

where the second inequality is again obtained by Fatou's lemma. But we have the identity

$$\int_{\mathbb{R}^2} \Phi_-(\omega_j(x))\,\mathrm{d}x = \int_{\mathbb{R}^2} \Phi_+(\omega_j(x))\,\mathrm{d}x - S(\omega_j) = \int_{\mathbb{R}^2} \Phi_+(\omega_j(x))\,\mathrm{d}x + E(\omega_j) - F(\omega_j),$$

where the first two terms in the right-hand side are bounded uniformly in j by (3-7), in view of Hypotheses 3.3 and Proposition 2.2, whereas $F(\omega_j)$ is bounded from below since (ω_j) is a maximizing sequence for F. We conclude that the right-hand side of (3-10) is finite, so that $S(\bar\omega) > -\infty$. □

To conclude the proof of Theorem 3.4, it remains to show that the free energy is upper semicontinuous along the maximizing sequence constructed in Proposition 3.5, namely

$$E(\bar{\omega}) + S(\bar{\omega}) \geq \limsup_{j\to+\infty}(E(\omega_j) + S(\omega_j)) = F_M. \tag{3-11}$$

This will imply that $E(\bar{\omega}) + S(\bar{\omega}) = F_M$, which is the desired result.

Proof of Theorem 3.4. Let $(\omega_j)_{j\in\mathbb{N}}$ be the maximizing sequence defined in Proposition 3.5, and $\bar{\omega} \in X_M$ be the limiting profile. Given any sufficiently large $R > 0$, we take the decomposition

$$\begin{aligned}\omega_j(x) &= \omega_j(x)\,\mathbf{1}_{\{|x|\leq R\}} + \omega_j(x)\,\mathbf{1}_{\{|x|>R\}} =: \omega^1_{jR}(x) + \omega^2_{jR}(x),\\ \bar{\omega}(x) &= \bar{\omega}(x)\,\mathbf{1}_{\{|x|\leq R\}} + \bar{\omega}(x)\,\mathbf{1}_{\{|x|>R\}} =: \bar{\omega}^1_R(x) + \bar{\omega}^2_R(x)\end{aligned}$$

for all $x \in \mathbb{R}^2$. We thus have

$$\begin{aligned}E(\omega_j) + S(\omega_j) &= E(\omega^1_{jR}) + S(\omega^1_{jR}) + 2E(\omega^1_{jR}, \omega^2_{jR}) + E(\omega^2_{jR}) + S(\omega^2_{jR}),\\ E(\bar{\omega}) + S(\bar{\omega}) &= E(\bar{\omega}^1_R) + S(\bar{\omega}^1_R) + 2E(\bar{\omega}^1_R, \bar{\omega}^2_R) + E(\bar{\omega}^2_R) + S(\bar{\omega}^2_R),\end{aligned}$$

where $E(\omega_1, \omega_2)$ is the bilinear form associated with the energy functional:

$$E(\omega_1, \omega_2) = -\frac{1}{4\pi}\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \log|x-y|\,\omega_1(x)\,\omega_2(y)\,\mathrm{d}x\,\mathrm{d}y.$$

The upper-semicontinuity property (3-11) can be deduced from the following assertions:

$$\limsup_{j\to+\infty}(E(\omega^1_{jR}) + S(\omega^1_{jR})) \leq E(\bar{\omega}^1_R) + S(\bar{\omega}^1_R), \tag{3-12}$$

$$\sup_{j\in\mathbb{N}}\big(2E(\omega^1_{jR}, \omega^2_{jR}) + E(\omega^2_{jR}) + S(\omega^2_{jR})\big) \leq \delta_1(R) \xrightarrow[R\to+\infty]{} 0, \tag{3-13}$$

$$2E(\bar{\omega}^1_R, \bar{\omega}^2_R) + E(\bar{\omega}^2_R) + S(\bar{\omega}^2_R) = \delta_2(R) \xrightarrow[R\to+\infty]{} 0. \tag{3-14}$$

Indeed, assuming that (3-12)–(3-14) hold, we obtain

$$\limsup_{j\to+\infty}(E(\omega_j) + S(\omega_j)) - (E(\bar{\omega}) + S(\bar{\omega})) \leq \delta_1(R) - \delta_2(R) \xrightarrow[R\to+\infty]{} 0.$$

It remains to verify the assertions (3-12)–(3-14) above. We recall that the functions ω_j, $\bar{\omega}$ are radially symmetric and nonincreasing in the radial direction. With a slight abuse of notation, we write $\omega_j(r)$ instead of $\omega_j(x)$ when $r = |x|$, and similarly for $\bar{\omega}$. Accordingly, using (2-41), we obtain the following expressions for the energy of ω_j and $\bar{\omega}$:

$$E(\omega_j) = -\int_0^\infty M_j(r)\log(r)\,r\omega_j(r)\,\mathrm{d}r, \quad E(\bar{\omega}) = -\int_0^\infty \overline{M}(r)\log(r)\,r\,\bar{\omega}(r)\,\mathrm{d}r, \tag{3-15}$$

where

$$M_j(r) = 2\pi\int_0^r s\omega_j(s)\,\mathrm{d}s, \quad \overline{M}(r) = 2\pi\int_0^r s\bar{\omega}(s)\,\mathrm{d}s, \quad r > 0. \tag{3-16}$$

Since $\omega_j \to \bar{\omega}$ in $L^1(\mathbb{R}^2)$, we see that $M_j(r) \to \overline{M}(r)$ uniformly in r as $j \to +\infty$. Moreover, since $\omega_j \in X_M$ satisfies (3-7), the quantity $M_j(r)$ converges to M as $r \to +\infty$ uniformly in j. In particular, we can choose $R \geq 1$ large enough so that $M_j(r) \geq M/2$ for all $j \in \mathbb{N}$ when $r \geq R$.

To prove (3-12), we first take the decomposition

$$E(\omega_{jR}^1) - E(\bar\omega_R^1) = -\int_0^R (M_j(r) - \overline{M}(r)) \log(r)\, r\omega_j(r)\,\mathrm{d}r - \int_0^R \overline{M}(r)\log(r)\, r(\omega_j(r) - \bar\omega(r))\,\mathrm{d}r,$$

and we deduce that

$$\begin{aligned}|E(\omega_{jR}^1) - E(\bar\omega_R^1)| \le \sup_{0\le r\le R} (|M_j(r) - \overline{M}(r)|) \int_0^R |\log(r)|\, r\omega_j(r)\,\mathrm{d}r \\ + \sup_{0\le r\le R} (|\log(r)|\, \overline{M}(r)) \int_0^R r|\omega_j(r) - \bar\omega(r)|\,\mathrm{d}r \xrightarrow[j\to+\infty]{} 0. \qquad (3\text{-}17)\end{aligned}$$

Here we used the convergence of ω_j to $\bar\omega$ in $L^1(\mathbb{R}^2)$, the a priori estimates (3-7), and the fact that $\log(r)\overline{M}(r)$ is bounded as $r \to 0$, as a consequence of (3-8). On the other hand, since the function $-\Phi$ is continuous and bounded from below, and since we integrate on the bounded domain $\{x \in \mathbb{R}^2 : |x| \le R\}$, we can apply Fatou's lemma to obtain

$$-S(\bar\omega_R^1) = \int_{|x|\le R} -\Phi(\bar\omega(x))\,\mathrm{d}x \le \liminf_{j\to+\infty} \int_{|x|\le R} -\Phi(\omega_j(x))\,\mathrm{d}x = -\limsup_{j\to+\infty} S(\omega_{jR}^1). \qquad (3\text{-}18)$$

Combining (3-17) and (3-18), we obtain (3-12).

We next prove (3-13). Recalling that $R \ge 1$, we first observe that

$$E(\omega_{jR}^2) = -\int_R^\infty M_j(r)\log(r)\, r\omega_j(r)\,\mathrm{d}r \le 0,$$

which means that the contribution of $E(\omega_{jR}^2)$ can be disregarded since we only need an upper bound. The other terms in (3-13) have the expressions

$$2E(\omega_{jR}^1, \omega_{jR}^2) = -M_j(R)\int_R^\infty \log(r)\, r\omega_j(r)\,\mathrm{d}r, \qquad S(\omega_{jR}^2) = 2\pi\int_R^\infty \Phi(\omega_j(r))\, r\,\mathrm{d}r.$$

Since ω_j is decreasing, we have $\omega_j(r) \le M_j(r)/(\pi r^2) \le M$ for $r \ge R$. So, using Hypotheses 3.3, we deduce that $\Phi(\omega_j) \le C_1\omega_j + C_2\omega_j \log(M/\omega_j)$, where $C_1 \in \mathbb{R}$ and $C_2 < M/(8\pi)$. It follows that

$$2E(\omega_{jR}^1, \omega_{jR}^2) + S(\omega_{jR}^2) \le 2\pi C_1 \int_R^\infty \omega_j(r) r\,\mathrm{d}r + \int_R^\infty \Delta_j(r)\omega_j(r) r\,\mathrm{d}r, \qquad (3\text{-}19)$$

where

$$\Delta_j(r) = 2\pi C_2 \log\frac{M}{\omega_j(r)} - M_j(R)\log(r).$$

In view of (3-7), the first term in the right-hand side of (3-19) converges to zero uniformly in j as $R \to +\infty$, and can therefore be absorbed in the quantity $\delta_1(R)$. To treat the second term, we fix a positive number $\alpha > 2$ such that $4\pi C_2\alpha \le M$, and we introduce the mutually disjoints sets

$$I(\alpha, R) = \{r \ge R : \omega_j(r) \ge Mr^{-\alpha}\}, \qquad I(\alpha, R)^c = \{r \ge R : \omega_j(r) < Mr^{-\alpha}\}. \qquad (3\text{-}20)$$

As $M_j(R) \ge M/2$, it follows from (3-20) that $\Delta_j(r) \le 0$ when $r \in I(\alpha, R)$, so the last integral in (3-19) can be restricted to the complement $I(\alpha, R)^c$. But on that set we have the upper bound $\omega_j(r) < Mr^{-\alpha}$, where $\alpha > 2$, and we easily deduce that $\int_{I(\alpha,R)^c} \Delta_j(r)\omega_j(r) r\,\mathrm{d}r$ converges to zero as $R \to +\infty$, uniformly in j. Altogether we obtain (3-13).

It remains to establish (3-14), which is an easy task. Indeed $\bar{\omega}$ is a fixed function which satisfies the estimates (3-8), so that $2E(\bar{\omega}_R^1, \bar{\omega}_R^2) + E(\bar{\omega}_R^2) \to 0$ as $R \to +\infty$. In addition, we proved in Proposition 3.5 that the integral defining $S(\bar{\omega})$ is absolutely convergent, and this implies that $S(\bar{\omega}_R^2) \to 0$ as $R \to +\infty$. We thus obtain (3-14), and the proof of Theorem 3.4 is complete. □

Example 3.6. We consider the family of algebraic vortices with parameter $\kappa > 1$:

$$\omega(r) = \frac{1}{(1+r^2)^\kappa}, \qquad M = 2\pi \int_0^\infty r\omega(r)\,\mathrm{d}r = \frac{\pi}{\kappa - 1}.$$

The associated stream function ψ satisfies $\psi(r) = \psi(0) + \int_0^r \psi'(s)\,\mathrm{d}s$, where

$$\psi(0) = \int_0^\infty \log(r)\frac{r}{(1+r^2)^\kappa}\,\mathrm{d}r, \qquad \psi'(r) = \frac{1}{2(\kappa-1)r}\left(1 - \frac{1}{(1+r^2)^{\kappa-1}}\right).$$

We have $\Phi(\omega) = \int_0^\omega \phi(s)\,\mathrm{d}s$, where $\phi(\omega(r)) = \psi(r)$. Explicitly, for a few values of κ, we find

$$\begin{aligned}
\kappa = \frac{3}{2}: &\quad \psi(r) = \log\bigl(1 + \sqrt{1+r^2}\bigr), && \phi(\omega) = \log\left(1 + \frac{1}{\omega^{1/3}}\right),\\
\kappa = 2: &\quad \psi(r) = \frac{1}{4}\log(1+r^2), && \phi(\omega) = \frac{1}{8}\log\frac{1}{\omega},\\
\kappa = 3: &\quad \psi(r) = \frac{1}{8}\left(\log(1+r^2) - \frac{1}{1+r^2}\right), && \phi(\omega) = \frac{1}{24}\log\frac{1}{\omega} - \frac{\omega^{1/3}}{8}.
\end{aligned}$$

In all cases, we observe that

$$\phi(\omega) = \Phi'(\omega) \sim \frac{1}{4\kappa(\kappa-1)}\log\frac{1}{\omega} = \frac{M}{4\pi\kappa}\log\frac{1}{\omega} \quad \text{as } \omega \to 0.$$

It follows that Hypotheses 3.3 are satisfied if and only if $\kappa > 2$.

Example 3.7. We next consider the Gaussian vortex $\omega(r) = e^{-r^2/4}$, where $M = 4\pi$. In that case we have $\psi(0) = \int_0^{+\infty} \log(r)e^{-r^2/4}\,\mathrm{d}r = 2\log(2) - \gamma_E$, so that the stream function satisfies

$$\psi(r) = \psi(0) + \int_0^r \frac{2}{s}(1 - e^{-s^2/4})\,\mathrm{d}s = 2\log(2) - \gamma_E + \mathrm{E}_{\mathrm{in}}\left(\frac{r^2}{4}\right),$$

where

$$\mathrm{E}_{\mathrm{in}}(z) = \int_0^z \frac{1-e^{-t}}{t}\,\mathrm{d}t = \sum_{k=1}^\infty \frac{(-1)^{k-1}}{k}\frac{z^k}{k!}, \qquad z \in \mathbb{C}.$$

We conclude that

$$\phi(\omega) = \Phi'(\omega) = 2\log(2) - \gamma_E + \mathrm{E}_{\mathrm{in}}\left(\log\frac{1}{\omega}\right).$$

In particular $\phi(\omega) \sim \log\log(1/\omega)$ as $\omega \to 0$, and Hypotheses 3.3 are satisfied in that case.

We do not have much information on the maximizer $\bar{\omega}$ whose existence is established in Theorem 3.4. We expect that, if Φ is as in Example 3.7, the maximizer is indeed the Gaussian vortex (2-39), but except for numerical evidence we have no proof so far. Similarly, we believe that the algebraic vortices (2-38) with $\kappa \geq 2$ are global maximizers, but this is known only in the particular case $\kappa = 2$, where maximality follows from the logarithmic HLS inequality (3-5).

The examples above also suggest that the decay rate of the maximizer $\bar\omega(x)$ as $|x|\to\infty$ strongly depends on the behavior of the function $\Phi(s)$ near $s=0$. Extending the techniques in the proof of Theorem 3.4, one should be able to prove that, if Φ is differentiable to the right at the origin, the corresponding maximizer $\bar\omega$ is compactly supported. It is also worth mentioning that the entropy function Φ associated with any radially symmetric decreasing vortex $\bar\omega$ through the relation $\bar\psi(x)=\Phi'(\bar\omega(x))$ is necessarily concave on the range of $\bar\omega$, whereas no concavity assumption is included in Hypotheses 3.3. This suggests that the maximizer $\bar\omega$ corresponding to a nonconcave function Φ should be discontinuous, so that its range does not include the intervals where Φ does not coincide with its concave hull.

4. Stability of viscous vortices

In this final section, we give a new proof of the nonlinear stability of the Oseen vortices, which are self-similar solutions of the Navier–Stokes equations in $\mathbb{R}^2$. Our approach relies on the functional-analytic tools developed in Section 2, in connection with Arnold's variational principle, although we now consider a dissipative equation for which the Casimir functions (1-9) are no longer conserved quantities. Let $w=w(y,\tau)\in\mathbb{R}$ denote the vorticity of the fluid at point $y\in\mathbb{R}^2$ and time $\tau>0$, and let $\phi=\phi(y,\tau)\in\mathbb{R}$ be the associated stream function. The vorticity formulation of the Navier–Stokes equations is

$$\partial_\tau w(y,\tau)+\{\phi,w\}(y,\tau)=\nu\Delta(y,\tau),\qquad \Delta\phi(y,\tau)=w(y,\tau), \tag{4-1}$$

where $\{\phi,w\}=\nabla^\perp\phi\cdot\nabla w$ is the Poisson bracket, $\nu>0$ is the viscosity parameter, and the Laplace operator Δ acts on the space variable $y\in\mathbb{R}^2$. As in [Gallay and Wayne 2002; 2005], we introduce self-similar variables $x=y/\sqrt{\nu\tau}$ and $t=\log(\tau/T)$, where $T>0$ is an arbitrary time scale. More precisely, we look for solutions of (4-1) in the form

$$w(y,\tau)=\frac{1}{\tau}\omega\biggl(\frac{y}{\sqrt{\nu\tau}},\log\frac{\tau}{T}\biggr),\qquad \phi(y,\tau)=\nu\psi\biggl(\frac{y}{\sqrt{\nu\tau}},\log\frac{\tau}{T}\biggr). \tag{4-2}$$

The evolution equation for the rescaled vorticity ω is

$$\partial_t\omega(x,t)+\{\psi,\omega\}(x,t)=\mathcal{L}\omega(x,t),\qquad \Delta\psi(x,t)=\omega(x,t), \tag{4-3}$$

where $\{\psi,\omega\}=\nabla^\perp\psi\cdot\nabla\omega$ and $\mathcal{L}$ is the Fokker–Planck operator

$$\mathcal{L}=\Delta+\tfrac12 x\cdot\nabla+1. \tag{4-4}$$

Let $\bar\omega$ be the vortex with Gaussian profile (2-39), namely

$$\bar\omega(x)=\frac{1}{4\pi}e^{-|x|^2/4},\qquad \bar u(x)=\nabla^\perp\bar\psi(x)=\frac{1}{2\pi}\frac{x^\perp}{|x|^2}(1-e^{-|x|^2/4}). \tag{4-5}$$

It is easy to verify that $\mathcal{L}\bar\omega=0$ and $\{\bar\psi,\bar\omega\}=0$. This implies that $\omega=\alpha\bar\omega$ is a stationary solution of (4-3) for any $\alpha\in\mathbb{R}$. This family of equilibria is known to be stable with respect to perturbations in various weighted L^2 spaces; see [Gallay and Wayne 2005; Gallay 2012]. We present here a new stability proof, which may be easier to adapt to more general situations.

4A. ***Nonlinear stability of Oseen vortices.*** Given any $\alpha \in \mathbb{R}$, we consider solutions of (4-3) of the form $\omega = \alpha\bar{\omega} + \tilde{\omega}$, $\psi = \alpha\bar{\psi} + \tilde{\psi}$. The perturbation $\tilde{\omega}$ satisfies the modified equation

$$\partial_t \tilde{\omega} + \alpha\{\bar{\psi}, \tilde{\omega}\} + \alpha\{\tilde{\psi}, \bar{\omega}\} + \{\tilde{\psi}, \tilde{\omega}\} = \mathcal{L}\tilde{\omega}, \tag{4-6}$$

where it is understood that the stream function $\tilde{\psi}$ is expressed in terms of $\tilde{\omega}$ via the formula (1-14), so that $\Delta\tilde{\psi} = \tilde{\omega}$. We assume henceforth that the perturbation $\tilde{\omega}$ satisfies the moment conditions

$$\int_{\mathbb{R}^2} \tilde{\omega}\,\mathrm{d}x = 0 \quad \text{and} \quad \int_{\mathbb{R}^2} x_j\tilde{\omega}\,\mathrm{d}x = 0 \quad \text{for } j = 1, 2, \tag{4-7}$$

which are preserved under the evolution defined by (4-6). As is shown at the end of [Gallay and Wayne 2005], this hypothesis does not restrict the generality, in the sense that stability with respect to general perturbations (with no moment conditions) can then deduced by a simple argument. As for the existence of solutions to (4-6), we have the following standard result:

Lemma 4.1. *The Cauchy problem for (4-6) is globally well-posed in the weighted L^2 space X defined by (2-4), where $\mathcal{A}(x) = 4|x|^{-2}(e^{|x|^2/4} - 1)$, and the subspace $\mathcal{X}_1 \subset X$ defined by (2-24) is invariant under the evolution.*

Proof. It is known that the vorticity equation (4-3) or (4-6) is globally well-posed in various weighted L^2 spaces; see, e.g., [Gallay and Wayne 2002; Gallay 2012; 2018]. The nearly Gaussian weight $\mathcal{A}$ is not explicitly considered in those references, but the arguments therein can be easily modified to cover that case too. If $\mathcal{A}^{1/2}\tilde{\omega} \in L^2(\mathbb{R}^2)$, then all moments of $\tilde{\omega}$ are well-defined, and a direct calculation shows that the conditions (4-7) are preserved under the evolution, so that (4-6) is globally well-posed in the subspace $\mathcal{X}_1$. □

Let $\tilde{\omega}_0 \in \mathcal{X}_1$, and let $\tilde{\omega} \in C^0([0, +\infty), \mathcal{X}_1)$ be the solution of (4-6) with initial data $\tilde{\omega}_0$. By parabolic regularization, we have $\tilde{\omega}(\,\cdot\,, t) \in Z_1 := Z \cap \mathcal{X}_1$ for all $t > 0$, where Z is the weighted Sobolev space

$$Z = \{\omega \in H^1(\mathbb{R}^2) : \mathcal{A}^{1/2}\omega \in L^2(\mathbb{R}^2),\ \mathcal{A}^{1/2}\nabla\omega \in L^2(\mathbb{R}^2)\}. \tag{4-8}$$

For later use, we introduce the following quadratic form on Z:

$$Q(\omega) = \int_{\mathbb{R}^2} (\mathcal{A}(x)|\nabla\omega(x)|^2 - \mathcal{B}(x)\omega(x)^2)\,\mathrm{d}x, \quad \omega \in Z, \tag{4-9}$$

where

$$\mathcal{B} = 1 + \frac{1}{2}\Big(\Delta\mathcal{A} - \frac{x}{2}\cdot\nabla\mathcal{A} + \mathcal{A}\Big) = 1 + \mathcal{A} - \frac{x\cdot\nabla\mathcal{A}}{|x|^2}. \tag{4-10}$$

We shall verify in Section A3 that $\mathcal{A}/2 \le \mathcal{B} \le 2\mathcal{A}$, so that the form Q is well-defined.

The following coercivity result plays a crucial role in our argument.

Theorem 4.2. *The quadratic form Q defined by (4-9) is coercive on the subspace $Z_1 = Z \cap \mathcal{X}_1$: there exists a constant $\delta > 0$ such that*

$$Q(\omega) \ge \delta \int_{\mathbb{R}^2} \mathcal{A}(x)\omega(x)^2\,\mathrm{d}x \quad \textit{for all } \omega \in Z_1. \tag{4-11}$$

The proof of Theorem 4.2 requires a careful analysis, which is postponed to Section 4B below. In particular, we shall see that the quadratic form Q is not positive on the whole space Z, because it takes negative values on a one-dimensional subspace made of radially symmetric functions. If we restrict ourselves to functions with zero mean, the form Q is nonnegative but vanishes on a two-dimensional subspace due to translation invariance. Therefore, all moment conditions (4-7) are necessary to establish the coercivity of Q.

Returning to the solution $\tilde{\omega} \in C^0([0,+\infty), \mathcal{X}_1)$ of (4-6), we define for all $t > 0$ the quantities

$$\begin{aligned}
\tilde{J}(t) &= \frac{1}{2}\int_{\mathbb{R}^2} \big(\mathcal{A}(x)\tilde{\omega}(x,t)^2 + \tilde{\psi}(x,t)\tilde{\omega}(x,t)\big)\,\mathrm{d}x = J(\tilde{\omega}(t)),\\
\tilde{Q}(t) &= \int_{\mathbb{R}^2} \big(\mathcal{A}(x)|\nabla\tilde{\omega}(x,t)|^2 - \mathcal{B}(x)\tilde{\omega}(x,t)^2\big)\,\mathrm{d}x = Q(\tilde{\omega}(t)),\\
\tilde{N}(t) &= \frac{1}{2}\int_{\mathbb{R}^2} \{\mathcal{A}(x), \tilde{\psi}(x,t)\}\tilde{\omega}(x,t)^2\,\mathrm{d}x =: N(\tilde{\omega}(t)).
\end{aligned} \tag{4-12}$$

The key observation is:

Proposition 4.3. *If $\tilde{\omega} \in C^0([0,+\infty), \mathcal{X}_1)$ is a solution of (4-6), the quantities defined in (4-12) satisfy*

$$\tilde{J}'(t) = -\tilde{Q}(t) - \tilde{N}(t) \quad \textit{for all } t > 0. \tag{4-13}$$

Proof. Using the evolution equation (4-6), we find

$$\begin{aligned}
\tilde{J}'(t) &= \int_{\mathbb{R}^2} \big(\mathcal{A}(x)\tilde{\omega}(x,t) + \tilde{\psi}(x,t)\big)\partial_t\tilde{\omega}(x,t)\,\mathrm{d}x\\
&= \int_{\mathbb{R}^2} (\mathcal{A}\tilde{\omega} + \tilde{\psi})\big(\mathcal{L}\tilde{\omega} - \alpha\{\bar{\psi},\tilde{\omega}\} - \alpha\{\tilde{\psi},\bar{\omega}\} - \{\tilde{\psi},\tilde{\omega}\}\big)(x,t)\,\mathrm{d}x.
\end{aligned} \tag{4-14}$$

We first consider the terms involving the diffusion operator $\mathcal{L}$ in (4-14). We observe that

$$\int_{\mathbb{R}^2} \tilde{\psi}(x,t)\mathcal{L}\tilde{\omega}(x,t)\,\mathrm{d}x = \int_{\mathbb{R}^2} \tilde{\omega}(x,t)^2\,\mathrm{d}x \tag{4-15}$$

because $\mathcal{L}\tilde{\omega} = \Delta\tilde{\omega} + \frac{1}{2}\operatorname{div}(x\tilde{\omega})$ and

$$\begin{aligned}
\int_{\mathbb{R}^2} \tilde{\psi}\,\Delta\tilde{\omega}\,\mathrm{d}x &= \int_{\mathbb{R}^2} (\Delta\tilde{\psi})\tilde{\omega}\,\mathrm{d}x = \int_{\mathbb{R}^2} \tilde{\omega}^2\,\mathrm{d}x,\\
\int_{\mathbb{R}^2} \tilde{\psi}\,\operatorname{div}(x\tilde{\omega})\,\mathrm{d}x &= -\int_{\mathbb{R}^2} (\Delta\tilde{\psi})(x\cdot\nabla\tilde{\psi})\,\mathrm{d}x = \frac{1}{2}\int_{\mathbb{R}^2} \operatorname{div}(x|\nabla\tilde{\psi}|^2)\,\mathrm{d}x = 0.
\end{aligned}$$

On the other hand, integrating by parts we obtain by direct calculation

$$\int_{\mathbb{R}^2} \mathcal{A}(x)\,\tilde{\omega}(x,t)\,\mathcal{L}\tilde{\omega}(x,t)\,\mathrm{d}x = -Q(\tilde{\omega}(t)) - \int_{\mathbb{R}^2} \tilde{\omega}(x,t)^2\,\mathrm{d}x. \tag{4-16}$$

We next compute the advection terms in (4-14), which are proportional to α. We claim that

$$I(\tilde{\omega}) := \int_{\mathbb{R}^2} (\mathcal{A}\tilde{\omega} + \tilde{\psi})(\{\bar{\psi},\tilde{\omega}\} + \{\tilde{\psi},\bar{\omega}\})\,\mathrm{d}x = 0. \tag{4-17}$$

This identity is not surprising, as it means that the quadratic form J is invariant under the evolution defined by the linearized Euler equation at $\bar{\omega}$; see (1-6) for an analogue in the finite-dimensional case. It can also be verified by direct calculations:

$$\int_{\mathbb{R}^2} \mathcal{A}\tilde{\omega}\{\bar{\psi}, \tilde{\omega}\}\,\mathrm{d}x = \frac{1}{2}\int_{\mathbb{R}^2} \mathcal{A}\{\bar{\psi}, \tilde{\omega}^2\}\,\mathrm{d}x = \frac{1}{2}\int_{\mathbb{R}^2} \{\mathcal{A}, \bar{\psi}\}\tilde{\omega}^2\,\mathrm{d}x = 0,$$
$$\int_{\mathbb{R}^2} \tilde{\psi}\{\tilde{\psi}, \bar{\omega}\}\,\mathrm{d}x = \int_{\mathbb{R}^2} \{\tilde{\psi}, \tilde{\psi}\}\bar{\omega}\,\mathrm{d}x = 0,$$
$$\int_{\mathbb{R}^2} (\mathcal{A}\tilde{\omega}\{\tilde{\psi}, \bar{\omega}\} + \tilde{\psi}\{\bar{\psi}, \tilde{\omega}\})\,\mathrm{d}x = \int_{\mathbb{R}^2} \tilde{\omega}(\mathcal{A}\{\tilde{\psi}, \bar{\omega}\} + \{\tilde{\psi}, \bar{\psi}\})\,\mathrm{d}x = 0.$$

Here we used the fact that $\{\mathcal{A}, \bar{\psi}\} = 0$, because $\mathcal{A}$ and $\bar{\psi}$ are radially symmetric. Moreover,

$$\mathcal{A}\{\tilde{\psi}, \bar{\omega}\} + \{\tilde{\psi}, \bar{\psi}\} = (\nabla\tilde{\psi})^{\perp} \cdot (\mathcal{A}\nabla\bar{\omega} + \nabla\bar{\psi}) = 0,$$

by the very definition of $\mathcal{A}$; see (2-3). This proves (4-17).

Finally, integrating by parts the last term in (4-14), we find

$$N(\tilde{\omega}) := \int_{\mathbb{R}^2} (\mathcal{A}\tilde{\omega} + \tilde{\psi})\{\tilde{\psi}, \tilde{\omega}\}\,\mathrm{d}x = \int_{\mathbb{R}^2} \mathcal{A}\tilde{\omega}\{\tilde{\psi}, \tilde{\omega}\}\,\mathrm{d}x = \frac{1}{2}\int_{\mathbb{R}^2} \{\mathcal{A}, \tilde{\psi}\}\tilde{\omega}^2\,\mathrm{d}x. \tag{4-18}$$

Combining (4-14)–(4-18), we obtain the desired result. □

To control the nonlinear term $N(\tilde{\omega})$, we use the following estimate.

Lemma 4.4. *There exists a constant $C_0 > 0$ such that, for all $\tilde{\omega} \in Z$, the nonlinear term (4-18) satisfies*

$$|N(\tilde{\omega})| \le C_0\, \|\mathcal{A}^{1/2}\tilde{\omega}\|_{L^2}^2 (\|\mathcal{A}^{1/2}\tilde{\omega}\|_{L^2} + \|\mathcal{A}^{1/2}\nabla\tilde{\omega}\|_{L^2}). \tag{4-19}$$

Proof. We have $|\{\mathcal{A}, \tilde{\psi}\}| \le C|\nabla\mathcal{A}||\nabla\tilde{\psi}| \le C|x|\mathcal{A}|\nabla\tilde{\psi}|$; hence

$$|N(\tilde{\omega})| \le C\int_{\mathbb{R}^2} |x||\nabla\tilde{\psi}|\mathcal{A}\,\tilde{\omega}^2\,\mathrm{d}x \le C\||x||\nabla\tilde{\psi}|\|_{L^\infty}\|\mathcal{A}^{1/2}\tilde{\omega}\|_{L^2}^2.$$

On the other hand, using Proposition B.1 in [Gallay and Wayne 2002], Hölder's inequality and Sobolev's embedding theorem, we find

$$\||x||\nabla\tilde{\psi}|\|_{L^\infty} \le C(\|\langle x\rangle\tilde{\omega}\|_{L^{3/2}} + \|\langle x\rangle\tilde{\omega}\|_{L^3}) \le C(\|\mathcal{A}^{1/2}\tilde{\omega}\|_{L^2} + \|\mathcal{A}^{1/2}\nabla\tilde{\omega}\|_{L^2}),$$

where $\langle x\rangle = (1 + |x|^2)^{1/2}$. Combining these estimates we arrive at (4-19). □

We are now able to state our final result:

Theorem 4.5. *There exist positive constants C_1, ϵ_0, and μ such that, for any $\alpha \in \mathbb{R}$ and any $\tilde{\omega}_0 \in \mathcal{X}_1$ satisfying $\|\tilde{\omega}_0\|_X \le \epsilon_0$, the solution of (4-6) with initial data $\tilde{\omega}_0$ satisfies*

$$\|\tilde{\omega}(t)\|_X^2 \le C_1\|\tilde{\omega}_0\|_X^2 e^{-\mu t} \quad \textit{for all } t \ge 0. \tag{4-20}$$

Proof. If $\tilde{\omega} \in C^0([0, +\infty), \mathcal{X}_1)$ is the solution of (4-6) with initial data $\tilde{\omega}_0$, we define

$$m_0(t) = \|\tilde{\omega}(t)\|_X^2 = \|\mathcal{A}^{1/2}\tilde{\omega}(t)\|_{L^2}^2 \quad (t \ge 0), \qquad m_1(t) = \|\mathcal{A}^{1/2}\nabla\tilde{\omega}(t)\|_{L^2}^2 \quad (t > 0).$$

For the Gaussian vortex, we proved in Section 2 that Hardy's inequality (2-25) holds for some $C_H < 1$. Thus, by Theorems 2.5 and 2.8, there exists a constant $\gamma \in (0, 1)$ such that

$$\frac{\gamma}{2} m_0(t) \le \tilde{J}(t) \le \frac{1}{2} m_0(t), \quad t \ge 0. \tag{4-21}$$

On the other hand, by Theorem 4.2, there exists $\delta > 0$ such that

$$\tilde{Q}(t) \ge \delta m_0(t) \quad \text{and} \quad \tilde{Q}(t) \ge m_1(t) - 2m_0(t), \quad t > 0, \tag{4-22}$$

where the second inequality follows from the definition (4-9) and the inequality $\mathcal{B} \le 2\mathcal{A}$. Taking a convex combination of both estimates in (4-22), we deduce

$$\tilde{Q}(t) \ge \mu(m_0(t) + m_1(t)), \quad t > 0, \tag{4-23}$$

where $\mu = \delta/(3+\delta)$. Finally, it follows from Lemma 4.4 and Young's inequality that

$$|\tilde{N}(t)| \le C_0 m_0(t)(m_0(t)^{1/2} + m_1(t)^{1/2}) \le \frac{\mu}{4}(m_0(t) + m_1(t)) + \frac{2C_0^2}{\mu} m_0(t)^2. \tag{4-24}$$

Now, as long as $m_0(t) \le \epsilon^2 := \mu^2/(8C_0^2)$, we have by (4-13), (4-21), (4-23), (4-24)

$$\tilde{J}'(t) = -\tilde{Q}(t) - \tilde{N}(t) \le -\frac{\mu}{2}(m_1(t) + m_0(t)) \le -\mu \tilde{J}(t),$$

which implies

$$\gamma m_0(t) \le 2\tilde{J}(t) \le 2\tilde{J}(0)e^{-\mu t} \le m_0(0)e^{-\mu t}.$$

As a consequence, if we assume that $\|\tilde{\omega}_0\|_X^2 = m_0(0) \le \epsilon_0^2 := \gamma\epsilon^2$, we have $m_0(t) \le \epsilon^2$ for all $t \ge 0$ and estimate (4-20) holds with $C_1 = \gamma^{-1}$. □

We briefly indicate here the meaning of our result for the Navier–Stokes equations in the original, unscaled variables. If $\omega = \alpha\bar{\omega} + \tilde{\omega}$, where $\tilde{\omega} \in C^0([0, +\infty), \mathcal{X}_1)$ is as in Theorem 4.5, the vorticity w defined by (4-2) satisfies, in particular, the estimate

$$\int_{\mathbb{R}^2} \left| w(y,\tau) - \frac{\alpha}{4\pi\tau} e^{-|y|^2/(4\tau)} \right| \mathrm{d}y = \mathcal{O}(\tau^{-\mu/2}) \quad \text{as } \tau \to +\infty,$$

which means that $w(\,\cdot\,,\tau)$ converges to a self-similar solution with Gaussian profile as $\tau \to +\infty$. As is shown in [Gallay 2012, Theorem 1.2], that property holds in fact for all solutions of the vorticity equation (4-1) in $L^1(\mathbb{R}^2)$, although it is not possible to specify any decay rate in the general case. Note that the evolution defined by (4-1) in $L^1(\mathbb{R}^2)$ preserves the total mass, so that we necessarily have $\int_{\mathbb{R}^2} w(y,\tau)\,\mathrm{d}y = \alpha$ for all $\tau > 0$.

Remark 4.6. Except for a slight difference in the definition of the function space X, Theorem 4.5 coincides with the well-known stability result [Gallay 2012, Proposition 4.5]. The approach originally developed by C. E. Wayne and the first author relies on conserved quantities related to symmetries of the problem, such as the second-order moment $I(\omega)$ in (1-15). In many respects, it is simpler than ours, and it provides an estimate of the form (4-20) with explicit constants C_1 and μ. Note also that, in the limit of large circulation numbers $|\alpha| \to \infty$, the enhanced dissipation effect due to fast rotation can be used to improve both the decay rate of the perturbations and the size of the basin of attraction of the vortex; see [Gallay 2018].

4B. ***Coercivity of the diffusive quadratic form.*** This section is entirely devoted to the proof of Theorem 4.2, which is a key ingredient in Theorem 4.5. We first observe that the functions $\mathcal{A}(x), \mathcal{B}(x)$ in (4-9) are both radially symmetric, with radial profiles $A(r)$, $B(r)$ given by the explicit expressions

$$A(r) = \frac{e^s - 1}{s}, \quad B(r) = \frac{1}{2s^2}(e^s(1+s) - 1 - 2s) + 1, \quad s = \frac{r^2}{4}. \tag{4-25}$$

On can also verify that B/A is a decreasing function of r satisfying $\frac{1}{2} \le B(r)/A(r) \le \frac{7}{4}$ for all $r > 0$; see Section A3.

We next follow an approach similar to that in Section 2. If $\omega \in Z$ is decomposed in Fourier series like in (2-14), we have

$$Q(\omega) = 2\pi \sum_{k\in\mathbb{Z}} \int_0^\infty \Big\{ A(r)\Big(|\omega_k'(r)|^2 + \frac{k^2}{r^2}\,|\omega_k(r)|^2\Big) - B(r)|\omega_k(r)|^2 \Big\} r\,\mathrm{d}r, \tag{4-26}$$

and we observe that $\omega \in Z_1$ if and only if

$$\int_0^\infty \omega_0(r)\, r\,\mathrm{d}r = 0 \quad \text{and} \quad \int_0^\infty \omega_{\pm 1}(r)\, r^2\,\mathrm{d}r = 0.$$

Introducing the new variables $w_k = A^{1/2}\omega_k \equiv e^\chi \omega_k$, where $\chi = \frac{1}{2}\log(A)$, we obtain after straightforward calculations

$$Q(\omega) = 2\pi \sum_{k\in\mathbb{Z}} \int_0^\infty \Big\{ |w_k'(r)|^2 + \frac{k^2}{r^2}|w_k(r)|^2 + W(r)|w_k(r)|^2 \Big\} r\,\mathrm{d}r, \tag{4-27}$$

where the potential W is defined by

$$W(r) = \chi''(r) + \frac{1}{r}\chi'(r) + \chi'(r)^2 - \frac{B(r)}{A(r)} = \frac{r}{2}\chi'(r) - \chi'(r)^2 - \frac{1}{2} - e^{-2\chi(r)}. \tag{4-28}$$

The coercivity estimate (4-11) is thus equivalent to the inequality

$$\int_0^\infty \Big\{ |w_k'(r)|^2 + \frac{k^2}{r^2}|w_k(r)|^2 + W(r)|w_k(r)|^2 \Big\} r\,\mathrm{d}r \ge \delta \int_0^\infty |w_k(r)|^2\, r\,\mathrm{d}r, \tag{4-29}$$

which should hold for all $k \in \mathbb{Z}$ under the conditions

$$\int_0^\infty w_0(r) e^{-\chi(r)}\, r\,\mathrm{d}r = 0 \quad \text{and} \quad \int_0^\infty w_{\pm 1}(r) e^{-\chi(r)}\, r^2\,\mathrm{d}r = 0. \tag{4-30}$$

For any $k \in \mathbb{Z}$, we denote by L_k the self-adjoint operator in $Y = L^2(\mathbb{R}_+, r\,\mathrm{d}r)$ defined by

$$L_k g = -\frac{1}{r}\partial_r(r\partial_r g) + \frac{k^2}{r^2} g + W g. \tag{4-31}$$

The domain of L_k is exactly the same as for the harmonic oscillator in $\mathbb{R}^2$, because the potential W defined by (4-28) satisfies

$$W(r) > \tfrac{1}{16}r^2 - \tfrac{3}{2} \quad \text{for all } r > 0, \quad \text{and} \quad W(r) \sim \begin{cases} -\frac{3}{2} & \text{as } r \to 0, \\ \frac{1}{16}r^2 & \text{as } r \to \infty; \end{cases} \tag{4-32}$$

see Section A3. Our goal is to prove the lower bound $L_k \geq \delta$ in the entire space Y when $|k| \geq 2$, and in the subspaces given by conditions (4-30) when $k = 0$ or $k = \pm 1$. We consider three cases separately.

Case 1: When $|k| \geq 2$, the desired inequality is simply obtained by comparing L_k with the usual harmonic operator. Indeed, we know from (4-31), (4-32) that

$$L_k > -\partial_r^2 - \frac{1}{r}\partial_r + \frac{k^2}{r^2} + \frac{r^2}{16} - \frac{3}{2} \geq \frac{|k|}{2} - 1, \tag{4-33}$$

where inequalities are between self-adjoint operators on Y. Thus $L_k \geq \frac{1}{2}$ when $|k| \geq 3$, and there exists $\delta > 0$ such that $L_k \geq \delta$ when $|k| = 2$, because the inequality in (4-32) is strict.

Case 2: When $|k| = 1$, the lower bound (4-33) is of no use, but it is easy to verify that $L_k \geq 0$ in that case. Indeed, we claim that $L_k g_1 = 0$, where $g_1(r) = e^{\chi(r)} r e^{-r^2/4}$. Since g_1 is a positive function vanishing at the origin and at infinity, this means that 0 is the lowest eigenvalue of L_k in Y when $k = \pm 1$. To prove the above claim, we first observe that, for any (smooth) function f on $\mathbb{R}_+$, we have the identity

$$\widetilde{L}_k f := e^{\chi} L_k(e^{\chi} f) = -\frac{1}{r}\partial_r(rA\partial_r f) + \frac{k^2}{r^2}Af - Bf, \tag{4-34}$$

because this is the property we used to go from (4-26) to (4-27). On the other hand, in view of (2-2) and (2-3), we have the identity

$$-\frac{1}{r}\partial_r(rA\partial_r\omega_*) = \omega_*, \tag{4-35}$$

which holds in fact for any vorticity profile ω_*, if A is defined by (2-3). In the case of the Lamb–Oseen vortex, if we differentiate the equality (4-35) with respect to r, we find that the function $f = -2\omega_*' = re^{-r^2/4}$ satisfies the relation

$$-\frac{1}{r}\partial_r(rA\partial_r f) + \frac{1}{r^2}Af - \Big(A'' + \frac{2}{r}A' - \frac{r}{2}A'\Big)f = f. \tag{4-36}$$

But $A'' + 2A'/r - rA'/2 = B - 1$ by (4-10), so combining (4-34) and (4-36) we conclude that $\widetilde{L}_k f = 0$ if $|k| = 1$, which is the desired result.

To get coercivity, we now restrict ourselves to the subspace $Y_1 \subset Y$ of all functions g satisfying $\langle g, h_1\rangle = 0$, where $h_1(r) = re^{-\chi(r)}$; see the second relation in (4-30). It is important to observe that h_1 is not proportional to g_1, so that Y_1 is *not* the orthogonal complement in Y of the eigenspace spanned by g_1. However, we have $\langle g_1, h_1\rangle = 8 \neq 0$, which means that the closed hyperplane Y_1 does not contain the eigenfunction g_1. In view of Remark 4.8 below, we conclude that there exists some $\delta > 0$ such that $L_k \geq \delta$ on Y_1 when $|k| = 1$.

Case 3: Finally, we consider the radially symmetric case where $k = 0$. The difficulty here is that the operator L_0 is not positive on the entire space Y. A numerical calculation indicates that L_0 has one negative eigenvalue $\mu_0 \approx -0.722$, and that the next eigenvalue $\mu_1 \approx 0.615$ is positive. So it is essential to use the first relation in (4-30), and to restrict our analysis to the subspace Y_0 of all $g \in Y$ such that $\langle g, h_0\rangle = 0$, where $h_0(r) = e^{-\chi(r)}$. Our strategy is to apply Lemma 4.7 below with $a = -\mu_0$, $b = \mu_1$, $\psi = h_0/\|h_0\|$, and $\phi = g_0/\|g_0\|$, where g_0 denotes a positive function in the kernel of $L_0 - \mu_0$. Estimate (4-41) can

be used to prove coercivity of L_0 on Y_0 if we have good lower bounds on the eigenvalues μ_0, μ_1 and on the scalar product $|\langle\phi, \psi\rangle|$, which measures the angle between the linear spaces spanned by g_0 and h_0.

We first estimate the lowest eigenvalue μ_0. We know from the previous step that $L_1 g_1 = 0$. Defining $g = cg_1/r = ce^{\chi}e^{-r^2/4}$, where $c = (2\log(2))^{-1/2}$ is a normalization factor chosen so that $\|g\| = 1$, we deduce that $L_0 g = (2/r)\partial_r g$. This gives the relation

$$\Big(L_0 + \frac{3}{4}\Big)g = R, \quad \text{where } R = \frac{2}{r}\partial_r g + \frac{3}{4}g = \Big(\frac{3}{4} - \frac{B-1}{A}\Big)g, \tag{4-37}$$

where we used the identity $(B-1)/A = 1 - A'/(rA) = 1 - 2\chi'/r$; see (4-10). In Section A3 below, we show that $B - 1 < \frac{3}{4}A$, so that $R > 0$. This means that the operator $L_0 + \frac{3}{4}$ admits a positive supersolution, and using Sturm–Liouville theory we conclude that $L_0 + \frac{3}{4} > 0$, so that $\mu_0 > -\frac{3}{4}$. Actually the function g is a remarkably accurate quasimode, in the sense that the remainder R in (4-37) is small. The norm of R in $Y = L^2(\mathbb{R}_+, r\,\mathrm{d}r)$ can be computed explicitly; see Section A4. The result is

$$\int_0^\infty R(r)^2 r\,\mathrm{d}r = \frac{1}{16\log(2)}(3 - \log(2) - 2\log(\pi)), \tag{4-38}$$

so that $\epsilon := \|R\|_Y \approx 0.0396$. Since L_0 is a self-adjoint operator, we deduce that L_0 has an eigenvalue in the interval $\big[-\frac{3}{4}, -\frac{3}{4} + \epsilon\big]$. Anticipating the fact (established below) that L_0 has a unique negative eigenvalue, we conclude that $\mu_0 \in \big[-\frac{3}{4}, -\frac{3}{4} + \epsilon\big]$.

We next estimate the second eigenvalue μ_1 of L_0. It is convenient here to observe that, if $g = e^{\chi} f$, the relation $L_0 g = \mu g$ is equivalent to the generalized eigenvalue problem $\widetilde{L}_0 f = \mu A f$, where $\widetilde{L}_k$ is defined in (4-34). The second eigenvalue of that problem is characterized by the inf-sup formula

$$\mu_1 = \inf_{f\in\mathcal{F}} \sup_{r>0} (\mathcal{R}[f])(r) = \sup_{f\in\mathcal{F}} \inf_{r>0} (\mathcal{R}[f])(r), \quad \text{where } \mathcal{R}[f] = \frac{\widetilde{L}_0 f}{Af}. \tag{4-39}$$

Here $\mathcal{F}$ denotes the class of all (smooth) functions $f : [0, +\infty) \to \mathbb{R}$ such that $f(0) = 1$, $f(r) \to 0$ as $r \to +\infty$, and f has exactly one zero in the interval $(0, +\infty)$. Our first trial function is $f(r) = e^{-s}(1-\alpha s)$, where $s = r^2/4$ and $\alpha = \log(2)^{-1}$. The value of α is chosen so that the Rayleigh quotient has no singularity:

$$\mathcal{R}[f] = \frac{e^{-s}(1 + (2-\alpha)s + 2\alpha s^2) - (1 + (1-\alpha)s + \alpha s^2)}{2s(1 - e^{-s})(1 - \alpha s)}, \quad s = \frac{r^2}{4}.$$

It happens that $\mathcal{R}[f]$ is a decreasing function on $\mathbb{R}_+$, with $\mathcal{R}[f](0) = -\frac{3}{4} + \alpha$ and $\mathcal{R}[f](+\infty) = \frac{1}{2}$. In view of (4-39), this implies that $\frac{1}{2} < \mu_1 < -\frac{3}{4} + \alpha \approx 0.69$. A better approximation is obtained using

$$f(r) = e^{-s}(1 - \alpha s)(1 + \beta s), \quad \text{where } \beta = \frac{\alpha(1 - 2e^{-1/\alpha})}{2\alpha - 1 + 2e^{-1/\alpha}(1-\alpha)}.$$

If $\frac{1}{2} < \alpha < \log(2)^{-1}$, then $\beta > 0$ and the Rayleigh quotient has no singularity in the interval $(0, +\infty)$. Taking for instance $\alpha = 1.4$ gives the excellent lower bound $\mu_1 \geq 0.6$.

Finally, we use the quasimode g in (4-37) and a standard perturbation argument to estimate the true eigenfunction corresponding to the lowest eigenvalue μ_0. We first look for a nonnormalized eigenfunction

of the form $g_0 = g - f$, where $f \perp g_0$. We have

$$0 = (L_0 - \mu_0)g_0 = (L_0 - \mu_0)g - (L_0 - \mu_0)f = R - \big(\mu_0 + \tfrac{3}{4}\big)g - (L_0 - \mu_0)f,$$

so that $f = (L_0 - \mu_0)^{-1}\big(R - \big(\mu_0 + \frac{3}{4}\big)g\big)$, where $(L_0 - \mu_0)^{-1}$ denotes the partial inverse of $L_0 - \mu_0$ on its range. The norm of that inverse is bounded by $1/d$, where $d = \mu_1 - \mu_0$ is the spectral gap. As $\|R\| = \epsilon$ and $|\mu_0 + \frac{3}{4}| \le \epsilon$, we conclude that $\|f\| \le 2\epsilon/d$. The normalized eigenfunction is

$$\phi = \frac{g_0}{\|g_0\|} = \frac{g - f}{\sqrt{1 - \|f\|^2}}.$$

Let $\psi = \hat{c}h_0 = \hat{c}e^{-\chi}$, where $\hat{c} = \sqrt{3}/\pi$ is a normalization factor chosen so that $\|\psi\| = 1$. A direct calculation shows

$$\langle \psi, g\rangle = c\hat{c}\int_0^\infty e^{-r^2/4}\, r\,\mathrm{d}r = 2c\hat{c} = \frac{1}{\pi}\sqrt{\frac{6}{\log(2)}} \approx 0.9365;$$

hence

$$\langle \psi, \phi\rangle = \frac{\langle \psi, g\rangle - \langle \psi, f\rangle}{\sqrt{1 - \|f\|^2}} \ge 2c\hat{c} - \frac{2\epsilon}{d}. \tag{4-40}$$

We use Lemma 4.7 below with $a = -\mu_0 \le \frac{3}{4}$, $d = a + b = \mu_1 - \mu_0 \ge 1.2$, and $\epsilon = \|R\| \le 0.04$. In view of (4-40), estimate (4-41) shows that there exists some $\delta > 0$ such that $\langle Lf, f\rangle \ge \delta\|f\|^2$ for all $f \in Y_0 = h_0^\perp$. This concludes the proof of Theorem 4.5. □

Finally, we state an elementary lemma that was used twice in the above proof.

Lemma 4.7. *Let X be a Hilbert space and $L : D(L) \to X$ be a self-adjoint operator in X. We assume that there exist $\phi \in D(L)$ with $\|\phi\| = 1$ and $a, b \in \mathbb{R}$ with $a + b \ge 0$ such that*

(i) *$L\phi = -a\phi$, and*

(ii) *$\langle Lg, g\rangle \ge b\|g\|^2$ for all $g \in D(L)$ with $g \perp \phi$.*

Then, for any $\psi \in X$ with $\|\psi\| = 1$, we have the lower bound

$$\langle Lf, f\rangle \ge ((a + b)|\langle \phi, \psi\rangle|^2 - a)\|f\|^2 \quad \text{for all } f \in D(L) \text{ with } f \perp \psi. \tag{4-41}$$

Proof. Given $f \in D(L)$, we take the decomposition $f = \langle f, \phi\rangle\phi + g$, so that $g \perp \phi$. Since $L\phi = -a\phi$, we find

$$\langle Lf, f\rangle = \langle Lg, g\rangle - a|\langle f, \phi\rangle|^2 \ge b\|g\|^2 - a|\langle f, \phi\rangle|^2 = b\|f\|^2 - (a + b)|\langle f, \phi\rangle|^2,$$

where the inequality follows from (ii). We now assume that $f \perp \psi$ and take the decomposition $\phi = \langle \phi, \psi\rangle\psi + h$. By Cauchy–Schwarz, we have

$$|\langle f, \phi\rangle|^2 = |\langle f, h\rangle|^2 \le \|f\|^2\|h\|^2 = \|f\|^2(1 - |\langle \phi, \psi\rangle|^2),$$

and combining both inequalities we arrive at (4-41). □

Remark 4.8. In the particular case where $a = 0$ and $b > 0$, the kernel of L is one-dimensional, and inequality (4-41) implies that the quadratic form of L is strictly positive on any closed hyperplane that does not contain the eigenfunction ϕ.

Appendix

A1. ***Integral inequalities involving logarithmic weights.***

Proof of Proposition 3.1. Let $B_1 = \{x \in \mathbb{R}^n : |x| < 1\}$ and $D_M = \{x \in \mathbb{R}^n : f(x) < M\}$. To prove (3-1), we must verify that

$$\int_{B_1} \left(\log \frac{1}{|x|}\right) f(x)\,\mathrm{d}x \lesssim M + \int_{\mathbb{R}^n \setminus D_M} \left(\log \frac{f(x)}{M}\right) f(x)\,\mathrm{d}x. \tag{A-1}$$

Let $\Omega_1 = \{x \in B_1 : f(x) \le M|x|^{-n/2}\}$ and $\Omega_2 = \{x \in B_1 : f(x) > M|x|^{-n/2}\} \subset \mathbb{R}^n \setminus D_M$. We have $B_1 = \Omega_1 \cup \Omega_2$ and

$$\int_{\Omega_1} \left(\log \frac{1}{|x|}\right) f(x)\,\mathrm{d}x \le M \int_{B_1} \frac{1}{|x|^{n/2}} \log \frac{1}{|x|}\,\mathrm{d}x = CM,$$

$$\int_{\Omega_2} \left(\log \frac{1}{|x|}\right) f(x)\,\mathrm{d}x \le \frac{2}{n} \int_{\Omega_2} \left(\log \frac{f(x)}{M}\right) f(x)\,\mathrm{d}x \le \frac{2}{n} \int_{\mathbb{R}^n \setminus D_M} \left(\log \frac{f(x)}{M}\right) f(x)\,\mathrm{d}x;$$

hence (A-1) follows by adding both inequalities. We next consider (3-2), which reads

$$\int_{D_M} \left(\log \frac{M}{f(x)}\right) f(x)\,\mathrm{d}x \lesssim M + \int_{\mathbb{R}^n \setminus B_1} (\log |x|) f(x)\,\mathrm{d}x. \tag{A-2}$$

Let $e = \exp(1)$ and

$$\Omega_3 = \left\{x \in D_M : f(x) \le \frac{M}{e(1+|x|)^{2n}}\right\}, \quad \Omega_4 = \left\{x \in D_M : f(x) > \frac{M}{e(1+|x|)^{2n}}\right\}.$$

Since $t \mapsto t \log(1/t)$ is increasing on $[0, e^{-1}]$ and $s \mapsto \log(s)$ is increasing on $[1, +\infty)$, we have

$$\int_{\Omega_3} \left(\log \frac{M}{f(x)}\right) f(x)\,\mathrm{d}x \le M \int_{\mathbb{R}^n} \frac{1}{e(1+|x|)^{2n}} \log(e(1+|x|)^{2n})\,\mathrm{d}x = CM,$$

$$\int_{\Omega_4} \left(\log \frac{M}{f(x)}\right) f(x)\,\mathrm{d}x \le \int_{\Omega_4} \log(e(1+|x|)^{2n}) f(x)\,\mathrm{d}x \le CM + 2n \int_{\mathbb{R}^n \setminus B_1} (\log |x|) f(x)\,\mathrm{d}x,$$

and (A-2) follows in the same way.

From now on, we assume that f is radially symmetric and nonincreasing in the radial direction. In particular, we have, for all $x \neq 0$,

$$f(x) \le \frac{1}{\alpha_n |x|^n} \int_{|y| \le |x|} f(y)\,\mathrm{d}y \le \frac{M}{\alpha_n |x|^n}, \quad \text{where } \alpha_n = \frac{\pi^{n/2}}{\Gamma(1+n/2)}. \tag{A-3}$$

Since $t \mapsto \log_+(t)$ is increasing, we deduce that

$$\int_{\mathbb{R}^n \setminus D_M} \left(\log \frac{f(x)}{M}\right) f(x)\,\mathrm{d}x \le \int_{\mathbb{R}^n} \left(\log_+ \frac{1}{\alpha_n |x|^n}\right) f(x)\,\mathrm{d}x \le CM + n \int_{B_1} \left(\log \frac{1}{|x|}\right) f(x)\,\mathrm{d}x,$$

which is the converse of (3-1). Note that, when $n \le 12$, the first term CM in the right-hand side can be dropped, because $\alpha_n > 1$. In a similar way, we find

$$\int_{\mathbb{R}^n \setminus B_1} (\log |x|) f(x)\,\mathrm{d}x \le \frac{1}{n} \int_{\mathbb{R}^n} \left(\log_+ \frac{M}{\alpha_n f(x)}\right) f(x)\,\mathrm{d}x \le CM + \int_{D_M} \left(\log \frac{M}{f(x)}\right) f(x)\,\mathrm{d}x,$$

which is the converse of (3-2). Again, the first term CM in the right-hand side can be dropped when $n \le 12$. □

Proof of Proposition 2.2. Throughout the proof we assume that $M := \|\omega\|_{L^1} > 0$. We take the decomposition $E(\omega) = E_1(\omega) + E_2(\omega)$, where

$$E_i(\omega) = \frac{1}{4\pi} \int_{\Omega_i} \log \frac{1}{|x-y|} \omega(x)\omega(y)\,\mathrm{d}x\,\mathrm{d}y, \quad i = 1, 2,$$

and $\Omega_1 = \{(x, y) \in \mathbb{R}^2 \times \mathbb{R}^2 : |x - y| < 1\}$, $\Omega_2 = \{(x, y) \in \mathbb{R}^2 \times \mathbb{R}^2 : |x - y| \ge 1\}$. We have to verify that the integrals defining the quantities E_1, E_2 are convergent under assumptions (2-6).

First of all, using inequality (A-1) above with $n = 2$, we obtain for all $x \in \mathbb{R}^2$

$$\int_{|y-x|<1} \log \frac{1}{|x-y|} |\omega(y)|\,\mathrm{d}y \le C \int_{\mathbb{R}^2} \left(1 + \log_+ \frac{|\omega(y)|}{M}\right) |\omega(y)|\,\mathrm{d}y.$$

If we multiply both sides by $|\omega(x)|$ and integrate over $x \in \mathbb{R}^2$, we thus find

$$|E_1(\omega)| \le CM \int_{\mathbb{R}^2} \left(1 + \log_+ \frac{|\omega(y)|}{M}\right) |\omega(y)|\,\mathrm{d}y. \tag{A-4}$$

On the other hand, we have $\log|x - y| \le \log(|x| + |y|) \le \log(1 + |x|) + \log(1 + |y|)$ when $|x - y| \ge 1$. This gives the bound

$$\begin{aligned} |E_2(\omega)| &\le \frac{1}{4\pi} \int_{\Omega_2} |\omega(x)|\,|\omega(y)| \big(\log(1 + |x|) + \log(1 + |y|)\big)\,\mathrm{d}x\,\mathrm{d}y \\ &\le \frac{M}{2\pi} \int_{\mathbb{R}^2} |\omega(y)| \log(1 + |y|)\,\mathrm{d}y. \end{aligned} \tag{A-5}$$

Combining (A-4) and (A-5), we arrive at (2-7).

Finally, we assume that $\omega \in C_c^2(\mathbb{R}^2)$ and $\int_{\mathbb{R}^2} \omega(x)\,\mathrm{d}x = 0$. The associated stream function $\psi \in C^2(\mathbb{R})$ defined by (1-14) satisfies $|\psi(x)| = \mathcal{O}(|x|^{-1})$ and $|u(x)| = |\nabla\psi(x)| = \mathcal{O}(|x|^{-2})$ as $|x| \to \infty$, so that $u \in L^2(\mathbb{R}^2)$. This allows us to integrate by parts in the first expression (1-13) of the energy, using the relation $\Delta\psi = \omega$, to obtain the elegant formula $E(\omega) = \frac{1}{2} \int_{\mathbb{R}^2} |u|^2\,\mathrm{d}x$. By a density argument, the conclusion remains valid for all integrable vorticities with zero average satisfying a assumptions (2-6). □

A2. ***Positivity of the potential V in some examples.*** For the algebraic vortex (2-38) with $\kappa = 1 + \nu > 1$, the potential V defined in (2-34) has the expression

$$V(r) = \frac{1}{r^2(1+r^2)^2} \big(3 - 2(\nu-1)r^2 + (\nu^2-1)r^4 - 2S - S^2\big), \quad \text{where } S = \frac{\nu r^2}{(1+r^2)^\nu - 1}.$$

If $\nu = 1$, then $S = 1$; hence $V \equiv 0$. Otherwise:

- If $\nu > 1$, we have $(1 + r^2)^\nu > 1 + \nu r^2$ for all $r > 0$, so that $S < 1$. We deduce

$$V(r) > \frac{\nu - 1}{(1+r^2)^2} (-2 + (\nu+1)r^2), \tag{A-6}$$

so that $V(r) > 0$ if $r^2 \geq 2/(\nu+1)$. In the region where $r^2 \leq 2/(\nu+1)$, we use the improved estimate

$$S = \frac{\nu r^2}{(1+r^2)^\nu - 1} < 1 - \frac{\nu-1}{2} r^2 + \frac{\nu^2-1}{12} r^4, \quad r > 0, \tag{A-7}$$

which can be established by a direct calculation. This gives the lower bound

$$V(r) > \frac{(\nu-1)r^2}{12(1+r^2)^2}\left(5\nu + 11 + (\nu^2-1)r^2 - \frac{(\nu-1)(\nu+1)^2}{12} r^4\right), \tag{A-8}$$

which implies that $V(r) > 0$ if $(\nu+1)r^2 \leq 2$.

- If $0 < \nu < 1$, the calculations are entirely similar, except that all inequalities in (A-6)–(A-8) are reversed. This shows that $V(r) < 0$ in that case.

For the Gaussian vortex (2-39), a direct calculation shows that

$$V(r) = \frac{3}{4s} - \frac{1}{2} + \frac{s}{4} - \frac{1/2}{e^s - 1} - \frac{s/4}{(e^s-1)^2}, \quad \text{where } s = \frac{r^2}{4}.$$

Using the lower bound $e^s - 1 \geq s(1 + s/2 + s^2/6)$, we obtain

$$\begin{aligned} V(r) &\geq \frac{1}{4s}\frac{1}{(1+s/2+s^2/6)^2}\left((3-2s+s^2)\left(1+\frac{s}{2}+\frac{s^2}{6}\right)^2 - 2\left(1+\frac{s}{2}+\frac{s^2}{6}\right) - 1\right) \\ &= \frac{1}{4}\frac{s}{(6+3s+s^2)^2}(15 + 12s + 12s^2 + 4s^3 + s^4) > 0. \end{aligned}$$

A3. ***Properties of the Gaussian vortex.*** Given the expressions of A, B in (4-25), we first verify that the ratio B/A is a decreasing function of r. We have

$$\frac{B(r)-1}{A(r)} = \frac{1}{2}\left(1 + h\left(\frac{r^2}{4}\right)\right), \quad \text{where } h(s) = \frac{1}{s} - \frac{1}{e^s - 1}. \tag{A-9}$$

Since

$$h'(s) = -\frac{(e^s-1)^2 - s^2 e^s}{s^2(e^s-1)^2} = -4e^s \frac{\sinh(s/2)^2 - (s/2)^2}{s^2(e^s-1)^2} < 0, \quad s > 0,$$

we see that h is strictly decreasing on $(0, +\infty)$ with $h(0) = \frac{1}{2}$ and $h(+\infty) = 0$. We conclude that $(B-1)/A$, hence also B/A, is a decreasing function of r, and that $\frac{1}{2} \leq B/A \leq \frac{7}{4}$.

We next prove the lower bound (4-32) on the potential W. Since $\chi = \log(A)/2$ with A as in (4-25), a direct calculation shows that the potential W defined by (4-28) has the expression

$$W(r) = \frac{s}{4} - \frac{1}{2} - \frac{1}{4s} - \frac{s - 1/2}{e^s - 1} - \frac{s/4}{(e^s-1)^2}, \quad \text{where } s = \frac{r^2}{4}.$$

Inequality (4-32) is thus equivalent to the positivity of the function G defined by

$$G(s) = 1 - \frac{1}{4s} - \frac{s-1/2}{e^s-1} - \frac{s/4}{(e^s-1)^2}, \quad s > 0. \tag{A-10}$$

If $s \geq \frac{1}{2}$, we use the lower bound $e^s - 1 \geq s(1+s/2)$ and obtain

$$G(s) \geq \frac{s}{4(2+s)^2}(7+4s) > 0.$$

If $0 < s < \frac{1}{2}$, the third term in the right-hand side of (A-10) has the opposite sign. To estimate the denominator, we use the upper bound $e^s - 1 \leq s(1+s/2)(1+s^2/5)$, which holds for $s \leq \frac{1}{2}$. This gives

$$G(s) \geq \frac{s}{4(2+s)^2(5+s^2)}(27+32s+15s^2+4s^3) > 0.$$

A4. ***Computing the norm of the quasimode (4-37).*** In this section we compute the L^2 norm of the function R defined by (4-37). We recall that $g = cA^{1/2}e^{-r^2/4}$, where $c = (2\log(2))^{-1/2}$, and using (A-9) we observe that

$$\frac{3}{4} - \frac{B-1}{A} = \frac{1}{4}\left(1 - 2h\left(\frac{r^2}{4}\right)\right), \quad \text{where } h(s) = \frac{1}{s} - \frac{1}{e^s - 1}.$$

It follows that

$$\|R\|_Y^2 = \frac{1}{16}\int_0^\infty \left(1 - 2h\left(\frac{r^2}{4}\right)\right)^2 g(r)^2 r\,\mathrm{d}r = \frac{1}{16\log(2)}\int_0^\infty (1-2h(s))^2 \frac{1}{s}(e^{-s} - e^{-2s})\,\mathrm{d}s.$$

Expanding $(1-2h(s))^2 = 1 - 4h(s) + 4h(s)^2$, we take the decomposition

$$\|R\|_Y^2 \equiv \int_0^\infty R(r)^2 r\,\mathrm{d}r = \frac{I_1 - 4I_2 + 4I_3}{16\log(2)}, \tag{A-11}$$

where the integrals I_1, I_2, I_3 are defined and computed below.

- Evaluation of I_1: $$I_1 = \int_0^\infty \frac{1}{s}(e^{-s} - e^{-2s})\,\mathrm{d}s = \log(2).$$
- Evaluation of I_2:
$$\begin{aligned} I_2 &= \int_0^\infty \frac{h(s)}{s}(e^{-s} - e^{-2s})\,\mathrm{d}s \\ &= \int_0^\infty \left(\frac{1}{s} - \frac{1}{e^s - 1}\right)\left\{\int_0^\infty e^{-st}\,\mathrm{d}t\right\}(e^{-s} - e^{-2s})\,\mathrm{d}s \\ &= \int_0^\infty \left\{\int_0^\infty \left(\frac{1}{s} - \frac{1}{e^s - 1}\right)(e^{-s(1+t)} - e^{-s(2+t)})\,\mathrm{d}s\right\}\mathrm{d}t \\ &= \int_0^\infty \left(\log\frac{2+t}{1+t} - \frac{1}{2+t}\right)\mathrm{d}t = (1+t)\log\frac{2+t}{1+t}\bigg|_{t=0}^{t=+\infty} = 1 - \log(2). \end{aligned}$$
- Evaluation of I_3:
$$\begin{aligned} I_3 &= \int_0^\infty \frac{h(s)^2}{s}(e^{-s} - e^{-2s})\,\mathrm{d}s \\ &= \int_0^\infty \left(\frac{1}{s} - \frac{1}{e^s - 1}\right)^2\left\{\int_0^\infty tse^{-st}\,\mathrm{d}t\right\}(e^{-s} - e^{-2s})\,\mathrm{d}s \\ &= \int_0^\infty \left\{\int_0^\infty \left(\frac{e^{-s(1+t)} - e^{-s(2+t)}}{s} - 2e^{-s(2+t)} + \frac{se^{-s(2+t)}}{e^s - 1}\right)\mathrm{d}s\right\} t\,\mathrm{d}t \\ &= \int_0^\infty \left(\log\frac{2+t}{1+t} - \frac{2}{2+t} + \psi^{(1)}(3+t)\right) t\,\mathrm{d}t, \end{aligned}$$

where $\psi^{(1)}$ denotes the trigamma function [Abramowitz and Stegun 1966, Section 6.4]:

$$\psi^{(1)}(z) = \int_0^\infty \frac{se^{-sz}}{1-e^{-s}}\,\mathrm{d}s = \frac{\mathrm{d}^2}{\mathrm{d}z^2}\log\Gamma(z), \qquad \mathrm{Re}(z) > 0.$$

It follows that

$$\begin{aligned} I_3 &= \frac{t^2+4}{2}\log(2+t) - \frac{t^2-1}{2}\log(1+t) - \frac{3t}{2} + t(\log\Gamma)'(3+t) - (\log\Gamma)(3+t)\Big|_{t=0}^{t=+\infty} \\ &= \frac{1}{4}(7 - 6\log(2) - 2\log(\pi)). \end{aligned}$$

Here we used Stirling's formula to compute an asymptotic expansion of $(\log\Gamma)(3+t)$ and its derivative as $t \to +\infty$. Inserting the values of I_1, I_2, I_3 into (A-11), we arrive at (4-38).

A5. ***The Poisson structure on*** $\mathcal{P}$**.** For two functions ϕ, ψ on $\mathbb{R}^2$ we use the familiar notation $\{\phi, \psi\} = \partial_1\phi\partial_2\psi - \partial_2\phi\partial_1\psi$. Now, if $\mathcal{F}$ and $\mathcal{G}$ are two functionals of $\mathcal{P}$, the standard way of defining their Poisson bracket is

$$\{\mathcal{F}, \mathcal{G}\}(\omega) = -\int_{\mathbb{R}^2} \omega\left\{\frac{\delta\mathcal{F}}{\delta\omega}, \frac{\delta\mathcal{G}}{\delta\omega}\right\}\mathrm{d}x, \tag{A-12}$$

where $\delta\mathcal{F}/\delta\omega$ is the usual "variational derivative" of $\mathcal{F}$, namely, the function on $\mathbb{R}^2$ defined by the relation

$$\left(\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{F}(\omega+\epsilon\eta)\right)\Big|_{\epsilon=0} = \int_{\mathbb{R}^2} \frac{\delta\mathcal{F}}{\delta\omega}(x)\eta(x)\,\mathrm{d}x$$

for all (smooth and compactly supported) increments η. In particular, the variational derivative of the energy functional (1-13) is $\delta E/\delta\omega = -\psi$, where ψ is the stream function (1-14). As an application, if ω evolves according to the Euler equation $\partial_t\omega + \{\psi, \omega\} = 0$, we have for any (smooth) functional $\mathcal{F}$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\omega) = -\int_{\mathbb{R}^2} \frac{\delta\mathcal{F}}{\delta\omega}\{\psi, \omega\}\,\mathrm{d}x = \int_{\mathbb{R}^2}\left\{\frac{\delta\mathcal{F}}{\delta\omega}, \frac{\delta E}{\delta\omega}\right\}\omega\,\mathrm{d}x = \{E, \mathcal{F}\}(\omega).$$

This is precisely the integrated form of the canonical equation $\partial_t\omega = \{E, \omega\}$.

Acknowledgments

Gallay is partially supported by the grant SingFlows ANR-18-CE40-0027 of the French National Research Agency (ANR). The research of Šverák is supported in part by grant DMS 1956092 from the National Science Foundation.

References

[Abramowitz and Stegun 1966] M. Abramowitz and I. A. Stegun (editors), *Handbook of mathematical functions, with formulas, graphs, and mathematical tables*, Dover, New York, 1966. MR Zbl

[Arnold 1965] V. I. Arnold, "On conditions for non-linear stability of plane stationary curvilinear flows of an ideal fluid", *Dokl. Akad. Nauk SSSR* **162**:5 (1965), 975–978. In Russian; translated in *Soviet Math. Dokl.* **6** (1965), 773–777. MR Zbl

[Arnold 1966a] V. Arnold, "Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits", *Ann. Inst. Fourier* (*Grenoble*) **16**:1 (1966), 319–361. MR Zbl

[Arnold 1966b] V. I. Arnold, "Sur un principe variationnel pour les écoulements stationnaires des liquides parfaits et ses applications aux problèmes de stabilité non linéaire", *J. Mécanique* **5** (1966), 29–43. Zbl

[Arnold and Khesin 1998] V. I. Arnold and B. A. Khesin, *Topological methods in hydrodynamics*, Appl. Math. Sci. **125**, Springer, 1998. MR Zbl

[Bedrossian et al. 2019] J. Bedrossian, M. Coti Zelati, and V. Vicol, "Vortex axisymmetrization, inviscid damping, and vorticity depletion in the linearized 2D Euler equations", *Ann. PDE* **5**:1 (2019), art. id. 4. MR Zbl

[Bianchi and Egnell 1991] G. Bianchi and H. Egnell, "A note on the Sobolev inequality", *J. Funct. Anal.* **100**:1 (1991), 18–24. MR Zbl

[Burton 2005] G. R. Burton, "Global nonlinear stability for steady ideal fluid flow in bounded planar domains", *Arch. Ration. Mech. Anal.* **176**:2 (2005), 149–163. MR Zbl

[Cao et al. 2019] D. Cao, J. Wan, and G. Wang, "Nonlinear orbital stability for planar vortex patches", *Proc. Amer. Math. Soc.* **147**:2 (2019), 775–784. MR Zbl

[Carlen and Loss 1992] E. Carlen and M. Loss, "Competing symmetries, the logarithmic HLS inequality and Onofri's inequality on S^n", *Geom. Funct. Anal.* **2**:1 (1992), 90–104. MR Zbl

[Coddington and Levinson 1955] E. A. Coddington and N. Levinson, *Theory of ordinary differential equations*, McGraw-Hill, New York, 1955. MR Zbl

[Deimling 1985] K. Deimling, *Nonlinear functional analysis*, Springer, 1985. MR Zbl

[Fusco et al. 2008] N. Fusco, F. Maggi, and A. Pratelli, "The sharp quantitative isoperimetric inequality", *Ann. of Math.* (2) **168**:3 (2008), 941–980. MR Zbl

[Gallay 2012] T. Gallay, "Stability and interaction of vortices in two-dimensional viscous flows", *Discrete Contin. Dyn. Syst. Ser. S* **5**:6 (2012), 1091–1131. MR Zbl

[Gallay 2018] T. Gallay, "Enhanced dissipation and axisymmetrization of two-dimensional viscous vortices", *Arch. Ration. Mech. Anal.* **230**:3 (2018), 939–975. MR Zbl

[Gallay and Wayne 2002] T. Gallay and C. E. Wayne, "Invariant manifolds and the long-time asymptotics of the Navier–Stokes and vorticity equations on $\mathbb{R}^2$", *Arch. Ration. Mech. Anal.* **163**:3 (2002), 209–258. MR Zbl

[Gallay and Wayne 2005] T. Gallay and C. E. Wayne, "Global stability of vortex solutions of the two-dimensional Navier–Stokes equation", *Comm. Math. Phys.* **255**:1 (2005), 97–129. MR Zbl

[Hartman 1964] P. Hartman, *Ordinary differential equations*, Wiley, New York, 1964. MR Zbl

[Kato 1966] T. Kato, *Perturbation theory for linear operators*, Grundlehren der Math. Wissenschaften **132**, Springer, 1966. MR Zbl

[Lieb and Loss 1997] E. H. Lieb and M. Loss, *Analysis*, Grad. Stud. in Math. **14**, Amer. Math. Soc., Providence, RI, 1997. MR Zbl

[Marchioro and Pulvirenti 1994] C. Marchioro and M. Pulvirenti, *Mathematical theory of incompressible nonviscous fluids*, Appl. Math. Sci. **96**, Springer, 1994. MR Zbl

[Mazya 2011] V. Mazya, *Sobolev spaces: with applications to elliptic partial differential equations*, 2nd ed., Grundlehren der Math. Wissenschaften **342**, Springer, 2011. MR Zbl

[Onofri 1982] E. Onofri, "On the positivity of the effective action in a theory of random surfaces", *Comm. Math. Phys.* **86**:3 (1982), 321–326. MR Zbl

[Onsager 1949] L. Onsager, "Statistical hydrodynamics", *Nuovo Cimento* (9) **6**:suppl. 2 (1949), 279–287. MR

[Reed and Simon 1978] M. Reed and B. Simon, *Methods of modern mathematical physics, IV: Analysis of operators*, Academic Press, New York, 1978. MR Zbl

[Rudin 1953] W. Rudin, *Principles of mathematical analysis*, McGraw-Hill, New York, 1953. MR Zbl

Received 14 Nov 2021. Revised 1 Jun 2022. Accepted 11 Jul 2022.

THIERRY GALLAY: thierry.gallay@univ-grenoble-alpes.fr
Institut Fourier, Université Grenoble Alpes, Gières, France

VLADIMÍR ŠVERÁK: sverak@math.umn.edu
School of Mathematics, University of Minnesota, Minneapolis, MN, United States

ANALYSIS AND PDE
Vol. 17 (2024), No. 2, pp. 723–748
DOI: 10.2140/apde.2024.17.723

EXPLICIT FORMULA OF RADIATION FIELDS OF FREE WAVES WITH APPLICATIONS ON CHANNEL OF ENERGY

LIANG LI, RUIPENG SHEN AND LIJUAN WEI

We give a few explicit formulas regarding the radiation fields of linear free waves. We then apply these formulas on the channel-of-energy theory. We characterize all the radial weakly nonradiative solutions in all dimensions and give a few new exterior energy estimates.

1. Introduction

1A. ***Background and topics.*** The semilinear wave equation

$$\partial_t^2 u - \Delta u = \pm |u|^{p-1} u, \quad (x,t) \in \mathbb{R}^d \times \mathbb{R},$$

especially the energy critical case $p = 1 + 4/(d-2)$, has been extensively studied by many mathematicians in the past few decades. Please see, for example, [Kapitanski 1994; Lindblad and Sogge 1995] for local existence and well-posedness, and [Ginibre, Soffer and Velo 1992; Grillakis 1990; 1992; Kenig and Merle 2008; Nakanishi 1999a; 1999b; Shatah and Struwe 1993; 1994] for global existence, regularity, scattering and blow-up. Since the semilinear wave equation can be viewed as a small perturbation of the homogenous linear wave equation in many situations, especially when we consider the asymptotic behaviors of solutions as spatial variables or time tends to infinity, it is important to first understand the asymptotic behaviors of solutions to the homogenous linear wave equation, i.e., free waves. This work is concerned with two important tools to understand the asymptotic behaviors of free waves: radiation fields and the channel of energy. We first introduce some necessary notation. Throughout this work we consider the homogenous linear wave equation with initial data in the energy space

$$\begin{cases} \partial_t^2 u - \Delta u = 0, \quad (x,t) \in \mathbb{R}^d \times \mathbb{R}, \\ u|_{t=0} = u_0 \in \dot{H}^1(\mathbb{R}^d), \\ u_t|_{t=0} = u_1 \in L^2(\mathbb{R}^d). \end{cases} \tag{1}$$

In this work we also use the notation $\boldsymbol{S}_L(u_0, u_1)$ to represent the free wave u defined above. If it is necessary to mention the velocity u_t, we use the notation

$$\boldsymbol{S}_L(t) \begin{pmatrix} u_0 \\ u_1 \end{pmatrix} = \begin{pmatrix} u(\cdot\,,t) \\ u_t(\cdot\,,t) \end{pmatrix} \in \dot{H}^1 \times L^2.$$

MSC2020: 35L05.
Keywords: wave equation, radiation field, channel of energy.

It is well known that the linear wave propagation preserves the $\dot{H}^1 \times L^2$ norm, i.e., the energy conservation law holds ($\nabla_{x,t} u = (\nabla u, u_t)$):

$$\int_{\mathbb{R}^d} |\nabla_{x,t} u(x,t)|^2 \, dx = \int_{\mathbb{R}^d} (|\nabla u_0|^2 + |u_1|^2) \, dx.$$

Now we make a brief review of radiation fields and the channel-of-energy method.

Radiation fields. The asymptotic behavior of free waves at the energy level can be characterized by the following theorem.

Theorem 1.1 (radiation field). *Assume that $d \geq 3$ and let u be a solution to the free wave equation $\partial_t^2 u - \Delta u = 0$, with initial data $(u_0, u_1) \in \dot{H}^1 \times L^2(\mathbb{R}^d)$. Then*

$$\lim_{t \to \pm\infty} \int_{\mathbb{R}^d} \left(|\nabla u(x,t)|^2 - |u_r(x,t)|^2 + \frac{|u(x,t)|^2}{|x|^2} \right) dx = 0$$

and there exist two functions $G_\pm \in L^2(\mathbb{R} \times \mathbb{S}^{d-1})$ so that

$$\lim_{t \to \pm\infty} \int_0^\infty \int_{\mathbb{S}^{d-1}} |r^{(d-1)/2} \partial_t u(r\theta, t) - G_\pm(r \mp t, \theta)|^2 \, d\theta \, dr = 0,$$
$$\lim_{t \to \pm\infty} \int_0^\infty \int_{\mathbb{S}^{d-1}} |r^{(d-1)/2} \partial_r u(r\theta, t) \pm G_\pm(r \mp t, \theta)|^2 \, d\theta \, dr = 0.$$

In addition, the maps $(u_0, u_1) \to \sqrt{2} G_\pm$ are a bijective isometries from $\dot{H}^1 \times L^2(\mathbb{R}^d)$ to $L^2(\mathbb{R} \times \mathbb{S}^{d-1})$.

This has been known for more than 50 years, at least in the 3-dimensional case. Please see [Friedlander 1962; 1980], for example. The version of the radiation field theorem given above and a proof for all dimensions $d \geq 3$ can be found in [Duyckaerts, Kenig and Merle 2019]. In addition, there is also a 2-dimensional version of the radiation field theorem. The statement in dimension 2 can be given in almost the same way as in the higher-dimensional case, except that the limit

$$\lim_{t \to \pm\infty} \int_{\mathbb{R}^2} \frac{|u(x,t)|^2}{|x|^2} \, dx = 0$$

no longer holds. A proof by Radon transform for all dimensions $d \geq 2$ can be found in [Katayama 2013], where the statement of the theorem is slightly different. Throughout this work we call the function $G_\pm$ radiation profiles and use the notation $\boldsymbol{T}_\pm$ for the linear map $(u_0, u_1) \to G_\pm$.

Channel of energy. The second tool is the channel-of-energy method, which plays an important role in the study of wave equations in the past decade. This method is first introduced in the 3-dimensional case in [Duyckaerts, Kenig and Merle 2011] and then in the 5-dimensional case in [Kenig, Lawrie and Schlag 2014]. This method was used in the proof of the soliton resolution conjecture of the energy critical wave equation with radial data in all odd dimensions $d \geq 3$ in [Duyckaerts, Kenig and Merle 2013; 2023]. It can also be used to show the nonexistence of minimal blow-up solutions in a compactness-rigidity argument in the energy super- or subcritical case. Please see, for example, [Duyckaerts, Kenig and Merle

2014; Shen 2013]. Roughly speaking, the channel-of-energy method discusses the amount of energy located in an exterior region as time tends to infinity:

$$\lim_{t\to\pm\infty}\int_{|x|>|t|+R}|\nabla_{x,t}u(x,t)|^2\,dx.$$

Here the constant R is nonnegative. Since the energy travels at a finite speed, the energy in the exterior region $\{x : |x| > |t| + R\}$ decays as $|t|$ increases. Thus the limits above are always well-defined. We can also give the exact value of the limit in terms of the radiation field:

$$\lim_{t\to\pm\infty}\int_{|x|>|t|+R}|\nabla_{x,t}u(x,t)|^2\,dx = 2\int_R^\infty\int_{\mathbb{S}^{d-1}}|G_\pm(s,\theta)|^2\,d\theta\,ds. \tag{2}$$

We first introduce a few already known results. We start with the odd dimensions.

Proposition 1.2 [Duyckaerts, Kenig and Merle 2012]. *Assume that $d \geq 3$ is an odd integer. All solutions to $\partial_t^2 u - \Delta u = 0$ satisfy*

$$\sum_{\pm}\lim_{t\to\pm\infty}\int_{|x|>|t|}|\nabla_{x,t}u(x,t)|^2\,dx = \int_{\mathbb{R}^d}|\nabla_{x,t}u(x,0)|^2\,dx. \tag{3}$$

As a result, we have:

Corollary 1.3. *Assume that $d \geq 3$ is odd. Then $u \equiv 0$ is the only free wave u satisfying*

$$\lim_{t\to\pm\infty}\int_{|x|>|t|}|\nabla_{x,t}u(x,t)|^2\,dx = 0.$$

In contrast, if $R > 0$, the subspace of $\dot{H}^1 \times L^2(\mathbb{R}^d)$ defined by

$$P(R) = \left\{(u_0,u_1)\in\dot{H}^1\times L^2(\mathbb{R}^d) : \lim_{t\to\pm\infty}\int_{|x|>R+|t|}|\nabla_{t,x}\boldsymbol{S}_L(u_0,u_1)(x,t)|^2\,dx = 0\right\} \tag{4}$$

does contain initial data (u_0, u_1) whose support is essentially bigger than $\{x : |x| \leq R\}$. The free waves u satisfying

$$\lim_{t\to\pm\infty}\int_{|x|>R+|t|}|\nabla_{t,x}u(x,t)|^2\,dx = 0$$

are usually called (R-weakly) nonradiative solutions. If the dimension is odd, these solutions are well-understood in the radial case:

Theorem 1.4 [Kenig, Lawrie, Liu and Schlag 2015]. *In any odd dimension $d \geq 1$, every radial solution u to (1) satisfies*

$$\max_{\pm}\lim_{t\to\pm\infty}\int_{r>|t|+R}|\nabla_{x,t}u(r,t)|^2 r^{d-1}\,dr \geq \tfrac{1}{2}\|\Pi^\perp_{P_{\mathrm{rad}}(R)}(u_0,u_1)\|_{\dot{H}^1\times L^2(r\geq R:r^{d-1}dr)}. \tag{5}$$

Here

$$P_{\mathrm{rad}}(R) \doteq \mathrm{Span}\left\{(r^{2k_1-d},0),(0,r^{2k_2-d}) : k_1,k_2\in\mathbb{N};\ 1\leq k_1\leq\frac{d+2}{4},\ 1\leq k_2\leq\frac{d}{4}\right\}.$$

$\Pi^\perp_{P_{\mathrm{rad}}(R)}$ is the orthogonal projection from $\dot{H}^1\times L^2(r\geq R : r^{d-1}dr)$ onto the complement of the finite-dimensional subspace $P_{\mathrm{rad}}(R)$.

Note the proof of Theorem 1.4 in [Kenig, Lawrie, Liu and Schlag 2015] uses the radial Fourier transform.

The case of even dimensions is much more complicated and subtle. Côte, Kenig and Schlag [2014] showed that in general the inequality

$$\sum_{\pm} \lim_{t\to\pm\infty} \int_{|x|>|t|} |\nabla_{x,t} u(x,t)|^2\,dx \geq C \int_{\mathbb{R}^d} |\nabla_{x,t} u(x,0)|^2\,dx$$

does not hold for any positive constant C in even dimensions. But a similar inequality holds in the radial case for either initial data $(u_0, 0)$ if $d = 0 \bmod 4$, or $(0, u_1)$ if $d = 2 \bmod 4$. More precisely we have

$$\lim_{t\to\pm\infty} \int_{|x|>|t|} |\nabla_{x,t} \boldsymbol{S}_L(u_0,0)(x,t)|^2\,dx \geq \frac{1}{2}\int_{\mathbb{R}^d} |\nabla u_0(x)|^2\,dx, \quad d = 4k,\ k\in\mathbb{N}, \tag{6}$$

$$\lim_{t\to\pm\infty} \int_{|x|>|t|} |\nabla_{x,t} \boldsymbol{S}_L(0,u_1)(x,t)|^2\,dx \geq \frac{1}{2}\int_{\mathbb{R}^d} |u_1(x)|^2\,dx, \quad d = 4k+2,\ k\in\mathbb{N}. \tag{7}$$

In addition, Duyckaerts, Kenig and Merle [2021] showed that the only nonradiative solution is still the zero solution in even dimensions $d \geq 4$; i.e., Corollary 1.3 still holds for even dimensions $d \geq 4$, even in the nonradial case. Much less is known about the exterior energy estimate in the region $\{x : |x| > R + |t|\}$ with $R > 0$. Duyckaerts, Kenig, Martel and Merle [2022] proves the exterior energy estimate in dimensions 4 and 6 if the initial data are radial:

$$\lim_{t\to\pm\infty} \int_{|x|>|t|+R} |\nabla_{x,t} \boldsymbol{S}_L(u_0,0)(x,t)|^2\,dx \geq \frac{3}{10}\|\Pi_R^{\perp} u_0\|^2_{\dot H^1(\{x\in\mathbb{R}^4:|x|>R\})},$$

$$\lim_{t\to\pm\infty} \int_{|x|>|t|+R} |\nabla_{x,t} \boldsymbol{S}_L(0,u_1)(x,t)|^2\,dx \geq \frac{3}{10}\|\pi_R^{\perp} u_1\|^2_{L^2(\{x\in\mathbb{R}^6:|x|>R\})}.$$

Here $\Pi_R^{\perp}$ is the orthogonal projection from $\dot H^1(\{x \in \mathbb{R}^4 : |x| > R\})$ onto the complement space of $\mathrm{Span}\{|x|^{-2}\}$. While $\pi_R^{\perp}$ is the orthogonal projection from $L^2(\{x \in \mathbb{R}^6 : |x| > R\})$ onto the complement space of $\mathrm{Span}\{|x|^{-4}\}$.

1B. ***Main idea.*** According to (2) we may obtain exterior energy estimates conveniently from the radiation profiles $G_\pm$. Please note that G_- and G_+ are not independent of each other. In fact the map $\boldsymbol{T}_+ \circ \boldsymbol{T}_-^{-1} : G_- \to G_+$ is a bijective isometry. If we could find explicit expressions of the maps

$$\boldsymbol{T}_+ \circ \boldsymbol{T}_-^{-1} : G_- \to G_+, \quad \boldsymbol{T}_-^{-1} : G_- \to (u_0, u_1), \quad \boldsymbol{S}_L \circ \boldsymbol{T}_-^{-1} : G_- \to u,$$

then we would be able to:

(a) Understand how the asymptotic behavior in one time direction determines the behavior in the other time direction. This is known in the odd-dimensional case, as shown (although not stated explicitly) in the proof of Proposition 1.2 in [Duyckaerts, Kenig and Merle 2012]. In this work we will try to figure out the even-dimensional case.

(b) Characterize (weakly) nonradiative solutions, especially in the radial case. We first determine all the radiation profiles G_- so that

$$G_-(s,\theta) = G_+(s,\theta) = 0, \quad s > R \quad \Longleftrightarrow \quad \lim_{t\to\pm\infty} \int_{|x|>|t|+R} |\nabla_{x,t} u(x,t)|^2\,dx = 0;$$

then we may obtain all the nonradiative solutions (as well as their initial data) by applying the formula of $\boldsymbol{T}_-^{-1}$. In particular we prove that radial nonradiative solutions in even dimensions can be characterized in the same way as in odd dimensions.

(c) Prove exterior energy estimates. We generalize the radial exterior energy estimates in 4 and 6 dimensions to all even dimensions; we also prove a nonradial exterior energy estimate in odd dimensions. In both applications (b) and (c) we follow the same roadmap:

$$\text{exterior energy} \quad \leftrightarrow \quad \text{radiation profile} \quad \leftrightarrow \quad \text{initial data.}$$

1C. ***Main results.*** Now we give the statement of our results. The details and proofs can be found in subsequent sections.

Theorem 1.5. *Let u be a finite-energy free wave with an even spatial dimension $d \geq 2$ and G_+, G_- be the radiation profiles associated with u. Then we may give an explicit formula of the operator $\boldsymbol{T}_+ \circ \boldsymbol{T}_-^{-1} : G_- \to G_+$*

$$G_+(s,\theta) = (-1)^{d/2}(\mathcal{H}G_-)(-s,-\theta).$$

Here $\mathcal{H}$ is the Hilbert transform in the first variable, i.e.,

$$(\mathcal{H}G_-)(-s,-\theta) = \text{p.v.}\,\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{G_-(\tau,-\theta)}{(-s)-\tau}\,d\tau.$$

Remark 1.6. A similar but simpler argument shows that if d is odd, then $\boldsymbol{T}_+ \circ \boldsymbol{T}_-^{-1} : G_- \to G_+$ can be explicitly given by

$$G_+(s,\theta) = (-1)^{(d-1)/2}G_-(-s,-\theta).$$

This can also be verified by assuming that the initial data is smooth and compactly supported, and considering the expression of G_- and G_+ in terms of (u_0, u_1) if d is odd. Please refer to [Duyckaerts, Kenig and Merle 2012]. Since we have $\mathcal{H}^2 = -1$. We may write the odd and even dimensions in a universal formula

$$G_+(s,\theta) = ((-\mathcal{H})^{d-1}G_-)(-s,-\theta).$$

Remark 1.7. It has been proved in Section 3.2 of [Duyckaerts, Kenig and Merle 2021] (in the language of Hankel and Laplace transforms) that the zero function is the only function $f \in L^2(\mathbb{R})$ satisfying

$$f(s) = 0, \quad s > 0, \qquad (\mathcal{H}f)(s) = 0, \quad s < 0.$$

It immediately follows that:

Corollary 1.8. *Assume $d \geq 2$. Let Ω be a region in $\mathbb{S}^{d-1}$. If a finite-energy solution u to the homogenous linear wave equation satisfies*

$$\lim_{t\to\pm\infty}\int_{|t|}^{\infty}\int_{\pm\Omega} |\nabla_{t,x}u(r\theta,t)|^2 r^{d-1}\,d\theta\,dr = 0,$$

then we have

$$\lim_{t\to\pm\infty}\int_{0}^{\infty}\int_{\pm\Omega} |\nabla_{t,x}u(r\theta,t)|^2 r^{d-1}\,d\theta\,dr = 0.$$

This is an angle-localized version of Corollary 1.3.

Applications on channel of energy. By following the idea described above, we obtain the following results about the channel of energy.

Proposition 1.9 (radial weakly nonradiative solutions). *Let $d \geq 2$ be an integer and $R > 0$ be a constant. If the initial data $(u_0, u_1) \in \dot{H}^1 \times L^2$ are radial, then the corresponding solution to the homogeneous linear wave equation u is R-weakly nonradiative, i.e.,*

$$\lim_{t\to\pm\infty} \int_{|x|>|t|+R} |\nabla_{t,x} u(x,t)|^2 \, dx = 0,$$

if and only if the restriction of (u_0, u_1) in the region $\{x \in \mathbb{R}^d : |x| > R\}$ is contained in

$$P_{\mathrm{rad}}(R) = \mathrm{Span}\left\{(r^{2k_1-d}, 0), (0, r^{2k_2-d}) : 1 \leq k_1 \leq \left\lfloor \frac{d+1}{4} \right\rfloor,\ 1 \leq k_2 \leq \left\lfloor \frac{d-1}{4} \right\rfloor\right\}.$$

Here the notation $\lfloor q \rfloor$ is the integer part of q. In particular, all radial R-weakly nonradiative solutions in dimension 2 are supported in $\{(x,t) : |x| \leq |t| + R\}$.

Remark 1.10. If d is odd, we have $\lfloor (d+1)/4 \rfloor = \lfloor (d+2)/4 \rfloor$ and $\lfloor (d-1)/4 \rfloor = \lfloor d/4 \rfloor$; thus our result here is the same as the already known result in odd dimension, as given in Theorem 1.4.

Proposition 1.11 (radial exterior estimates in even dimensions). *Let $d = 4k$ with $k \in \mathbb{N}$ and $R > 0$. If initial data $u_0 \in \dot{H}^1(\mathbb{R}^d)$ are radial, then the corresponding solution u to the homogenous linear wave equation with initial data $(u_0, 0)$ satisfies*

$$\lim_{t\to\infty} \int_{|x|>R+|t|} |\nabla u(x,t)|^2 \, dx = \lim_{t\to\infty} \int_{|x|>R+|t|} |u_t(x,t)|^2 \, dx \geq \tfrac{1}{4} \|\Pi^{\perp}_{Q_k(R)} u_0\|^2_{\dot{H}^1(\{x:|x|>R\})}.$$

Here $\Pi^{\perp}_{Q_k(R)}$ is the orthogonal projection from $\dot{H}^1(\{x \in \mathbb{R}^d : |x| > R\})$ onto the complement of the k-dimensional linear space

$$Q_k(R) = \mathrm{Span}\left\{\frac{1}{|x|^{4k-2k_1}} : 1 \leq k_1 \leq k\right\}.$$

Similarly if the dimension d satisfies $d = 4k+2 \geq 2$, with $k \in \{0\} \cup \mathbb{N}$ and the initial data $u_1 \in L^2(\mathbb{R}^d)$ are radial, then the corresponding solution u to the homogenous linear wave equation with initial data $(0, u_1)$ satisfies

$$\lim_{t\to\infty} \int_{|x|>R+|t|} |\nabla u(x,t)|^2 \, dx = \lim_{t\to\infty} \int_{|x|>R+|t|} |u_t(x,t)|^2 \, dx \geq \tfrac{1}{4} \|\Pi^{\perp}_{Q'_k(R)} u_1\|^2_{L^2(\{x:|x|>R\})}.$$

Here $\Pi^{\perp}_{Q'_k(R)}$ is the orthogonal projection from $L^2(\{x \in \mathbb{R}^d : |x| > R\})$ onto the complement of the k-dimensional linear space

$$Q'_k(R) = \mathrm{Span}\left\{\frac{1}{|x|^{4k+2-2k_1}} : 1 \leq k_1 \leq k\right\}.$$

Remark 1.12. Given $u_0 \in \dot{H}^1(\mathbb{R}^{4k})$ or $u_1 \in L^2(\mathbb{R}^{4k+2})$, the orthogonal projection of u_0 or u_1 onto the finite-dimensional space $Q_k(R)$ or $Q'_k(R)$ gradually vanishes as $R \to 0^+$. Therefore if we make $R \to 0^+$ in Proposition 1.11, we immediately obtain (6) and (7).

Remark 1.13. At the same time as this work was done, Kenig et al. proved radial exterior estimates similar to Proposition 1.11 for even dimensions $d \geq 8$ by using the already-known result in dimension 4 and an induction argument.

Proposition 1.14 (nonradial exterior energy estimates). *Let $d \geq 3$ be an odd integer and $R > 0$ be a constant. Then the following identity holds for all $(u_0, u_1) \in \dot{H}^1 \times L^2(\mathbb{R}^d)$:*

$$\sum_{\pm} \lim_{t \to \pm\infty} \int_{|x|>R+|t|} |\nabla_{t,x} \boldsymbol{S}_L(t)(u_0, u_1)(x, t)|^2 \, dx = \|\Pi^{\perp}_{P(R)}(u_0, u_1)\|^2_{\dot{H}^1 \times L^2(\mathbb{R}^d)}.$$

Here $\Pi^{\perp}_{P(R)}$ is the orthogonal projection from $\dot{H}^1 \times L^2(\mathbb{R}^d)$ onto the complement of the closed linear space

$$P(R) = \left\{ (u_0, u_1) \in \dot{H}^1 \times L^2(\mathbb{R}^d) : \lim_{t \to \pm\infty} \int_{|x|>R+|t|} |\nabla_{t,x} \boldsymbol{S}_L(u_0, u_1)(x, t)|^2 \, dx = 0 \right\}.$$

Structure of this work. This work is organized as follows. In Section 2 we deduce an explicit formula of $\boldsymbol{T}_-^{-1}$ in all dimensions. Then in Section 3 we prove the explicit formula of $\boldsymbol{T}_+ \circ \boldsymbol{T}_-^{-1}$ given in Theorem 1.5. The rest of the paper is devoted to the applications in the channel of energy. We characterize radial weakly nonradiative solutions in Section 4, prove radial exterior energy estimate for all even dimensions in Section 5 and finally give a short proof of nonradial exterior energy estimate in odd-dimensional space in Section 6. The Appendix is concerned with Hilbert transform of a family of special functions, since the Hilbert transform is involved in the even dimensions.

Notation. In this work we use the notation $C(d)$ for a nonzero constant determined solely by the dimension d. It may represent different constants in different places. This avoids the trouble of keeping track of the constants when unnecessary.

2. From radiation profile to solution

Now we assume that $G_-(r, \theta)$ is smooth and compactly supported and give an explicit formula of the operator $\boldsymbol{T}_-^{-1}$. We consider the odd dimensions first.

2A. *Odd dimensions.*

Lemma 2.1. *Assume that $d \geq 3$ is odd. Let G_- be a smooth function with* $\operatorname{supp} G_- \subset [-R, R] \times \mathbb{S}^{d-1}$. *Then $(u_0, u_1) = \boldsymbol{T}_-^{-1} G_-$ satisfies*

$$u_0(x) = \frac{1}{(2\pi)^{(d-1)/2}} \int_{\mathbb{S}^{d-1}} G_-^{(\mu-1)}(x \cdot \omega, \omega) \, d\omega, \tag{8}$$

$$u_1(x) = \frac{1}{(2\pi)^{(d-1)/2}} \int_{\mathbb{S}^{d-1}} G_-^{(\mu)}(x \cdot \omega, \omega) \, d\omega. \tag{9}$$

Here the notation $G_-^{(k)}$ represents the partial derivative

$$G_-^{(k)}(s, \theta) = \frac{\partial^k G_-(s, \theta)}{\partial s^k}.$$

Remark 2.2. This formula in 3-dimensional case was previously known. Please refer to [Friedlander 1973], for example.

Proof. Let $(u_0, u_1) = \boldsymbol{T}_-^{-1} G_-$ and $u = \boldsymbol{S}_L(u_0, u_1)$. Given a large time $t > 0$, we choose approximated data $(v_{0,t}, v_{1,t}) \approx (u(\cdot, -t), u_t(\cdot, -t))$ as

$$v_{1,t}(r\theta) = r^{-\mu} G_-(r - t, \theta), \qquad r > 0,\ \theta \in \mathbb{S}^{d-1}, \tag{10}$$

$$v_{0,t}(r\theta) = -\chi\left(\frac{r}{t}\right) \int_r^{+\infty} r'^{-\mu} G_-(r' - t, \theta)\, dr', \quad r > 0,\ \theta \in \mathbb{S}^{d-1}. \tag{11}$$

Here $\mu = (d-1)/2$ and $\chi : \mathbb{R} \to [0, 1]$ is a smooth center cut-off function satisfying

$$\chi(s) = \begin{cases} 1, & s > \frac{1}{2}, \\ 0, & s < \frac{1}{4}. \end{cases}$$

It is clear that the data $(v_{0,t}, v_{1,t})$ are smooth and compactly supported in $\{x : |x| < R+t\}$. A straightforward calculation shows that

$$\int_0^\infty \int_{\mathbb{S}^{d-1}} |r^\mu v_{1,t}(r\theta) - G_-(r - t, \theta)|^2\, d\theta\, dr = 0,$$

$$\int_0^\infty \int_{\mathbb{S}^{d-1}} |r^\mu \partial_r v_{0,t}(r\theta) - G_-(r - t, \theta)|^2\, d\theta\, dr \lesssim \frac{1}{t},$$

$$\int_{\mathbb{R}^d} (|\nabla v_{0,t}(x)|^2 - |\partial_r v_{0,t}(x)|^2)\, dx \lesssim \frac{1}{t}.$$

Thus by the radiation field we have

$$\lim_{t \to +\infty} \|(v_{0,t}, v_{1,t}) - (u(\cdot, -t), u_t(\cdot, -t))\|_{\dot{H}^1 \times L^2(\mathbb{R}^d)} = 0.$$

Since the linear propagation operator $\boldsymbol{S}_L(t)$ preserves the $\dot{H}^1 \times L^2$ norm, we have

$$\lim_{t \to +\infty} \left\| \begin{pmatrix} u_0 \\ u_1 \end{pmatrix} - \boldsymbol{S}_L(t) \begin{pmatrix} v_{0,t} \\ v_{1,t} \end{pmatrix} \right\|_{\dot{H}^1 \times L^2(\mathbb{R}^d)} = 0. \tag{12}$$

Next we use the explicit expression of the linear propagation operator (see, for instance, [Evans 1998]) and write $v = \boldsymbol{S}_L(v_0, v_1)$ in terms of (v_0, v_1) when the initial data are sufficiently smooth:

$$\begin{aligned} v(x, t) &= c_d \cdot \frac{\partial}{\partial t}\left(\frac{1}{t}\frac{\partial}{\partial t}\right)^{\mu-1} \left(t^{d-2} \int_{\mathbb{S}^{d-1}} v_0(x + t\omega)\, d\omega\right) + c_d \cdot \left(\frac{1}{t}\frac{\partial}{\partial t}\right)^{\mu-1} \left(t^{d-2} \int_{\mathbb{S}^{d-1}} v_1(x + t\omega)\, d\omega\right) \\ &= c_d t^\mu \int_{\mathbb{S}^{d-1}} [((w \cdot \nabla)^\mu v_0)(x + t\omega) + ((w \cdot \nabla)^{\mu-1} v_1)(x + t\omega)]\, d\omega \\ &\quad + \sum_{0 \le k < \mu} A_{d,k} t^k \int_{\mathbb{S}^{d-1}} ((w \cdot \nabla)^k v_0)(x + t\omega)\, d\omega + \sum_{0 \le k < \mu - 1} B_{d,k} t^{k+1} \int_{\mathbb{S}^{d-1}} ((w \cdot \nabla)^k v_1)(x + t\omega)\, d\omega. \end{aligned}$$

Here $c_d = 1/(2(2\pi)^{(d-1)/2})$, $A_{d,k}$, $B_{d,k}$ (and $A'_{d,k}$, $B'_{d,k}$ below) are all constants. We may differentiate and obtain

$$\begin{aligned} v_t(x, t) &= c_d t^\mu \int_{\mathbb{S}^{d-1}} [((w \cdot \nabla)^{\mu+1} v_0)(x + t\omega) + ((w \cdot \nabla)^\mu v_1)(x + t\omega)]\, d\omega \\ &\quad + \sum_{1 \le k \le \mu} A'_{d,k} t^{k-1} \int_{\mathbb{S}^{d-1}} ((w \cdot \nabla)^k v_0)(x + t\omega)\, d\omega + \sum_{0 \le k \le \mu - 1} B'_{d,k} t^k \int_{\mathbb{S}^{d-1}} ((w \cdot \nabla)^k v_1)(x + t\omega)\, d\omega. \end{aligned}$$

Now we plug in $(v_0, v_1) = (v_{0,t}, v_{1,t})$ with large time t. We observe that

$$|(\omega\cdot\nabla)^k v_{j,t}(x+tw)| \lesssim t^{-\mu}, \quad j=0,1,\ k\geq 0, \tag{13}$$

and $(r=|x+t\omega|, \theta=(x+t\omega)/|x+t\omega|,\ k=\mu-1,\mu)$

$$((\omega\cdot\nabla)^{k+1}v_{0,t})(x+t\omega) = (\omega\cdot\theta)^{k+1}r^{-\mu}G_-^{(k)}(r-t,\theta)+O(t^{-\mu-1}),$$
$$((\omega\cdot\nabla)^{k}v_{1,t})(x+t\omega) = (\omega\cdot\theta)^{k}r^{-\mu}G_-^{(k)}(r-t,\theta)+O(t^{-\mu-1}).$$

Thus

$$\begin{pmatrix} w_{0,t} \\ w_{1,t}\end{pmatrix} = \boldsymbol{S}_L(t)\begin{pmatrix} v_{0,t} \\ v_{1,t}\end{pmatrix}$$

satisfies

$$w_{0,t} = c_d\int_{\mathbb{S}^{d-1}}(\omega\cdot\theta)^{\mu-1}(1+\omega\cdot\theta)G_-^{(\mu-1)}(r-t,\theta)\,d\omega + O\Big(\frac{1}{t}\Big),$$
$$w_{1,t} = c_d\int_{\mathbb{S}^{d-1}}(\omega\cdot\theta)^{\mu}(1+\omega\cdot\theta)G_-^{(\mu)}(r-t,\theta)\,d\omega + O\Big(\frac{1}{t}\Big).$$

Please note that the implicit constants in (13), $O(t^{-\mu-1})$ and $O(1/t)$ above, may depend on x but remain uniformly bounded if x is contained in a compact subset of $\mathbb{R}^d$. Next we observe the facts

$$\theta(\omega) = \omega + O\Big(\frac{1}{t}\Big), \quad r(\omega)-t = x\cdot\omega + O\Big(\frac{1}{t}\Big),$$

and further simplify the formulas

$$w_{0,t} = 2c_d\int_{\mathbb{S}^{d-1}} G_-^{(\mu-1)}(x\cdot\omega,\omega)\,d\omega + O\Big(\frac{1}{t}\Big),$$
$$w_{1,t} = 2c_d\int_{\mathbb{S}^{d-1}} G_-^{(\mu)}(x\cdot\omega,\omega)\,d\omega + O\Big(\frac{1}{t}\Big).$$

Finally we make $t\to+\infty$, utilize (12) and obtain

$$u_0 = 2c_d\int_{\mathbb{S}^{d-1}} G_-^{(\mu-1)}(x\cdot\omega,\omega)\,d\omega,$$
$$u_1 = 2c_d\int_{\mathbb{S}^{d-1}} G_-^{(\mu)}(x\cdot\omega,\omega)\,d\omega.$$

We plug in the value of c_d and finish the proof. □

Remark 2.3. An explicit formula of the free wave $u = \boldsymbol{S}_L\boldsymbol{T}_-^{-1}G_-$ can be given by

$$u(x,t) = \frac{1}{(2\pi)^{(d-1)/2}}\int_{\mathbb{S}^{d-1}} G_-^{(\mu-1)}(x\cdot\omega+t,\omega)\,d\omega.$$

This can be verified by a straightforward calculation. One may check

- The function u above is a smooth solution to the homogenous linear wave equation.
- The initial data of u are exactly those given in Lemma 2.1.

We may differentiate and obtain

$$u_t(x,t) = \frac{1}{(2\pi)^{(d-1)/2}} \int_{\mathbb{S}^{d-1}} G_-^{(\mu)}(x\cdot\omega + t, \omega)\, d\omega,$$
$$\nabla u(x,t) = \frac{1}{(2\pi)^{(d-1)/2}} \int_{\mathbb{S}^{d-1}} G_-^{(\mu)}(x\cdot\omega + t, \omega)\, \omega\, d\omega.$$

2B. ***Even dimensions.*** The formula of $\boldsymbol{T}_-^{-1}$ in even dimensions is a little more complicated.

Lemma 2.4. *Assume that $d \geq 2$ is even and $G_- \in C_0^\infty(\mathbb{R}\times\mathbb{S}^{d-1})$. Then the operator $\boldsymbol{T}_-^{-1}$ is given explicitly by*

$$u_0(x) = \frac{\sqrt{2}}{(2\pi)^{d/2}} \cdot \int_0^\infty \int_{\mathbb{S}^{d-1}} \frac{G_-^{(d/2-1)}(x\cdot\omega - \rho, \omega)}{\sqrt{\rho}}\, d\omega\, d\rho,$$
$$u_1(x) = \frac{\sqrt{2}}{(2\pi)^{d/2}} \cdot \int_0^\infty \int_{\mathbb{S}^{d-1}} \frac{G_-^{(d/2)}(x\cdot\omega - \rho, \omega)}{\sqrt{\rho}}\, d\omega\, d\rho.$$

Proof. Without loss of generality let us assume $\operatorname{supp} G_- \subset [-R_1, R_1]\times\mathbb{S}^{d-1}$. It is sufficient to show that given any $R_2 > 0$, the formula above holds for almost every $x \in B(0, R_2)$. Let us use the notation $(u_0, u_1) = \boldsymbol{T}_-^{-1}(G_-)$ and $u = \boldsymbol{S}_L(u_0, u_1)$. We consider the approximated data

$$v_{1,t}(r\theta) = r^{-\mu} G_-(r-t, \theta),$$
$$v_{0,t}(r\theta) = -\chi\left(\frac{r}{t}\right) \int_r^{+\infty} r'^{-\mu} G_-(r'-t, \theta)\, dr', \quad r > 0,\ \theta \in \mathbb{S}^{d-1}.$$

and

$$\begin{pmatrix} w_{0,t} \\ w_{1,t} \end{pmatrix} = \boldsymbol{S}_L(t) \begin{pmatrix} v_{0,t} \\ v_{1,t} \end{pmatrix}.$$

Here χ is the center cut-off function as given in the previous subsection. A basic calculation shows

$$\lim_{t\to+\infty} \|(v_{0,t}, v_{1,t}) - (u(\cdot, -t), u_t(\cdot, -t))\|_{\dot{H}^1\times L^2(\mathbb{R}^d)} = 0.$$

Thus

$$\lim_{t\to+\infty} \|(w_{0,t}, w_{1,t}) - (u_0, u_1)\|_{\dot{H}^1\times L^2(\mathbb{R}^d)} = 0. \tag{14}$$

Let us first recall the explicit formula of $v = \boldsymbol{S}_L(v_0, v_1)$ in the even-dimensional case:

$$\begin{aligned} v(x,t) &= c_d \cdot \frac{\partial}{\partial t}\left(\frac{1}{t}\frac{\partial}{\partial t}\right)^{(d-2)/2} \left(t^{d-1} \int_{\mathbb{B}^d} \frac{v_0(x+ty)}{\sqrt{1-|y|^2}}\, dy\right) + c_d \cdot \left(\frac{1}{t}\frac{\partial}{\partial t}\right)^{(d-2)/2} \left(t^{d-1} \int_{\mathbb{B}^d} \frac{v_1(x+ty)}{\sqrt{1-|y|^2}}\, dy\right) \\ &= c_d \cdot t^{d/2} \int_{\mathbb{B}^d} \frac{((y\cdot\nabla)^{d/2} v_0)(x+ty) + ((y\cdot\nabla)^{d/2-1} v_1)(x+ty)}{\sqrt{1-|y|^2}}\, dy \\ &\quad + \sum_{0\leq k<d/2} A_{d,k} t^k \int_{\mathbb{B}^d} \frac{(y\cdot\nabla)^k v_0(x+ty)}{\sqrt{1-|y|^2}}\, dy + \sum_{0\leq k<d/2-1} B_{d,k} t^{k+1} \int_{\mathbb{B}^d} \frac{(y\cdot\nabla)^k v_1(x+ty)}{\sqrt{1-|y|^2}}\, dy. \end{aligned}$$

Here $\mathbb{B}_d$ is the unit ball in $\mathbb{R}^d$ and $c_d = (2\pi)^{-d/2}$ is a constant. The notations $A_{d,k}$, $B_{d,k}$ (and $A'_{d,k}$, $B'_{d,k}$ below) represent constants. We differentiate and obtain

$$v_t(x,t) = c_d \cdot t^{d/2} \int_{\mathbb{B}^d} \frac{((y\cdot\nabla)^{d/2+1} v_0)(x+ty) + ((y\cdot\nabla)^{d/2} v_1)(x+ty)}{\sqrt{1-|y|^2}}\, dy$$
$$+ \sum_{1\le k\le d/2} A'_{d,k} t^{k-1} \int_{\mathbb{B}^d} \frac{(y\cdot\nabla)^k v_0(x+ty)}{\sqrt{1-|y|^2}}\, dy + \sum_{0\le k\le d/2-1} B'_{d,k} t^k \int_{\mathbb{B}^d} \frac{(y\cdot\nabla)^k v_1(x+ty)}{\sqrt{1-|y|^2}}\, dy.$$

We plug in $(v_0, v_1) = (v_{0,t}, v_{1,t})$ and observe

$$|(y\cdot\nabla)^k v_{0,t}| \le t^{-(d-1)/2}, \quad |(y\cdot\nabla)^k v_{1,t}| \le t^{-(d-1)/2}.$$

This gives the approximation

$$w_{0,t}(x) = c_d \cdot t^{d/2} \int_{\mathbb{B}^d} \frac{((y\cdot\nabla)^{d/2} v_{0,t})(r\theta) + ((y\cdot\nabla)^{d/2-1} v_{1,t})(r\theta)}{\sqrt{1-|y|^2}}\, dy + O(t^{-1/2}),$$
$$w_{1,t}(x) = c_d \cdot t^{d/2} \int_{\mathbb{B}^d} \frac{((y\cdot\nabla)^{d/2+1} v_{0,t})(r\theta) + ((y\cdot\nabla)^{d/2} v_{1,t})(r\theta)}{\sqrt{1-|y|^2}}\, dy + O(t^{-1/2}).$$

Here $r = |x+ty|$ and $\theta = (x+ty)/|x+ty|$. Furthermore, we observe ($k = d/2, d/2-1$)

$$((y\cdot\nabla)^{k+1} v_{0,t})(r\theta) = (y\cdot\theta)^{k+1} r^{-(d-1)/2} G_-^{(k)}(r-t,\theta) + O(t^{-(d+1)/2}),$$
$$((y\cdot\nabla)^{k} v_{1,t})(r\theta) = (y\cdot\theta)^{k} r^{-(d-1)/2} G_-^{(k)}(r-t,\theta) + O(t^{-(d+1)/2}),$$

and write

$$w_{0,t}(x) = c_d \cdot t^{d/2} \int_{\mathbb{B}^d} \frac{(y\cdot\theta)^{d/2-1}(y\cdot\theta+1) r^{-(d-1)/2} G_-^{(d/2-1)}(r-t,\theta)}{\sqrt{1-|y|^2}}\, dy + O(t^{-1/2}),$$
$$w_{1,t}(x) = c_d \cdot t^{d/2} \int_{\mathbb{B}^d} \frac{(y\cdot\theta)^{d/2}(y\cdot\theta+1) r^{-(d-1)/2} G_-^{(d/2)}(r-t,\theta)}{\sqrt{1-|y|^2}}\, dy + O(t^{-1/2}).$$

Next we observe that if $|y| < 1-(R_1+R_2)/t$, then we have $r \le t|y| + |x| < t - R_1$; thus $G_-^{(k)}(r-t,\theta) = 0$. As a result, we may restrict the domain of the integral to

$$\mathbb{B}_t = \left\{ y \in \mathbb{B}^d : |y| \ge 1 - \frac{R_1+R_2}{t} \right\}.$$

Because in the region we have

$$\theta = \frac{y}{|y|} + O\Big(\frac{1}{t}\Big), \quad y\cdot\theta = 1 + O\Big(\frac{1}{t}\Big), \quad r = t + O(1),$$

we can simplify the formulas

$$w_{0,t}(x) = 2c_d \cdot t^{1/2} \int_{\mathbb{B}_t} \frac{G_-^{(d/2-1)}(r-t, y/|y|)}{\sqrt{1-|y|^2}}\, dy + O(t^{-1/2}),$$
$$w_{1,t}(x) = 2c_d \cdot t^{1/2} \int_{\mathbb{B}_t} \frac{G_-^{(d/2)}(r-t, y/|y|)}{\sqrt{1-|y|^2}}\, dy + O(t^{-1/2}).$$

Next we utilize the change of variables

$$y = \Big(1 - \frac{\rho}{t}\Big)\omega, \quad (\rho, \omega) \in (0, R_1 + R_2) \times \mathbb{S}^{d-1},$$

and the approximations

$$r - t = x \cdot \omega - \rho + O\Big(\frac{1}{t}\Big), \quad \sqrt{1 - |y|^2} = \Big(1 + O\Big(\frac{1}{t}\Big)\Big)\sqrt{\frac{2\rho}{t}}, \quad dy = \Big(1 + O\Big(\frac{1}{t}\Big)\Big)t^{-1}\, d\rho\, d\omega$$

to obtain

$$w_{0,t}(x) = \sqrt{2}c_d \cdot \int_0^{R_1+R_2}\int_{\mathbb{S}^{d-1}} \frac{G_-^{(d/2-1)}(x \cdot \omega - \rho, \omega)}{\sqrt{\rho}}\, d\omega\, d\rho + O(t^{-1/2}),$$

$$w_{1,t}(x) = \sqrt{2}c_d \cdot \int_0^{R_1+R_2}\int_{\mathbb{S}^{d-1}} \frac{G_-^{(d/2)}(x \cdot \omega - \rho, \omega)}{\sqrt{\rho}}\, d\omega\, d\rho + O(t^{-1/2}).$$

Finally we recall (14), let $t \to +\infty$ and conclude

$$u_0(x) = \sqrt{2}c_d \cdot \int_0^{R_1+R_2}\int_{\mathbb{S}^{d-1}} \frac{G_-^{(d/2-1)}(x \cdot \omega - \rho, \omega)}{\sqrt{\rho}}\, d\omega\, d\rho,$$

$$u_1(x) = \sqrt{2}c_d \cdot \int_0^{R_1+R_2}\int_{\mathbb{S}^{d-1}} \frac{G_-^{(d/2)}(x \cdot \omega - \rho, \omega)}{\sqrt{\rho}}\, d\omega\, d\rho. \qquad \square$$

Remark 2.5. If $d \geq 4$, the convergence (14) implies that $(w_{0,t}, w_{1,t})$ converges to (u_0, u_1) in $L^{2d/(d-2)} \times L^2$ by Sobolev embedding. We may combine this convergence with the local uniform convergence given above to verify the identities above. This argument breaks down in dimension 2. We given another argument below in dimension 2. Given any test function $\varphi \in C_0^\infty(\mathbb{R}^2)$, integration by parts gives an identity

$$\int w_{0,t}(x)\nabla\varphi(x)\, dx = -\int \nabla w_{0,t}(x)\varphi(x)\, dx.$$

We recall the local uniform convergence of $w_{0,t}$ given above and the L^2 convergence of $\nabla w_{0,t} \to \nabla u_0$ and then obtain

$$\int \Big(\sqrt{2}c_2 \cdot \int_0^\infty\int_{\mathbb{S}^1} \frac{G_-(x \cdot \omega - \rho, \omega)}{\sqrt{\rho}}\, d\omega\, d\rho\Big)\nabla\varphi(x)\, dx = -\int \nabla u_0(x)\varphi(x)\, dx.$$

This finishes the proof. Finally we would like to mention that we have

$$\lim_{|x|\to+\infty} \sqrt{2}c_2 \cdot \int_0^\infty\int_{\mathbb{S}^1} \frac{G_-(x \cdot \omega - \rho, \omega)}{\sqrt{\rho}}\, d\omega\, d\rho = 0.$$

Corollary 2.6. *If $G_- \in C_0^\infty(\mathbb{R} \times \mathbb{S}^{d-1})$, then $u = \boldsymbol{S}_L \boldsymbol{T}_-^{-1}(G_-)$ is given by*

$$u(x,t) = \frac{\sqrt{2}}{(2\pi)^{d/2}} \int_0^\infty\int_{\mathbb{S}^{d-1}} \frac{G_-^{(d/2-1)}(x \cdot \omega - \rho + t, \omega)}{\sqrt{\rho}}\, d\omega\, d\rho.$$

Thus

$$u_t(x,t) = \frac{\sqrt{2}}{(2\pi)^{d/2}} \int_0^\infty\int_{\mathbb{S}^{d-1}} \frac{G_-^{(d/2)}(x \cdot \omega - \rho + t, \omega)}{\sqrt{\rho}}\, d\omega\, d\rho.$$

Proof. A basic calculation shows that $u(x,t)$ solves the free wave equation with initial data given in Lemma 2.4. $\square$

2C. ***Universal formula.*** Now let us give a universal formula of $\boldsymbol{T}_-^{-1}$ for all dimensions. We first define two convolution operators ($1/\sqrt{\pi x}$ is understood as zero if $x<0$)

$$\mathcal{Q}f = \frac{1}{\sqrt{\pi x}} * f, \quad \mathcal{Q}'f = \frac{1}{\sqrt{-\pi x}} * f.$$

Their Fourier symbols are

$$\frac{1-i(\xi/|\xi|)}{2\sqrt{\pi|\xi|}} \quad \text{and} \quad \frac{1+i(\xi/|\xi|)}{2\sqrt{\pi|\xi|}},$$

respectively. Let us also use the notation $\mathcal{D}=d/dx$ and recall that its Fourier symbol is $2\pi i\xi$. A simple calculation of symbols shows

$$\mathcal{Q}^2\mathcal{D}=1, \quad \mathcal{Q}'^2\mathcal{D}=-1, \quad \mathcal{Q}\mathcal{Q}'\mathcal{D}=\mathcal{H}. \tag{15}$$

As a result, we may understand $\mathcal{Q}$ as $\mathcal{D}^{-1/2}$ and rewrite $u=\boldsymbol{S}_L\boldsymbol{T}_-^{-1}G_-$ in the form

$$\begin{aligned} u(x,t) &= \frac{1}{(2\pi)^{(d-1)/2}}\int_{\mathbb{S}^{d-1}} (\mathcal{Q}G_-^{(d/2-1)})(x\cdot\omega+t,\omega)\,d\omega \\ &= \frac{1}{(2\pi)^{\mu}}\int_{\mathbb{S}^{d-1}} \mathcal{D}^{\mu-1}G_-(x\cdot\omega+t,\omega)\,d\omega. \end{aligned} \tag{16}$$

Here $\mu=(d-1)/2$. This formula holds for both odd and even dimensions.

3. Between radiation profiles

In this section we give an explicit expression of the operator $\boldsymbol{T}_+\circ\boldsymbol{T}_-^{-1}$ in the even-dimensional case, without the radial assumption.

Theorem 3.1. *Assume that $d\geq 2$ is an even integer. The operator $\boldsymbol{T}_+\circ\boldsymbol{T}_-^{-1}$ can be explicitly given by the formula*

$$G_+(s,\theta)=(\boldsymbol{T}_+\boldsymbol{T}_-^{-1}G_-)(s,\theta)=(-1)^{d/2}(\mathcal{H}G_-)(-s,-\theta).$$

Here $\mathcal{H}$ is the Hilbert transform in the first variable, i.e.,

$$(\mathcal{H}G_-)(-s,-\theta)=\text{p.v.}\,\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{G_-(\tau,-\theta)}{-\tau-s}\,d\tau.$$

Proof. Since $\boldsymbol{T}_+\circ\boldsymbol{T}_-^{-1}$ is a bijective isometry from $L^2(\mathbb{R}\times\mathbb{S}^{d-1})$ to itself. We only need to prove this formula for smooth and compactly supported data G_-. Without loss of generality let us assume $\operatorname{supp} G_-\subset[-R_1,R_1]\times\mathbb{S}^{d-1}$. Let us also fix a positive constant $R_2>0$. If $(s,\theta)\in(-R_2,R_2)\times\mathbb{S}^{d-1}$, then we may apply Corollary 2.6 and obtain

$$(t+s)^{(d-1)/2}\partial_t u((t+s)\theta,t)=\sqrt{2}c_d(t+s)^{(d-1)/2}\int_0^\infty\int_{\mathbb{S}^{d-1}}\frac{G_-^{(d/2)}((t+s)\theta\cdot\omega-\rho+t,\omega)}{\sqrt{\rho}}\,d\omega\,d\rho.$$

Let $M \gg R_1 + R_2 + 1$ be a large constant. We may split the integral above into two parts:

$$J_1 = \sqrt{2} c_d (t+s)^{(d-1)/2} \int_0^\infty \int_{\theta\cdot\omega < -1+M/t} \frac{G_-^{(d/2)}((t+s)\theta\cdot\omega - \rho + t, \omega)}{\sqrt{\rho}} \, d\omega \, d\rho,$$

$$J_2 = \sqrt{2} c_d (t+s)^{(d-1)/2} \int_0^\infty \int_{\theta\cdot\omega \geq -1+M/t} \frac{G_-^{(d/2)}((t+s)\theta\cdot\omega - \rho + t, \omega)}{\sqrt{\rho}} \, d\omega \, d\rho.$$

We may find an upper bound of J_2. In this region we have

$$(t+s)\theta\cdot\omega + t \geq M - R_2 \quad \Longrightarrow \quad G_-((t+s)\theta\cdot\omega - \rho + t) = 0 \quad \text{if } \rho < \frac{M}{2}.$$

Thus we may integrate by parts and obtain

$$J_2 = C(d)(t+s)^{(d-1)/2} \int_0^\infty \int_{\theta\cdot\omega \geq -1+M/t} \frac{G_-((t+s)\theta\cdot\omega - \rho + t, \omega)}{\rho^{(d+1)/2}} \, d\omega \, d\rho.$$

Therefore when t is sufficiently large

$$\begin{aligned}
|J_2| &\lesssim t^{(d-1)/2} \int_{\theta\cdot\omega \geq -1+M/t} \int_{(t+s)\theta\cdot\omega + t - R_1}^{(t+s)\theta\cdot\omega + t + R_1} \frac{|G_-((t+s)\theta\cdot\omega - \rho + t, \omega)|}{\rho^{(d+1)/2}} \, d\rho \, d\omega \\
&\lesssim t^{(d-1)/2} \int_{\theta\cdot\omega \geq -1+M/t} \int_{(t+s)\theta\cdot\omega + t - R_1}^{(t+s)\theta\cdot\omega + t + R_1} \frac{|G_-((t+s)\theta\cdot\omega - \rho + t, \omega)|}{|(t+s)\theta\cdot\omega + t|^{(d+1)/2}} \, d\rho \, d\omega \\
&\lesssim t^{(d-1)/2} \int_{\theta\cdot\omega \geq -1+M/t} \frac{1}{|t\theta\cdot\omega + t|^{(d+1)/2}} \, d\omega \lesssim \frac{1}{M}.
\end{aligned}$$

In the integral region of J_1, we have the approximation $\omega = -\theta + O(t^{-1/2})$. Thus we have

$$J_1 = \sqrt{2} c_d t^{(d-1)/2} \int_0^\infty \int_{\theta\cdot\omega < -1+M/t} \frac{G_-^{(d/2)}((t+s)\theta\cdot\omega - \rho + t, -\theta)}{\sqrt{\rho}} \, d\omega \, d\rho + O(t^{-1/2}).$$

Next we utilize the change of variables (please refer to Figure 1 for a geometrical meaning)

$$\omega = \Big(-1 + \frac{\rho'}{t}\Big)\theta + \sqrt{\Big(\frac{\rho'}{t}\Big)\Big(2 - \frac{\rho'}{t}\Big)}\varphi, \quad \rho' \in [0, M], \quad \varphi \in \mathbb{S}^{d-2} = \{\varphi \in \mathbb{S}^{d-1} : \varphi \perp \theta\},$$

$$d\omega = \Big[1 + O\Big(\frac{1}{t}\Big)\Big]\Big(\frac{2\rho'}{t}\Big)^{d/2-1} d\mathbb{S}^{d-2}(\varphi) \cdot \frac{d\rho'}{\sqrt{2\rho' t}} = \Big[1 + O\Big(\frac{1}{t}\Big)\Big](2\rho')^{(d-3)/2} t^{-(d-1)/2} \, d\mathbb{S}^{d-2}(\varphi) \, d\rho'$$

and obtain

$$J_1 = \frac{1}{2\pi^{d/2}} \int_0^\infty \int_0^M \int_{\mathbb{S}^{d-2}} G_-^{(d/2)}(\rho' - \rho - s, -\theta) \rho'^{(d-3)/2} \rho^{-1/2} \, d\varphi \, d\rho' \, d\rho + O(t^{-1/2}).$$

We observe that the integrand is independent of φ and integrate by parts

$$J_1 = \frac{(-1)^{d/2-1}}{\pi} \int_0^\infty \int_0^M \frac{G'_-(\rho' - \rho - s, -\theta)}{\sqrt{\rho\rho'}} \, d\rho' \, d\rho + O(t^{-1/2}).$$

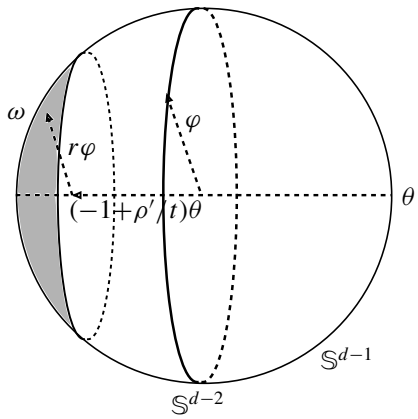

Figure 1. Change of variables, where $r = \sqrt{(\rho'/t)(2-\rho'/t)}$.

We next change the variables $\tau = \rho' - \rho$, $\eta = \rho' + \rho$, and write

$$\begin{aligned} J_1 &= \frac{(-1)^{d/2-1}}{\pi} \int_{-\infty}^{M} \int_{|\tau|}^{2M-\tau} \frac{G'_-(\tau - s, -\theta)}{\sqrt{\eta^2 - \tau^2}}\, d\eta\, d\tau + O(t^{-1/2}) \\ &= \frac{(-1)^{d/2-1}}{\pi} \int_{-\infty}^{M} G'_-(\tau - s, -\theta)\big[\ln(2M - \tau + \sqrt{4M^2 - 4M\tau}) - \ln|\tau|\big]\, d\tau + O(t^{-1/2}) \\ &= \frac{(-1)^{d/2-1}}{\pi} \int_{-R_1-R_2}^{R_1+R_2} G'_-(\tau - s, -\theta)\big[\ln(2M - \tau + \sqrt{4M^2 - 4M\tau}) - \ln|\tau|\big]\, d\tau + O(t^{-1/2}). \end{aligned}$$

The integrals above can be split into two parts:

$$\begin{aligned} I_1 &= \int_{-R_1-R_2}^{R_1+R_2} G'_-(\tau - s, -\theta)\big[\ln(2M - \tau + \sqrt{4M^2 - 4M\tau})\big]\, d\tau \\ &= \int_{-R_1-R_2}^{R_1+R_2} G'_-(\tau - s, -\theta)\big[\ln(2M - \tau + \sqrt{4M^2 - 4M\tau}) - \ln(4M)\big]\, d\tau \\ &= \int_{-R_1-R_2}^{R_1+R_2} G'_-(\tau - s, -\theta)\, O\Big(\frac{1}{M}\Big)\, d\tau = O\Big(\frac{1}{M}\Big) \end{aligned}$$

and

$$\begin{aligned} I_2 &= -\lim_{\varepsilon \to 0^+} \int_{\varepsilon < |\tau| < R_1+R_2} G'_-(\tau - s, -\theta) \ln|\tau|\, d\tau \\ &= \lim_{\varepsilon \to 0^+} \int_{\varepsilon < |\tau| < R_1+R_2} \frac{G_-(\tau - s, -\theta)}{\tau}\, d\tau = -\pi(\mathcal{H}G_-)(-s, -\theta). \end{aligned}$$

In summary we have

$$J_1 = (-1)^{d/2}(\mathcal{H}G_-)(-s, -\theta) + O\Big(\frac{1}{M}\Big) + O(t^{-1/2}).$$

Now we may combine J_1 and J_2

$$(t+s)^{(d-1)/2}\partial_t u((t+s)\theta, t) = (-1)^{d/2}(\mathcal{H}G_-)(-s, -\theta) + O\Big(\frac{1}{M}\Big) + O(t^{-1/2}).$$

Because the implicit constants in O's do not depend on $s \in [-R_2, R_2]$ or $\theta \in \mathbb{S}^{d-1}$, we may let $t \to +\infty$ then $M \to +\infty$ to conclude

$$\lim_{t\to+\infty} \int_{-R_2}^{R_2} \int_{\mathbb{S}^{d-1}} |(t+s)^{(d-1)/2} \partial_t u((t+s)\theta, t) - (-1)^{d/2} (\mathcal{H}G_-)(-s, -\theta)|^2 \, d\theta \, ds = 0. \qquad \square$$

4. Radial weakly nonradiative solutions

In this section we prove Proposition 1.9. First of all, we briefly show that any initial data in $P_{\mathrm{rad}}(R)$ leads to an R-weakly nonradiative solution. By linearity we only need to consider the case $(u_0, u_1) = (r^{2k_1-d}, 0)$ or $(u_0, u_1) = (0, r^{2k_2-d})$. If $(u_0, u_1) = (r^{2k_1-d}, 0)$, then a basic calculation shows that if we choose $C_1, C_2, \dots, C_{k_1-1}$ inductively, the solution

$$u_{k_1}(x,t) = \frac{1}{|x|^{d-2k_1}} + \frac{C_1 t^2}{|x|^{d-2k_1+2}} + \cdots + \frac{C_{k_1-1} t^{2k_1-2}}{|x|^{d-2}}$$

solves the linear wave equation with initial data $(|x|^{2k_1-d}, 0)$ in the region $\mathbb{R}^d \setminus \{0\}$. By finite speed of propagation, we have

$$\boldsymbol{S}_L(u_0, u_1)(x,t) = u_{k_1}(x,t), \qquad |x| > R + |t|.$$

A simple calculation shows that this is indeed a nonradiative solution. The case $(u_0, u_1) = (0, r^{2k_2-d})$ can be dealt with in the same manner by considering the solution

$$u_{k_2}(x,t) = \frac{t}{|x|^{d-2k_1}} + \frac{C_1 t^3}{|x|^{d-2k_1+2}} + \cdots + \frac{C_{k_2-1} t^{2k_1-1}}{|x|^{d-2}}.$$

Thus it is sufficient to show initial data of any nonradiative solution are contained in the space $P_{\mathrm{rad}}(R)$. We first consider the odd dimensions.

4A. ***Odd dimensions.*** Assume that $u = \boldsymbol{S}_L(u_0, u_1)$ is a radial R-weakly nonradiative solution. Let $G_- = \boldsymbol{T}_-(u_0, u_1)$. By radial assumption G_- is independent of the angle $\omega \in \mathbb{S}^{d-1}$. Let us first consider smooth functions G_-. We may calculate ($r > R,\ e_1 = (1, 0, \dots, 0) \in \mathbb{R}^d$)

$$u_0(re_1) = (2\pi)^{-\mu} \int_{\mathbb{S}^{d-1}} G_-^{(\mu-1)}(r\omega_1)\, d\omega = \frac{\sigma_{d-2}}{(2\pi)^\mu} \int_{-1}^{1} G_-^{(\mu-1)}(r\omega_1)(1-\omega_1^2)^{\mu-1}\, d\omega_1.$$

Here ω_1 is the first variable of $\mathbb{R}^d \supset \mathbb{S}^{d-1}$ and σ_{d-2} is the area of the sphere $\mathbb{S}^{d-2}$. We may integrate by parts and rescale:

$$\begin{aligned} u_0(re_1) &= \frac{(-1)^{\mu-1}\sigma_{d-2}}{(2\pi)^\mu r^{\mu-1}} \int_{-1}^{1} G_-(r\omega_1)[\partial_{\omega_1}^{\mu-1}(1-\omega_1^2)^{\mu-1}]\, d\omega_1 \\ &= \sum_{k=0}^{\lfloor(\mu-1)/2\rfloor} \frac{A_{d,k}}{r^{\mu-1}} \int_{-1}^{1} G_-(r\omega_1)\omega_1^{\mu-1-2k}\, d\omega_1 = \sum_{k=0}^{\lfloor(d-3)/4\rfloor} \frac{A_{d,k}}{r^{d-2-2k}} \int_{-R}^{R} G_-(s) s^{(d-3)/2-2k}\, ds \\ &= \sum_{k=1}^{\lfloor(d+1)/4\rfloor} \frac{A_{d,k}}{r^{d-2k}} \int_{-R}^{R} G_-(s) s^{(d+1)/2-2k}\, ds. \end{aligned}$$

Here the $A_{d,k}$ are nonzero constants. Similarly we have

$$\begin{aligned} u_1(re_1) &= (2\pi)^{-\mu}\int_{\mathbb{S}^{d-1}} G_-^{(\mu)}(r\omega_1)\,d\omega = \frac{\sigma_{d-2}}{(2\pi)^\mu}\int_{-1}^{1} G_-^{(\mu)}(r\omega_1)(1-\omega_1^2)^{\mu-1}\,d\omega_1 \\ &= \frac{(-1)^\mu\sigma_{d-2}}{(2\pi)^\mu r^\mu}\int_{-1}^{1} G_-(r\omega_1)[\partial_{\omega_1}^\mu(1-\omega_1^2)^{\mu-1}]\,d\omega_1 = \sum_{k=0}^{\lfloor(\mu-2)/2\rfloor}\frac{B_{d,k}}{r^\mu}\int_{-1}^{1} G_-(r\omega_1)\omega_1^{\mu-2-2k}\,d\omega_1 \\ &= \sum_{k=1}^{\lfloor(d-1)/4\rfloor}\frac{B_{d,k}}{r^{d-2k}}\int_{-R}^{R} G_-(s)s^{(d-1)/2-2k}\,ds. \end{aligned}$$

Here the $B_{d,k}$ are nonzero constants. Since smooth functions are dense in $L^2([-R,R])$, we have:

Proposition 4.1. *There exist constants $\{A_{d,k}\}_{1\le k\le\lfloor(d+1)/4\rfloor}$, $\{B_{d,k}\}_{1\le k\le\lfloor(d-1)/4\rfloor}$, so that for any $G_-\in L^2(\mathbb{R})$ supported in $[-R,R]$, the initial data $(u_0,u_1)=\boldsymbol{T}_-^{-1}G_-$ satisfy $(r>R)$*

$$\begin{aligned} u_0(r) &= \sum_{k=1}^{\lfloor(d+1)/4\rfloor}\left(A_{d,k}\int_{-R}^{R} G_-(s)s^{(d+1)/2-2k}\,ds\right)r^{-d+2k}, \\ u_1(r) &= \sum_{k=1}^{\lfloor(d-1)/4\rfloor}\left(B_{d,k}\int_{-R}^{R} G_-(s)s^{(d-1)/2-2k}\,ds\right)r^{-d+2k}. \end{aligned}$$

This clearly shows that if $u=\boldsymbol{S}_L(u_0,u_1)$ is a radial R-weakly nonradiative solution, then $(u_0,u_1)\in P_{\mathrm{rad}}(R)$.

4B. *Even dimensions.* The even dimensions involve the Hilbert transform, and thus are much more difficult to handle. The general idea is the same. If the initial data (u_0,u_1) are radial, then $G_\pm(s)=\boldsymbol{T}_\pm(u_0,u_1)$ is independent to the angle. We also have $G_+(s)=(-1)^{d/2}\mathcal{H}G_-(-s)$. Thus $\boldsymbol{S}_L(u_0,u_1)$ is R-weakly nonradiative if and only if G_- is contained in the space

$$\mathcal{P}_{\mathrm{rad}}=\{G_-\in L^2(\mathbb{R}): G_-(s)=0,\ s>R,\ (\mathcal{H}G_-)(s)=0,\ s<-R\}.$$

Now recall the operators $\mathcal{Q}$, $\mathcal{Q}'$ and $\mathcal{D}$ defined in Section 2C. We claim:

Lemma 4.2. $\mathcal{Q}'\mathcal{P}_{\mathrm{rad}}=\dot H_0^{1/2}(-R,R)$. *Here $\dot H_0^{1/2}(-R,R)$ is the completion of $C_0^\infty(-R,R)$ equipped with the $\dot H^{1/2}(\mathbb{R})$ norm.*

Proof. In order to avoid technical difficulties, we use an approximation technique. Given any $G_-\in\mathcal{P}_{\mathrm{rad}}$, we may utilize a local smoothing kernel to generate a sequence G_k, so that:

(a) $G_k\in\mathcal{P}_{\mathrm{rad}}(R+1/k)$.

(b) $G_k\in H^n(\mathbb{R})$ for all $n\ge 0$ and thus $G_k\in C^\infty(\mathbb{R})$.

(c) G_k converges to G_- in $L^2(\mathbb{R})$.

Let us consider the properties of the function $g_k=\mathcal{Q}'G_k\in C^\infty(\mathbb{R})$. According to part (a), $G_k(s)=0$ if $s>R+1/k$. We may use the convolution expression of $\mathcal{Q}'$ to obtain that g_k vanishes in the interval $(R+1/k,+\infty)$. Similarly $g_k=\mathcal{Q}\mathcal{H}G_k$ vanishes in the interval $(-\infty,-R-1/k)$. We recall that

$\mathcal{Q}' : L^2(\mathbb{R}) \to \dot{H}^{1/2}(\mathbb{R})$ is an isometry up to a constant. Thus $g_k \to g = \mathcal{Q}' G_-$ in $\dot{H}^{1/2}(\mathbb{R})$. This verifies $g \in \dot{H}_0^{1/2}(-R, R)$. We also need to show that given any $g \in \dot{H}_0^{1/2}(-R, R)$, then $\mathcal{Q}'^{-1} g \in \mathcal{P}_{\mathrm{rad}}$. It is sufficient to consider $g \in C_0^\infty(-R, R)$ by smooth approximation. A simple calculation of Fourier symbols shows that $\mathcal{Q}'^{-1} = -\mathcal{Q}'\mathcal{D}$ and $\mathcal{H}\mathcal{Q}'^{-1} = \mathcal{Q}\mathcal{D}$. A combination of these identities with the convolution expressions of $\mathcal{Q}$ and $\mathcal{Q}'$ immediately verifies $\mathcal{Q}'^{-1} g \in \mathcal{P}_{\mathrm{rad}}$. □

We also need to use the following explicit formula of $\boldsymbol{T}_-$ for radial data.

Lemma 4.3. *Assume $G \in C^\infty(\mathbb{R})$ so that $|G(s)| \lesssim |s|^{-3/2}$ for $|s| \gg 1$. Then the corresponding radial free wave $u = \boldsymbol{S}_L \boldsymbol{T}_-^{-1} G$ satisfies*

$$u(r, t) = C(d) \cdot r^{1-d/2} \int_{-1}^{1} \mathcal{Q}G(r\omega_1 + t) P_d(w_1)(1 - w_1^2)^{-1/2}\, d\omega_1. \tag{17}$$

Here P_d is an even or odd polynomial of degree $d/2 - 1$ defined by

$$\left(\frac{\partial}{\partial w_1}\right)^{d/2-1} (1 - w_1^2)^{(d-3)/2} = P_d(w_1)(1 - w_1^2)^{-1/2}.$$

Proof. If $G \in C_0^\infty(\mathbb{R})$, we use polar coordinates and integrate by parts:

$$\begin{aligned}
u(r, t) &= C(d) \int_0^\infty \int_{\mathbb{S}^{d-1}} \frac{G^{(d/2-1)}(r\omega_1 - \rho + t)}{\sqrt{\rho}}\, d\omega\, d\rho \\
&= C(d) \int_0^\infty \int_{-1}^{1} \frac{G^{(d/2-1)}(r\omega_1 - \rho + t)}{\sqrt{\rho}} (1 - w_1^2)^{(d-3)/2}\, d\omega_1\, d\rho \\
&= C(d) \cdot r^{1-d/2} \int_0^\infty \int_{-1}^{1} \frac{G(r\omega_1 - \rho + t)}{\sqrt{\rho}} P_d(w_1)(1 - w_1^2)^{-1/2}\, d\omega_1\, d\rho \\
&= C(d) \cdot r^{1-d/2} \int_{-1}^{1} \mathcal{Q}G(r\omega_1 + t) P_d(w_1)(1 - w_1^2)^{-1/2}\, d\omega_1.
\end{aligned}$$

This verifies the formula if $G \in C_0^\infty(\mathbb{R})$. In order to deal with profile G without compact support, we use standard smooth cut-off techniques. More precisely, we may choose $G_k \in C_0^\infty(\mathbb{R})$ so that $G_k \to G$ in $L^2(\mathbb{R})$ and

$$|G_k(s) - G(s)| = 0, \quad s < k, \qquad |G_k(s) - G(s)| \lesssim |s|^{-3/2}, \quad s \geq k.$$

Thus we have $\|\mathcal{Q}G - \mathcal{Q}G_k\|_{L^\infty} \lesssim 1/k$. This means we have the uniform convergence for all (r, t) in any compact subset of $\mathbb{R}^+ \times \mathbb{R}$:

$$\begin{aligned}
u_k(r, t) &= \frac{C(d)}{r^{d/2-1}} \int_{-1}^{1} \mathcal{Q}G_k(r\omega_1 + t) P_d(w_1)(1 - w_1^2)^{-1/2}\, d\omega_1 \\
&\rightrightarrows \frac{C(d)}{r^{d/2-1}} \int_{-1}^{1} \mathcal{Q}G(r\omega_1 + t) P_d(w_1)(1 - w_1^2)^{-1/2}\, d\omega_1.
\end{aligned}$$

Combining this with the convergence $u_k \to u$ in $\dot{H}^1$ we finish the proof. □

Remark 4.4. If $d \geq 4$ and $G \in L^2(\mathbb{R})$, then formula (17) still holds. This follows standard smooth approximation and/or cut-off techniques. Let $G_k \in C_0^\infty(\mathbb{R})$ so that $G_k \to G$ in $L^2(\mathbb{R})$. Thus $\mathcal{Q}G_k \to \mathcal{Q}G$ in $\dot{H}^{1/2}(\mathbb{R})$. Finally we observe the fact $P_d(w_1)(1-w_1^2)^{-1/2} \in \dot{H}^{-1/2}(\mathbb{R})$, obtain a locally uniform convergence $u_k(r,t) \to u(r,t)$ and conclude the proof.

Now we are ready to give an expression of $u = \boldsymbol{S}_L \boldsymbol{T}_-^{-1} G_-$ when $G_- \in \mathcal{P}_{\mathrm{rad}}(R)$.

Lemma 4.5. *Assume $G_- \in \mathcal{P}_{\mathrm{rad}}(R)$. Then the following identity holds*:

$$u(r,t) = \frac{C(d)}{r^{d/2}} \int_{-R}^{R} \mathcal{Q}' G_-(s) W_d\left(\frac{s-t}{r}\right) ds.$$

Here $W_d(\sigma)$ is the Hilbert transform (the function below is understood as zero if $|w_1| > 1$)

$$W_d(\sigma) \doteq \mathcal{H}\left(\frac{P_d(w_1)}{\sqrt{1-w_1^2}}\right) = \mathcal{H}\left[\left(\frac{d}{dw_1}\right)^{d/2-1} (1-w_1^2)^{(d-3)/2}\right].$$

Proof. By Lemma 4.2, we have $\mathcal{Q}'G_- \in \dot{H}_0^{1/2}(-R,R)$. We claim that it is sufficient to consider the case $\mathcal{Q}'G_- \in C_0^\infty(-R,R)$. In fact, we may choose $G_k \in \mathcal{P}_{\mathrm{rad}}(R)$ so that $\mathcal{Q}'G_k \in C_0^\infty(-R,R)$ so that

$$\mathcal{Q}'G_k \to \mathcal{Q}'G_- \quad \text{in } \dot{H}^{1/2}(-R,R) \quad \Longrightarrow \quad G_k \to G_- \quad \text{in } L^2(\mathbb{R}).$$

Now we observe a few important facts: we have the embedding $\dot{H}_0^{1/2}(-R,R) \hookrightarrow L^p(-R,R)$ for all $1 \leq p < +\infty$ and

$$\frac{P_d(w_1)}{\sqrt{1-w_1^2}} \in L^p(\mathbb{R}) \quad \Longrightarrow \quad W_d(\sigma) \in L^p(\mathbb{R}), \quad p \in (1,2).$$

As a result, if the identity

$$u_k(r,t) = \frac{C(d)}{r^{d/2}} \int_{-R}^{R} \mathcal{Q}' G_k(s) W_d\left(\frac{s-t}{r}\right) ds, \quad k \geq 1,$$

holds, then we may make $k \to +\infty$ in the identity above and verify that a similar identity holds for u and G_-. In fact the left-hand side converges in the space $\dot{H}^1(\mathbb{R}^d)$ for any given time t, while the right-hand side converges uniformly for (r,t) in any compact subset of $\mathbb{R}^+ \times \mathbb{R}$. Now we assume $g = \mathcal{Q}'G_- \in C_0^\infty(-R,R)$. Then $G_- = \mathcal{Q}'^{-1} g = -\mathcal{Q}'\mathcal{D}g$ satisfies the assumption of Lemma 4.3. As a result we have

$$\begin{aligned} u(r,t) &= C(d) \cdot r^{1-d/2} \int_{-1}^{1} \mathcal{Q}\mathcal{Q}'\mathcal{D}g(r\omega_1 + t) P_d(w_1)(1-w_1^2)^{-1/2}\, d\omega_1 \\ &= C(d) \cdot r^{1-d/2} \int_{-1}^{1} \mathcal{H}g(r\omega_1 + t) P_d(w_1)(1-w_1^2)^{-1/2}\, d\omega_1 \\ &= \frac{C(d)}{r^{d/2-1}} \int_{-\infty}^{\infty} g(r\sigma + t) W_d(\sigma)\, d\sigma. \end{aligned}$$

Here we use the facts $\mathcal{Q}\mathcal{Q}'\mathcal{D} = \mathcal{H}$ and

$$\int \mathcal{H}f \cdot \overline{\mathcal{H}g}\, dx = \int f \cdot \bar{g}\, dx, \quad \mathcal{H}(\mathcal{H}g(r\omega_1 + t))(\sigma) = (\mathcal{H}^2 g)(r\sigma + t) = -g(r\sigma + t).$$

Finally we apply change of variables $s = r\sigma + t$, recall the support of g and finish the proof. □

Now let us consider the Hilbert transform W_d. The key observation is the following technical lemma. This result has been known for many years; see [Solmon 1987], for instance. But we still give a brief proof in the Appendix for the purpose of completeness.

Lemma 4.6. *Assume that $P(x)$ is a polynomial of degree κ. Let W be the Hilbert transform*

$$W = \mathcal{H}\left(\frac{P(x)}{\sqrt{1-x^2}}\right).$$

Then $W(\sigma)$ is equal to a polynomial of degree $\kappa - 1$ if $\sigma \in (-1, 1)$. In particular, $W_2(\sigma) = 0$ for $\sigma \in (-1, 1)$; if $d \geq 4$, then the function $W_d(\sigma)$ is equal to an even or odd polynomial of degree $d/2 - 2$ in the interval $(-1, 1)$.

Proof of Proposition 1.11. According to Lemma 4.5, we have already obtained

$$u(r,t) = \frac{C(d)}{r^{d/2}} \int_{-R}^{R} \mathcal{Q}' G_-(s) W_d\left(\frac{s-t}{r}\right) ds.$$

Here $\mathcal{Q}' G_- \in \dot{H}_0^{1/2}(-R, R) \hookrightarrow L^p(-R, R)$ for all $1 < p < +\infty$. If we also have $r > |t| + R$, then

$$\left|\frac{s-t}{r}\right| < 1 \quad \text{for all } s \in (-R, R).$$

If $d = 2$, Lemma 4.6 immediately gives $u(r,t) \equiv 0$ if $r > |R| + t$ since we always have $W_2((s-t)/r) = 0$. In the higher-dimensional case $d \geq 4$, Lemma 4.6 guarantees that

$$W_d(s) = \sum_{l=1}^{\lfloor d/4 \rfloor} A_l s^{d/2-2l}, \quad -1 < s < 1,$$

is a polynomial. We plug this in the expression of u and obtain

$$u(r,t) = C(d) \sum_{l=1}^{\lfloor d/4 \rfloor} \frac{A_l}{r^{d-2l}} \int_{-R}^{R} \mathcal{Q}' G_-(s)(s-t)^{d/2-2l}\, ds, \quad r > R + |t|. \tag{18}$$

This immediately gives $(u_0, u_1) \in P_{\text{rad}}(R)$.

5. Exterior energy estimates of even dimensions

In this section we prove Proposition 1.11. It suffices to consider the case $d = 4k$. The proof of $d = 4k+2$ is almost the same. Again we switch to the space of radiation profiles $G_- \in L^2(\mathbb{R} \times \mathbb{S}^{d-1})$. We start with:

Lemma 5.1. *The image of radial data in the form of $(u_0, 0)$ can be characterized by*

$$\begin{aligned}\{\boldsymbol{T}_-(u_0, 0) : u_0 \in \dot{H}^1_{\text{rad}}(\mathbb{R}^d)\} &= \{G_- \in L^2(\mathbb{R}) : \mathcal{H}G_-(-s) = -G_-(s)\} \\ &= \left\{\frac{G(s) - \mathcal{H}G(-s)}{2} : G \in L^2(\mathbb{R})\right\}.\end{aligned}$$

Proof. First of all, if $u_0 \in \dot{H}^1_{\text{rad}}(\mathbb{R}^d)$, then the free wave $u = \boldsymbol{S}_L(u_0, u_1)$ is radial and satisfies

$$u(x,t) = u(x,-t), \quad u_t(x,t) = -u_t(x,-t).$$

Therefore G_-, G_+ are radial, i.e., independent of ω and satisfy $G_+(s) = -G_-(s)$. We may apply Theorem 1.5 and obtain $G_+(s) = \mathcal{H}G_-(-s)$. As a result, G_- satisfies the identity $\mathcal{H}G_-(-s) = -G_-(s)$. Next, let us assume G_- satisfies this identity. Then we have

$$G_-(s) = \frac{G_-(s) - \mathcal{H}G_-(-s)}{2} \in \left\{ \frac{G(s) - \mathcal{H}G(-s)}{2} : G \in L^2(\mathbb{R}) \right\}.$$

Finally, if $G_-(s) = (G(s) - \mathcal{H}G(-s))/2$, we show there exists $u_0 \in \dot{H}^1_{\mathrm{rad}}(\mathbb{R}^d)$, so that $G_- = \boldsymbol{T}_-(u_0, 0)$. In fact, we consider radial initial data $(u_0, u_1) = \boldsymbol{T}_-^{-1}G$ and free wave $u = \boldsymbol{S}_L(u_0, u_1)$. We may reverse the time and obtain $u(x, -t) = \boldsymbol{S}_L(u_0, -u_1)(x, t)$. Thus

$$\boldsymbol{T}_-(u_0, -u_1)(s) = -\boldsymbol{T}_+(u_0, u_1)(s) = -\mathcal{H}G(-s).$$

Therefore we have

$$\boldsymbol{T}_-(2u_0, 0)(s) = \boldsymbol{T}_-(u_0, u_1) + \boldsymbol{T}_-(u_0, -u_1) = G(s) - \mathcal{H}G(-s) = 2G_-(s),$$

which completes the proof. □

The key observation is the following:

Lemma 5.2. *Given $g \in L^2(\mathbb{R}^+)$, there exists a function G with $\|G\|_{L^2(\mathbb{R})} \le 2\|g\|_{L^2(\mathbb{R}^+)}$ so that*

$$G(s) - \mathcal{H}G(-s) = 2g(s), \quad s > 0, \qquad \left\| \frac{G(s) - \mathcal{H}G(-s)}{2} \right\|_{L^2(\mathbb{R})} \le \sqrt{2}\|g\|_{L^2(\mathbb{R}^+)}.$$

Proof. Let us first find a function G with $\|G\|_{L^2(\mathbb{R})} \le 2\|g\|_{L^2}$ so that

$$G(s) - \frac{G(s) + \mathcal{H}G(-s)}{2} = g(s), \quad s > 0.$$

We define a linear bounded operator $\boldsymbol{T}$ from $L^2(\mathbb{R}^+)$ to itself. In the formula below we extend the domain of G to $\mathbb{R}$ by assuming $G(s) = 0$ if $s < 0$ before we apply the Hilbert transform:

$$(\boldsymbol{T}G)(s) = \frac{G(s) + \mathcal{H}G(-s)}{2} = \frac{G(s)}{2} - \frac{1}{2\pi}\int_0^\infty \frac{G(\tau)}{s+\tau}\,d\tau, \quad s > 0.$$

We may further rewrite it as

$$\boldsymbol{T}G = \frac{G}{2} - \frac{1}{2\pi}\boldsymbol{L}^2 G.$$

Here $\boldsymbol{L}$ is the Laplace transform

$$\boldsymbol{L}G(s) = \int_0^\infty G(\tau)e^{-s\tau}\,d\tau,$$

which is self-adjoint operator in $L^2(\mathbb{R}^+)$ with an operator norm $\sqrt{\pi}$. More details about the Laplace transform can be found in [Lax 2002]. As a result, we have

$$\begin{aligned}
\|\boldsymbol{T}G\|^2_{L^2(\mathbb{R}^+)} &= \tfrac{1}{4}\langle G - \tfrac{1}{\pi}\boldsymbol{L}^2 G, G - \tfrac{1}{\pi}\boldsymbol{L}^2 G\rangle \\
&= \tfrac{1}{4}\|G\|^2_{L^2} + \tfrac{1}{4\pi^2}\|\boldsymbol{L}^2 G\|^2_{L^2} - \tfrac{1}{4\pi}\langle G, \boldsymbol{L}^2 G\rangle - \tfrac{1}{4\pi}\langle \boldsymbol{L}^2 G, G\rangle \\
&\le \tfrac{1}{4}\|G\|^2_{L^2} + \tfrac{1}{4\pi}\|\boldsymbol{L}G\|^2_{L^2} - \tfrac{1}{2\pi}\langle \boldsymbol{L}G, \boldsymbol{L}G\rangle = \tfrac{1}{4}\|G\|^2_{L^2} - \tfrac{1}{4\pi}\|\boldsymbol{L}G\|^2_{L^2}.
\end{aligned}$$

Thus the operator norm of $\boldsymbol{T}$ is less than or equal to $\frac{1}{2}$. This means that the function

$$G = \sum_{j=0}^{\infty} \boldsymbol{T}^j g \in L^2(\mathbb{R}^+)$$

satisfies the equation $G - \boldsymbol{T}G = g$ and $\|G\|_{L^2(\mathbb{R}^+)} \le 2\|g\|_{L^2(\mathbb{R}^+)}$. Finally we naturally extend the domain of G to $\mathbb{R}$ by defining $G(s) = 0$ if $s < 0$. We have

$$\frac{G(s) - \mathcal{H}G(-s)}{2} = \begin{cases} g(s), & s > 0, \\ -\frac{1}{2}\mathcal{H}G(-s), & s < 0. \end{cases}$$

Therefore we may find an upper bound of the L^2 norm

$$\left\| \frac{G(s) - \mathcal{H}G(-s)}{2} \right\|_{L^2(\mathbb{R})}^2 \le \|g\|_{L^2(\mathbb{R}^+)}^2 + \tfrac{1}{4}\|\mathcal{H}G\|_{L^2(\mathbb{R})}^2 \le 2\|g\|_{L^2(\mathbb{R}^+)}^2. \qquad \square$$

Proof of Proposition 1.11. Let $G_- = \boldsymbol{T}_-(u_0, 0)$ and $g(s)$ be its cut-off version:

$$g(s) = \begin{cases} G_-(s), & s > R, \\ 0, & s < R. \end{cases}$$

Then radiation field implies that the free wave $u = \boldsymbol{S}_L(u_0, 0)$ satisfies

$$\lim_{t\to-\infty} \int_{|x|>R+|t|} |\nabla u(x,t)|^2\, dx = \lim_{t\to-\infty} \int_{|x|>R+|t|} |u_t(x,t)|^2\, dx = \sigma_{4k-1}\|g\|_{L^2(\mathbb{R}^+)}^2. \tag{19}$$

Here again σ_{4k-1} is the area of the sphere $\mathbb{S}^{4k-1}$. According to Lemmas 5.1 and 5.2, there exists a function $\tilde{u}_0 \in \dot{H}^1_{\mathrm{rad}}(\mathbb{R}^{4k})$, so that

$$\boldsymbol{T}_-(\tilde{u}_0, 0)(s) = g(s), \quad s > 0, \qquad \|\tilde{u}_0\|_{\dot{H}^1(\mathbb{R}^{4k})}^2 \le 4\sigma_{4k-1}\|g\|_{L^2(\mathbb{R}^+)}^2.$$

Therefore $\boldsymbol{T}_-(u_0 - \tilde{u}_0, 0)$ vanishes if $s > R$. A combination of this fact with the time symmetry gives

$$\lim_{t\to\pm\infty} \int_{|x|>|t|+R} |\nabla_{t,x}\boldsymbol{S}_L(u_0 - \tilde{u}_0, 0)(x,t)|^2\, dx = 0.$$

As a result, we may apply Proposition 1.9 and conclude $u_0 - \tilde{u}_0 \in Q_k(R)$. This means

$$\|\Pi^{\perp}_{Q_k(R)} u_0\|_{\dot{H}^1(\{x:|x|>R\})}^2 \le \|\tilde{u}_0\|_{\dot{H}^1(\{x:|x|>R\})}^2 \le 4\sigma_{4k-1}\|g\|_{L^2(\mathbb{R}^+)}^2.$$

A combination of this inequality and identity (19) immediately verifies the conclusion of Proposition 1.11 in the negative time direction. The positive time direction follows the time symmetry.

6. Nonradial exterior energy estimates

In this section we give a short proof of Proposition 1.14. We start with:

Lemma 6.1. *Let $d \ge 3$ be an odd integer. Then*

$$\sum_{\pm} \lim_{t\to\pm\infty} \int_{|x|>R+|t|} |\nabla_{t,x}\boldsymbol{S}_L(u_0, u_1)(x,t)|^2\, dx = 2\int_{|s|>R} \int_{\mathbb{S}^{d-1}} |\boldsymbol{T}_-(u_0, u_1)(s,\theta)|^2\, d\theta\, ds. \tag{20}$$

In particular, we have (see (4) for the definition of $P(R)$)

$$\boldsymbol{T}_-(P(R)) = \mathcal{P}(R) \doteq \{G_- \in L^2(\mathbb{R}\times\mathbb{S}^{d-1}) : \operatorname{supp} G_- \subset [-R, R]\times\mathbb{S}^{d-1}\}.$$

Proof. Let u be the solution of linear wave equation with initial data (u_0, u_1). Then by radiation field (Theorem 1.1) we have

$$\lim_{t\to-\infty}\int_{|x|>|t|+R} |\nabla_{t,x}u|^2\,dx = 2\int_R^\infty\int_{\mathbb{S}^{d-1}} |G_-(s,\theta)|^2\,d\theta\,ds,$$
$$\lim_{t\to-\infty}\int_{|x|<|t|-R} |\nabla_{t,x}u|^2\,dx = 2\int_{-\infty}^{-R}\int_{\mathbb{S}^{d-1}} |G_-(s,\theta)|^2\,d\theta\,ds.$$

In addition, we may apply the energy conservation law, Proposition 1.2 and obtain

$$\begin{aligned}\lim_{t\to-\infty}\int_{|x|<|t|-R} |\nabla_{t,x}u|^2\,dx &= \int_{\mathbb{R}^d}(|\nabla u_0|^2+|u_1|^2)\,dx - \lim_{t\to-\infty}\int_{|x|>|t|-R}|\nabla_{t,x}u|^2\,dx\\ &= \lim_{t\to+\infty}\int_{|x|>t+R}|\nabla_{t,x}u|^2\,dx.\end{aligned}$$

Combining these identities we have

$$\sum_{\pm}\lim_{t\to\pm\infty}\int_{|x|>R+|t|}|\nabla_{t,x}u(x,t)|^2\,dx = 2\int_{|s|>R}\int_{\mathbb{S}^{d-1}}|G_-(s,\theta)|^2\,d\theta\,ds.$$

Finally $(u_0, u_1)\in P(R)$ is equivalent to saying

$$\int_{|s|>R}\int_{\mathbb{S}^{d-1}}|G_-(s,\theta)|^2\,d\theta\,ds = 0,$$

namely $\operatorname{supp} G_- \subset [-R, R]\times\mathbb{S}^{d-1}$. □

Now we are ready to prove Proposition 1.14. Since $\sqrt{2}\boldsymbol{T}_-$ is a bijective isometry from $\dot{H}^1\times L^2(\mathbb{R}^d)$ to $L^2(\mathbb{R}\times\mathbb{S}^{d-1})$, we have

$$\boldsymbol{\Pi}^\perp_{P(R)}(u_0,u_1) = \boldsymbol{T}_-^{-1}\boldsymbol{\Pi}^\perp_{\boldsymbol{T}_-(P(R))}\boldsymbol{T}_-(u_0,u_1).$$

We next use the expression of $\mathcal{P}(R) = \boldsymbol{T}_-(P(R))$:

$$\begin{aligned}\|\boldsymbol{\Pi}^\perp_{P(R)}(u_0,u_1)\|^2_{\dot{H}^1\times L^2} &= 2\|\boldsymbol{\Pi}^\perp_{\mathcal{P}(R)}\boldsymbol{T}_-(u_0,u_1)\|^2_{L^2(\mathbb{R}\times\mathbb{S}^{d-1})}\\ &= 2\int_{|s|>R}\int_{\mathbb{S}^{d-1}}|\boldsymbol{T}_-(u_0,u_1)(s,\theta)|^2\,d\theta\,ds.\end{aligned}$$

Combining this with (20) we finish the proof.

Appendix

In this section we give a brief proof of Lemma 4.6 for completeness. We first prove this lemma for two special cases, $P(x) = 1$ and $P(x) = 1 - x^2$. We start with $P(x) = 1$. A straightforward calculation gives

$$\begin{aligned}
\pi W(s) &= \text{p.v.} \int_{-1}^{1} \frac{(1-x^2)^{-1/2}}{s-x}\,dx \\
&= \text{p.v.} \int_{-1}^{1} \frac{(1-s^2)^{-1/2}}{s-x}\,dx + \int_{-1}^{1} \frac{(1-x^2)^{-1/2}-(1-s^2)^{-1/2}}{s-x}\,dx \\
&= (1-s^2)^{-1/2} \ln\left|\frac{1+s}{1-s}\right| + \int_{-1}^{1} \frac{(1-s^2)-(1-x^2)}{(s-x)\sqrt{1-x^2}\sqrt{1-s^2}(\sqrt{1-x^2}+\sqrt{1-s^2})}\,dx \\
&= (1-s^2)^{-1/2} \ln\left|\frac{1+s}{1-s}\right| + \frac{-s}{\sqrt{1-s^2}} \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}(\sqrt{1-x^2}+\sqrt{1-s^2})}\,dx.
\end{aligned}$$

Next we apply the change of variables $x = 2z/(1+z^2)$. We have

$$\sqrt{1-x^2} = \frac{1-z^2}{1+z^2}\,dx = \frac{2(1-z^2)}{(1+z^2)^2}\,dz.$$

Thus

$$\begin{aligned}
\int_{-1}^{1} \frac{1}{\sqrt{1-x^2}(\sqrt{1-x^2}+\sqrt{1-s^2})}\,dx &= \int_{-1}^{1} \frac{2\,dz}{1-z^2+\sqrt{1-s^2}(1+z^2)} \\
&= \frac{2}{s} \int_{0}^{1} \left(\frac{1}{(1+\sqrt{1-s^2})/s - z} + \frac{1}{(1+\sqrt{1-s^2})/s + z} \right) dz \\
&= \frac{2}{s} \ln\left| \left(\frac{1+\sqrt{1-s^2}}{s} + 1/\frac{1+\sqrt{1-s^2}}{s} - 1 \right) \right| = \frac{1}{s} \ln\left|\frac{1+s}{1-s}\right|. \qquad (21)
\end{aligned}$$

This immediately gives $W(s) = 0$ for $s \in (-1, 1)$. Next we consider the case $P(x) = 1 - x^2$. In this case we calculate the Hilbert transform of $\sqrt{1-x^2}$:

$$\begin{aligned}
\pi W(s) &= \text{p.v.} \int_{-1}^{1} \frac{\sqrt{1-x^2}}{s-x}\,dx \\
&= \text{p.v.} \int_{-1}^{1} \frac{\sqrt{1-s^2}}{s-x}\,dx + \int_{-1}^{1} \frac{\sqrt{1-x^2}-\sqrt{1-s^2}}{s-x}\,dx \\
&= \sqrt{1-s^2} \ln\left|\frac{1+s}{1-s}\right| + \int_{-1}^{1} \frac{(1-x^2)-(1-s^2)}{(s-x)(\sqrt{1-x^2}+\sqrt{1-s^2})}\,dx \\
&= \sqrt{1-s^2} \ln\left|\frac{1+s}{1-s}\right| + s \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}+\sqrt{1-s^2}}\,dx \\
&= \sqrt{1-s^2} \ln\left|\frac{1+s}{1-s}\right| + \pi s + s \int_{-1}^{1} \left(\frac{1}{\sqrt{1-x^2}+\sqrt{1-s^2}} - \frac{1}{\sqrt{1-x^2}} \right) dx \\
&= \sqrt{1-s^2} \ln\left|\frac{1+s}{1-s}\right| + \pi s - s\sqrt{1-s^2} \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}(\sqrt{1-x^2}+\sqrt{1-s^2})}\,dx = \pi s.
\end{aligned}$$

Here we use the integral (21) again.

Induction. Now we are ready to prove Lemma 4.6 by induction. It is clear that we only need to show the Hilbert transform of $f_\kappa(x) = x^\kappa (1-x^2)^{-1/2}$ is a polynomial of degree $\kappa - 1$ in the interval $(-1, 1)$. The

cases of $\kappa = 0, 2$ have been done. Now let us consider the case of $f_1(x) = x(1-x^2)^{-1/2}$. We observe that ($s \in (-1, 1)$)

$$\mathcal{H} f_1 = \mathcal{H}\frac{d}{dx}(-\sqrt{1-x^2}) = -\frac{d}{ds}\mathcal{H}(\sqrt{1-x^2}) = -1.$$

This proves the case $\kappa = 1$. Now let us assume that the cases $\kappa = 0, 1, 2, \dots, n$ are done and consider the case $\kappa = n+1$. Here $n \geq 2$. We have

$$x^{n+1}(1-x^2)^{-1/2} = -x^{n-1}(1-x^2)^{1/2} + x^{n-1}(1-x^2)^{-1/2}.$$

The Hilbert transform of the second term in the right-hand side is known to be a polynomial of degree $n-2$. Thus we only need to consider the first term. We have

$$\begin{aligned}\frac{d}{ds}\mathcal{H}(x^{n-1}(1-x^2)^{1/2}) &= \mathcal{H}\frac{d}{dx}(x^{n-1}(1-x^2)^{1/2}) \\ &= \mathcal{H}\{[-nx^n + (n-1)x^{n-2}](1-x^2)^{-1/2}\}.\end{aligned}$$

This is a polynomial of degree $n-1$ by induction hypothesis. A simple integration then finishes the proof of the case $\kappa = n+1$. Generally speaking, the derivative with respect to s as given above is in the weak sense. But since the derivative is known to be a polynomial in $(-1, 1)$, we can integrate as usual.

Acknowledgement

Shen is financially supported by National Natural Science Foundation of China, projects 12071339 and 11771325.

References

[Côte, Kenig and Schlag 2014] R. Côte, C. E. Kenig, and W. Schlag, "Energy partition for the linear radial wave equation", *Math. Ann.* **358**:3-4 (2014), 573–607. MR Zbl

[Duyckaerts, Kenig and Merle 2011] T. Duyckaerts, C. Kenig, and F. Merle, "Universality of blow-up profile for small radial type II blow-up solutions of the energy-critical wave equation", *J. Eur. Math. Soc.* **13**:3 (2011), 533–599. MR Zbl

[Duyckaerts, Kenig and Merle 2012] T. Duyckaerts, C. Kenig, and F. Merle, "Universality of the blow-up profile for small type II blow-up solutions of the energy-critical wave equation: the nonradial case", *J. Eur. Math. Soc.* **14**:5 (2012), 1389–1454. MR Zbl

[Duyckaerts, Kenig and Merle 2013] T. Duyckaerts, C. Kenig, and F. Merle, "Classification of radial solutions of the focusing, energy-critical wave equation", *Cambridge J. Math.* **1**:1 (2013), 75–144. MR Zbl

[Duyckaerts, Kenig and Merle 2014] T. Duyckaerts, C. Kenig, and F. Merle, "Scattering for radial, bounded solutions of focusing supercritical wave equations", *Int. Math. Res. Not.* **2014**:1 (2014), 224–258. MR Zbl

[Duyckaerts, Kenig and Merle 2019] T. Duyckaerts, C. Kenig, and F. Merle, "Scattering profile for global solutions of the energy-critical wave equation", *J. Eur. Math. Soc.* **21**:7 (2019), 2117–2162. MR Zbl

[Duyckaerts, Kenig and Merle 2021] T. Duyckaerts, C. Kenig, and F. Merle, "Decay estimates for nonradiative solutions of the energy-critical focusing wave equation", *J. Geom. Anal.* **31**:7 (2021), 7036–7074. MR Zbl

[Duyckaerts, Kenig and Merle 2023] T. Duyckaerts, C. Kenig, and F. Merle, "Soliton resolution for the radial critical wave equation in all odd space dimensions", *Acta Math.* **230**:1 (2023), 1–92. MR

[Duyckaerts, Kenig, Martel and Merle 2022] T. Duyckaerts, C. Kenig, Y. Martel, and F. Merle, "Soliton resolution for critical co-rotational wave maps and radial cubic wave equation", *Comm. Math. Phys.* **391**:2 (2022), 779–871. MR Zbl

[Evans 1998] L. C. Evans, *Partial differential equations*, Grad. Stud. Math. **19**, Amer. Math. Soc., Providence, RI, 1998. MR Zbl

[Friedlander 1962] F. G. Friedlander, "On the radiation field of pulse solutions of the wave equation", *Proc. Roy. Soc. London Ser. A* **269** (1962), 53–65. MR Zbl

[Friedlander 1973] F. G. Friedlander, "An inverse problem for radiation fields", *Proc. Lond. Math. Soc.* (3) **27**:3 (1973), 551–576. MR Zbl

[Friedlander 1980] F. G. Friedlander, "Radiation fields and hyperbolic scattering theory", *Math. Proc. Cambridge Philos. Soc.* **88**:3 (1980), 483–515. MR Zbl

[Ginibre, Soffer and Velo 1992] J. Ginibre, A. Soffer, and G. Velo, "The global Cauchy problem for the critical nonlinear wave equation", *J. Funct. Anal.* **110**:1 (1992), 96–130. MR Zbl

[Grillakis 1990] M. G. Grillakis, "Regularity and asymptotic behaviour of the wave equation with a critical nonlinearity", *Ann. of Math.* (2) **132**:3 (1990), 485–509. MR Zbl

[Grillakis 1992] M. G. Grillakis, "Regularity for the wave equation with a critical nonlinearity", *Comm. Pure Appl. Math.* **45**:6 (1992), 749–774. MR Zbl

[Kapitanski 1994] L. Kapitanski, "Weak and yet weaker solutions of semilinear wave equations", *Comm. Partial Differential Equations* **19**:9-10 (1994), 1629–1676. MR Zbl

[Katayama 2013] S. Katayama, "Asymptotic behavior for systems of nonlinear wave equations with multiple propagation speeds in three space dimensions", *J. Differential Equations* **255**:1 (2013), 120–150. MR Zbl

[Kenig and Merle 2008] C. E. Kenig and F. Merle, "Global well-posedness, scattering and blow-up for the energy-critical focusing non-linear wave equation", *Acta Math.* **201**:2 (2008), 147–212. MR Zbl

[Kenig, Lawrie and Schlag 2014] C. E. Kenig, A. Lawrie, and W. Schlag, "Relaxation of wave maps exterior to a ball to harmonic maps for all data", *Geom. Funct. Anal.* **24**:2 (2014), 610–647. MR Zbl

[Kenig, Lawrie, Liu and Schlag 2015] C. Kenig, A. Lawrie, B. Liu, and W. Schlag, "Channels of energy for the linear radial wave equation", *Adv. Math.* **285** (2015), 877–936. MR Zbl

[Lax 2002] P. D. Lax, *Functional analysis*, Wiley, New York, 2002. MR Zbl

[Lindblad and Sogge 1995] H. Lindblad and C. D. Sogge, "On existence and scattering with minimal regularity for semilinear wave equations", *J. Funct. Anal.* **130**:2 (1995), 357–426. MR Zbl

[Nakanishi 1999a] K. Nakanishi, "Scattering theory for the nonlinear Klein–Gordon equation with Sobolev critical power", *Int. Math. Res. Not.* **1999**:1 (1999), 31–60. MR Zbl

[Nakanishi 1999b] K. Nakanishi, "Unique global existence and asymptotic behaviour of solutions for wave equations with non-coercive critical nonlinearity", *Comm. Partial Differential Equations* **24**:1-2 (1999), 185–221. MR Zbl

[Shatah and Struwe 1993] J. Shatah and M. Struwe, "Regularity results for nonlinear wave equations", *Ann. of Math.* (2) **138**:3 (1993), 503–518. MR Zbl

[Shatah and Struwe 1994] J. Shatah and M. Struwe, "Well-posedness in the energy space for semilinear wave equations with critical growth", *Int. Math. Res. Not.* **1994**:7 (1994), 303–309. MR Zbl

[Shen 2013] R. Shen, "On the energy subcritical, nonlinear wave equation in $\mathbb{R}^3$ with radial data", *Anal. PDE* **6**:8 (2013), 1929–1987. MR Zbl

[Solmon 1987] D. C. Solmon, "Asymptotic formulas for the dual Radon transform and applications", *Math. Z.* **195**:3 (1987), 321–343. MR Zbl

Received 28 Nov 2021. Accepted 11 Aug 2022.

LIANG LI: 17864193561@163.com
Center for Applied Mathematics, Tianjin University, Tianjin, China

RUIPENG SHEN: srpgow@163.com
Center for Applied Mathematics, Tianjin University, Tianjin, China

LIJUAN WEI: lijuanwei8@163.com
Center for Applied Mathematics, Tianjin University, Tianjin, China

ANALYSIS AND PDE
Vol. 17 (2024), No. 2, pp. 749–756
DOI: 10.2140/apde.2024.17.749

ON L^∞ ESTIMATES FOR MONGE–AMPÈRE AND HESSIAN EQUATIONS ON NEF CLASSES

BIN GUO, DUONG H. PHONG, FREID TONG AND CHUWEN WANG

The PDE approach developed earlier by the first three authors for L^∞ estimates for fully nonlinear equations on Kähler manifolds is shown to apply as well to Monge–Ampère and Hessian equations on nef classes. In particular, one obtains a new proof of the estimates of Boucksom, Eyssidieux, Guedj and Zeriahi (2010) and Fu, Guo and Song (2020) for the Monge–Ampère equation, together with their generalization to Hessian equations.

1. Introduction

The goal of this short note is to show that the PDE approach introduced in [Guo et al. 2023a; 2023b] for L^∞ and Trudinger-type estimates for general classes of fully nonlinear equations on a compact Kähler manifold applies as well to Monge–Ampère and Hessian equations on nef classes.

The key to the approach in [Guo et al. 2023a; 2023b] is an estimate of Trudinger-type, obtained by comparing the solution φ of the given equation to the solution of an auxiliary Monge–Ampère equation with the energy of the sublevel set function $-\varphi+s$ on the right-hand side. We shall see that, in the present case of nef classes, the argument can still be made to work by replacing φ by $\varphi-V$, where V is the envelope of the nef class. Applied to the Monge–Ampère equation, this gives a PDE proof of the estimates obtained earlier for nef classes by Boucksom, Eyssidieux, Guedj and Zeriahi [Boucksom et al. 2010] and Fu, Guo and Song [Fu et al. 2020]. The estimates which we obtain with this method applied to Hessian equations seem new.

We note that the use of an auxiliary Monge–Ampère equation was instrumental in the recent progress of Chen and Cheng [2021] on the constant scalar curvature Kähler metrics problem. There the auxiliary equation involved the entropy, and not the energy, of sublevel set functions as in our case. More generally, auxiliary equations have often been used in the theory of partial differential equations, notably by De Giorgi [1961] and more recently by Dinew and Kołodziej [Demailly et al. 2014; Dinew and Kołodziej 2014] in their approach to Hölder estimates for the complex Monge–Ampère equation.

This work was supported in part by the National Science Foundation under grant DMS-1855947. Tong is supported by Harvard's Center for Mathematical Sciences and Applications.
MSC2020: primary 53C56; secondary 34G20.
Keywords: Monge–Ampere equations, Hessian equations.

2. The Monge–Ampère equation

We begin with the Monge–Ampère equation. Let (X, ω) be a compact Kähler manifold, and, without loss of generality, let us assume that $\int_X \omega^n = 1$. Let χ be a closed $(1,1)$-form on X. We assume the cohomology class $[\chi]$ is nef and let $\nu \in \{0, 1, \dots, n\}$ be the numerical dimension of $[\chi]$, i.e.,

$$\nu = \max\{k \mid [\chi]^k \neq 0 \text{ in } H^{k,k}(X, \mathbb{C})\}.$$

When $\nu = n$ we say the class $[\chi]$ is *big*.

Let $\hat{\omega}_t = \chi + t\omega$ for $t \in (0, 1]$. The form $\hat{\omega}_t$ may not be positive but its class is Kähler. We consider the family of complex Monge–Ampère equations

$$(\hat{\omega}_t + i\partial\bar{\partial}\varphi_t)^n = c_t e^F \omega^n, \quad \sup_X \varphi_t = 0, \tag{2-1}$$

where $c_t = [\hat{\omega}_t^n] = O(t^{n-\nu})$ is a normalizing constant and $F \in C^\infty(X)$ satisfies $\int_X e^F \omega^n = \int_X \omega^n$. This equation admits a unique smooth solution φ_t by Yau's theorem [1978].

The form χ is not assumed to be semipositive, so the usual L^∞ estimate of φ_t may not hold [Kołodziej 1998]. As in [Boucksom et al. 2010; Fu et al. 2020], we need to modify the solution φ_t by an envelope V_t of the class $[\hat{\omega}_t]$, defined as

$$V_t = \sup\{v \mid v \in \mathrm{PSH}(X, \hat{\omega}_t),\ v \leq 0\}.$$

Then we have:

Theorem 1. *Consider (2-1), and assume that the cohomology class of χ is nef. For any $s > 0$, let $\Omega_s = \{\varphi_t - V_t \leq -s\}$ be the sublevel set of $\varphi_t - V_t$.*

(a) *There are constants $C = C(n, \omega, \chi) > 0$ and $\alpha_0 = \alpha_0(n, \omega, \chi) > 0$ such that*

$$\int_{\Omega_s} \exp\left\{\alpha_0 \left(\frac{-(\varphi_t - V_t + s)}{A_s^{1/(1+n)}}\right)^{(n+1)/n}\right\} \omega^n \leq C \exp(CE_t), \tag{2-2}$$

where $A_s = \int_{\Omega_s} (-\varphi_t + V_t - s) e^F \omega^n$ and $E_t = \int_X (-\varphi_t + V_t) e^F \omega^n$.

(b) *Fix $p > n$. There is a constant $C(n, p, \omega, \chi, \|e^F\|_{L^1(\log L)^p})$ such that, for all $t \in (0, 1]$, we have*

$$0 \leq -\varphi_t + V_t \leq C(n, p, \omega, \chi, \|e^F\|_{L^1(\log L)^p}). \tag{2-3}$$

We remark that the estimates in Theorem 1 continue to hold for a family of Kähler metrics (maybe with distinct complex structures) which satisfy a uniform α-invariant-type estimate.

Proof. We would like to find an auxiliary equation with smooth coefficients, so that its solvability can be guaranteed by Yau's theorem. For this, we need a lemma due to Berman [2019] on a smooth approximation for V_t; see also Lemma 4 below. Fix a time $t \in (0, 1]$.

Lemma 2. *Let u_β be the smooth solution to the complex Monge–Ampère equation*

$$(\hat{\omega}_t + i\partial\bar{\partial}u_\beta)^n = e^{\beta u_\beta}\omega^n.$$

Then u_β converges uniformly to V_t as $\beta \to \infty$.

We remark that by [Chu et al. 2018], V_t is a $C^{1,1}$ function on X, although this fact is not used in this note. We now return to the proof of Theorem 1(a).

We choose a sequence of smooth positive functions $\tau_k : \mathbb{R} \to \mathbb{R}_+$ such that $\tau_k(x)$ decreases to $x \cdot \chi_{\mathbb{R}_+}(x)$ as $k \to \infty$. Fix a smooth function u_β as in Lemma 2. The function u_β depends on t, but for simplicity we omit the subscript t. We solve the following auxiliary Monge–Ampère equation on X,

$$(\hat{\omega}_t + i\partial\bar{\partial}\psi_{t,k})^n = c_t \frac{\tau_k(-\varphi_t + u_\beta - s)}{A_{s,k,\beta}} e^F \omega^n, \quad \sup_X \psi_{t,k} = 0, \tag{2-4}$$

where

$$A_{s,k,\beta} = \int_X \tau_k(-\varphi_t + u_\beta - s) e^F \omega^n.$$

Since $\psi_{t,k} \le V_t$ and u_β converges uniformly to V_t, by taking β large enough, we may assume $\psi_{t,k} < u_\beta + 1$.

Define a function

$$\Phi = -\varepsilon(-\psi_{t,k} + u_\beta + 1 + \Lambda)^{n/(n+1)} - (\varphi_t - u_\beta + s),$$

with the constants

$$\varepsilon^{n+1} = A_{s,k,\beta} n^{-n}(n+1)^n, \quad \Lambda = n^{n+1}(n+1)^{-n-1}\varepsilon^{n+1}. \tag{2-5}$$

As a smooth function on the compact manifold X, we know Φ must achieve its maximum at some $x_0 \in X$. If $x_0 \in X \backslash \Omega_s^\circ$, then

$$\Phi(x_0) \le -(\varphi_t - u_\beta + s) \le -V_t + u_\beta \le \epsilon_\beta,$$

where $\epsilon_\beta \to 0$ as $\beta \to \infty$. On the other hand, if $x_0 \in \Omega_s^\circ$, we calculate (Δ_t denotes the Laplacian with respect to the metric $\omega_t = \hat{\omega}_t + i\partial\bar{\partial}\varphi_t$)

$$\begin{aligned}
0 &\ge \Delta_t \Phi(x_0) \\
&= -\varepsilon \frac{n}{n+1}(-\psi_{t,k} + u_\beta + \Lambda + 1)^{-1/(n+1)} \operatorname{tr}_{\omega_t}(-i\partial\bar{\partial}\psi_{t,k} + i\partial\bar{\partial}u_\beta) - \operatorname{tr}_{\omega_t}(i\partial\bar{\partial}\varphi_t - i\partial\bar{\partial}u_\beta) \\
&\qquad + \frac{n\varepsilon}{(n+1)^2}(-\psi_{t,k} + u_\beta + 1 + \Lambda)^{-(n+2)/(n+1)} \operatorname{tr}_{\omega_t} i\partial(\psi_{t,k} - u_\beta) \wedge \bar{\partial}(\psi_{t,k} - u_\beta) \\
&\ge \frac{n\varepsilon}{n+1}(-\psi_{t,k} + u_\beta + \Lambda + 1)^{-1/(n+1)} \operatorname{tr}_{\omega_t}(\hat{\omega}_{t,\psi_{t,k}} - \hat{\omega}_{t,u_\beta}) - n + \operatorname{tr}_{\omega_t} \hat{\omega}_{t,u_\beta} \\
&\ge \frac{n\varepsilon}{n+1}(-\psi_{t,k} + u_\beta + \Lambda + 1)^{-1/(n+1)} n \left(\frac{\hat{\omega}^n_{t,\psi_{t,k}}}{\omega_t^n}\right)^{1/n} \\
&\qquad - n + \left(1 - \frac{n\varepsilon}{n+1}(-\psi_{t,k} + u_\beta + \Lambda + 1)^{-1/(n+1)}\right) \operatorname{tr}_{\omega_t} \hat{\omega}_{t,u_\beta} \\
&\ge \frac{n^2\varepsilon}{n+1}(-\psi_{t,k} + u_\beta + \Lambda + 1)^{-1/(n+1)} (\tau_k(-\varphi_t + u_\beta - s) A^{-1}_{s,k,\beta})^{1/n} \\
&\qquad - n + \left(1 - \frac{n\varepsilon}{n+1}\Lambda^{-1/(n+1)}\right) \operatorname{tr}_{\omega_t} \hat{\omega}_{t,u_\beta} \\
&\ge \frac{n^2\varepsilon}{n+1}(-\psi_{t,k} + u_\beta + \Lambda + 1)^{-1/(n+1)} (-\varphi_t + u_\beta - s)^{1/n} A^{-1/n}_{s,k,\beta} - n.
\end{aligned}$$

Therefore, at $x_0 \in \Omega_s^\circ$,

$$-(\varphi_t - u_\beta + s) \le \left(\frac{n+1}{n\varepsilon}\right)^n A_{s,k,\beta}(-\psi_{t,k} + u_\beta + \Lambda + 1)^{n/(n+1)} = \varepsilon(-\psi_{t,k} + u_\beta + \Lambda + 1)^{n/(n+1)},$$

i.e., $\Phi(x_0) \le 0$. Combining the two cases, we conclude that $\sup_X \Phi \le \epsilon_\beta \to 0$ as $\beta \to \infty$. It then follows that, on Ω_s,

$$(-\varphi_t + u_\beta - s)^{(n+1)/n} \le C_n A_{s,k,\beta}^{1/n}(-\psi_{t,k} + u_\beta + 1 + A_{s,k,\beta}) + \epsilon_\beta^{(n+1)/n}.$$

Letting $\beta \to \infty$, we have

$$(-\varphi_t + V_t - s)^{(n+1)/n} \le C_n A_{s,k}^{1/n}(-\psi_{t,k} + V_t + 1 + A_{s,k}),$$

where $A_{s,k} = \int_X \tau_k(-\varphi_t + V_t + s)e^F \omega^n$. Observe that $V_t \le 0$ by definition and, by the α-invariant estimate [Hörmander 1966; Tian 1987], there exists an $\alpha_0(n, \omega, \chi)$ such that

$$\int_{\Omega_s} \exp\left(\alpha_0 \frac{(-\varphi_t + V_t - s)^{(n+1)/n}}{A_{s,k}^{1/n}}\right)\omega^n \le \int_{\Omega_s} \exp(\alpha_0 C_n(-\psi_{t,k} + 1 + A_{s,k}))\omega^n \le Ce^{CA_{s,k}}. \tag{2-6}$$

Letting $k \to \infty$, we obtain

$$\int_{\Omega_s} \exp\left(\alpha_0 \frac{(-\varphi_t + V_t - s)^{(n+1)/n}}{A_s^{1/n}}\right)\omega^n \le Ce^{CA_s}.$$

Theorem 1(a) is proved by noting that $A_s \le E_t$ for any $s > 0$.

Once Theorem 1(a) has been proved, part (b) can be proved by following closely the arguments in [Guo et al. 2023a].

Fix $p > n$, and define $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ by $\eta(x) = (\log(1 + x))^p$. Note that η is a strictly increasing function with $\eta(0) = 0$, and let η^{-1} be its inverse function. Write

$$v := \frac{\alpha_0}{2}\left(\frac{-\varphi_t + V_t - s}{A_s^{1/(n+1)}}\right)^{(n+1)/n}. \tag{2-7}$$

Then by the generalized Young's inequality with respect to η, for any $z \in \Omega_s$,

$$\begin{aligned} v(z)^p e^{F(z)} &\le \int_0^{\exp(F(z))} \eta(x)\,dx + \int_0^{v(z)^p} \eta^{-1}(y)\,dy \le \exp(F(z))(1+|F(z)|)^p + \int_0^{v(z)^p} (e^{y^{1/p}} - 1)\,dy \\ &\le \exp(F(z))(1+|F(z)|)^p + p\int_0^{v(z)} e^y y^{p-1}\,dy \le \exp(F(z))(1+|F(z)|)^p + C(p)\exp(2v(z)). \end{aligned}$$

We integrate both sides in the inequality above over $z \in \Omega_s$ and get by Theorem 1(a) that

$$\int_{\Omega_s} v(z)^p e^{F(z)}\omega^n \le \int_{\Omega_s} e^F(1 + |F(z)|)^p \omega^n + \int_{\Omega_s} e^{2v(z)}\omega^n \le \|e^F\|_{L^1(\log L)^p} + C + Ce^{CE_t},$$

where the constant $C > 0$ depends only on n, ω_X and χ. In view of the definition of v, this implies

$$\int_{\Omega_s} (-\varphi_t + V_t - s)^{(n+1)p/n} e^{F(z)}\omega^n \le 2^p \alpha_0^{-p} A_s^{p/n}(\|e^F\|_{L^1(\log L)^p} + C + Ce^{CE_t}). \tag{2-8}$$

From the definition of A_s, it follows from Hölder's inequality that

$$A_s = \int_{\Omega_s} (-\varphi_t + V_t - s) e^F \omega^n \le \left(\int_{\Omega_s} (-\varphi_t + V_t - s)^{(n+1)p/n} e^F \omega^n \right)^{n/((n+1)p)} \cdot \left(\int_{\Omega_s} e^F \omega^n \right)^{1/q}$$
$$\le A_s^{1/(n+1)} (2^p \alpha_0^{-p} (\|e^F\|_{L^1(\log L)^p} + C + Ce^{CE_t}))^{n/((n+1)p)} \cdot \left(\int_{\Omega_s} e^F \omega^n \right)^{1/q},$$

where $q > 1$ satisfies $n/(p(n+1)) + 1/q = 1$, i.e., $q = p(n+1)/(p(n+1) - n)$. The inequality above yields

$$A_s \le (2^p \alpha_0^{-p} (\|e^F\|_{L^1(\log L)^p} + C + Ce^{CE_t}))^{1/p} \cdot \left(\int_{\Omega_s} e^F \omega^n \right)^{(1+n)/(qn)}. \tag{2-9}$$

Observe that the exponent of the integral on the right-hand of (2-9) satisfies

$$\frac{1+n}{qn} = \frac{pn + p - n}{pn} = 1 + \delta_0 > 1$$

for $\delta_0 := (p-n)/(pn) > 0$. For convenience of notation, set

$$B_0 := (2^p \alpha_0^{-p} (\|e^F\|_{L^1(\log L)^p} + C + Ce^{CE_t}))^{1/p}. \tag{2-10}$$

From (2-9) we then get

$$A_s \le B_0 \left(\int_{\Omega_s} e^F \omega^n \right)^{1+\delta_0}. \tag{2-11}$$

If we define $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ by $\phi(s) := \int_{\Omega_s} e^F \omega^n$, then (2-11) and the definition of A_s implies

$$r\phi(s+r) \le B_0 \phi(s)^{1+\delta_0} \quad \text{for all } r \in [0, 1] \text{ and } s \ge 0. \tag{2-12}$$

Since ϕ is clearly nonincreasing and continuous, a De Giorgi-type iteration argument shows that there is some S_∞ such that $\phi(s) = 0$ for any $s \ge S_\infty$. This finishes the proof of the L^∞ estimate of $\varphi_t - V_t$, combining with a bound on E_t by $\|e^F\|_{L^1(\log L)^1}$ which follows from Jensen's inequality; see Lemma 6 in [Guo et al. 2023a]. □

Finally, we note the recent advances in the theory of envelopes in [Guedj and Lu 2021; 2023], which can provide an approach to L^∞ estimates for Monge–Ampère equations on Hermitian manifolds.

3. Complex Hessian equations

We explain in this section how the proof of Theorem 1 can be modified to give a similar result for a degenerate family of complex Hessian equations. With the same notations as above, we consider the σ_k-equations

$$(\hat{\omega}_t + i\partial\bar{\partial}\varphi_t)^k \wedge \omega^{n-k} = c_t e^F \omega^n, \quad \sup_X \varphi_t = 0. \tag{3-1}$$

Define the envelope corresponding to the Γ_k-cone

$$\widetilde{V}_{t,k} = \sup\{v \mid v \in \mathrm{SH}_k(X, \omega, \hat{\omega}_t) \cap C^2,\ v \le 0\},$$

where $v \in \mathrm{SH}_k(X, \omega, \hat{\omega}_t) \cap C^2$ indicates that the vector of eigenvalues of the linear transformation $\omega^{-1} \cdot (\hat{\omega}_t + i\partial\bar{\partial}v)$ lies in the Γ_k-cone, which is the convex cone in $\mathbb{R}^n$ given by

$$\Gamma_k = \{\lambda \in \mathbb{R}^n \mid \sigma_1(\lambda) > 0, \dots, \sigma_k(\lambda) > 0\},$$

where $\sigma_j(\lambda)$ denotes the j-th elementary symmetric polynomial of $\lambda \in \mathbb{R}^n$.

Let

$$E_t(\varphi_t) = \int_X (-\varphi_t + \widetilde{V}_{t,k}) e^{nF/k} \omega^n$$

be the entropy associated to (3-1) as in [Guo et al. 2023a], and let $\bar{E}_t$ be an upper bound of $E_t(\varphi_t)$. Then the following L^∞ estimate holds for the solution φ_t to (3-1).

Theorem 3. *Let φ_t be the solution to (3-1). There exists a constant depending on*

$$\bar{E}_t, \quad \|e^{(n/k)F}\|_{L^1(\log L)^p}, \quad \frac{c_t}{[\hat{\omega}_t^k][\omega^{n-k}]} \quad \text{and} \quad p > n$$

such that

$$0 \le -\varphi_t + \widetilde{V}_{t,k} \le C.$$

This theorem can be derived using a similar argument as in Section 2 with suitable modifications for σ_k equations — see [Guo et al. 2023a] — so we omit the details. The only novel ingredient is the smooth approximation of $\widetilde{V}_{t,k}$ as in Lemma 2. One can adapt the method in [Berman 2019] to derive this required approximation. For the convenience of the reader, we present a sketch of the proof.

Lemma 4. *Fix $t \in (0, 1]$. There exists a sequence of smooth functions $u_\beta \in \mathrm{SH}_k(X, \omega, \hat{\omega}_t)$ converging uniformly to $\widetilde{V}_{t,k}$ as $\beta \to \infty$.*

Proof. Let $u_\beta \in \mathrm{SH}_k(X, \omega, \hat{\omega}_t)$ be the solution to the σ_k-equations

$$(\hat{\omega}_t + i\partial\bar{\partial}u_\beta)^k \wedge \omega^{n-k} = c_t e^{\beta u_\beta} \omega^n, \tag{3-2}$$

which admits a unique smooth solution by [Dinew and Kołodziej 2017]. We claim that there is a constant $C_t > 0$ such that

$$\sup_X |u_\beta - \widetilde{V}_{t,k}| \le \frac{C_t \log \beta}{\beta},$$

from which the lemma follows.

By the maximum principle, at a maximum point of u_β we have $i\partial\bar{\partial}u_\beta \le 0$, so

$$\beta u_\beta \le \log \frac{\hat{\omega}_t^k \wedge \omega^{n-k}}{c_t \omega^n} \le C_t,$$

that is, $u_\beta - C_t/\beta \le 0$. By the definition of $\widetilde{V}_{t,k}$, it follows that

$$u_\beta - \frac{C_t}{\beta} \le \widetilde{V}_{t,k}. \tag{3-3}$$

On the other hand, we fix a smooth $u \le 0$ such that $\hat{\omega}_t + i\partial\bar{\partial}u > 0$. Such a u exists because $[\hat{\omega}_t]$ is a Kähler class by assumption. For any $v \in \mathrm{SH}_k(X, \omega, \hat{\omega}_t) \cap C^2$ with $v \le 0$, we consider the barrier function

$$\tilde{u} = \frac{1}{\beta}u + \left(1 - \frac{1}{\beta}\right)v - \frac{C_t' \log \beta}{\beta},$$

where $C_t' > 0$ is a large constant to be determined. By direct calculation, we have

$$(\hat{\omega}_t + i\partial\bar{\partial}\tilde{u})^k \wedge \omega^{n-k} \ge \frac{1}{\beta^k}(\hat{\omega}_t + i\partial\bar{\partial}u)^k \wedge \omega^{n-k} \ge e^{\beta\tilde{u}}\omega^n,$$

where the last inequality holds if we choose C_t' large enough such that

$$e^{-C_t' \log \beta} \le \frac{1}{\beta^k} \min_X \frac{(\hat{\omega}_t + i\partial\bar{\partial}u)^k \wedge \omega^{n-k}}{\omega^n}.$$

Therefore, we get

$$(\hat{\omega}_t + i\partial\bar{\partial}\tilde{u})^k \wedge \omega^{n-k} \ge e^{\beta(\tilde{u} - u_\beta)}(\hat{\omega}_t + i\partial\bar{\partial}u_\beta)^k \wedge \omega^{n-k}.$$

At the maximum point of $\tilde{u} - u_\beta$, we have $(\hat{\omega}_t + i\partial\bar{\partial}\tilde{u})^k \wedge \omega^{n-k} \le (\hat{\omega}_t + i\partial\bar{\partial}u_\beta)^k \wedge \omega^{n-k}$. This shows that $\tilde{u} - u_\beta \le 0$ on X. Taking the supremum over all such v in $\tilde{u}$, it follows that

$$\left(1 - \frac{1}{\beta}\right)\widetilde{V}_{t,k} \le u_\beta + \frac{C_t \log \beta}{\beta}.$$

The lemma follows from this and (3-3). □

References

[Berman 2019] R. J. Berman, "From Monge–Ampère equations to envelopes and geodesic rays in the zero temperature limit", *Math. Z.* **291**:1-2 (2019), 365–394. MR Zbl

[Boucksom et al. 2010] S. Boucksom, P. Eyssidieux, V. Guedj, and A. Zeriahi, "Monge–Ampère equations in big cohomology classes", *Acta Math.* **205**:2 (2010), 199–262. MR Zbl

[Chen and Cheng 2021] X. Chen and J. Cheng, "On the constant scalar curvature Kähler metrics, I: A priori estimates", *J. Amer. Math. Soc.* **34**:4 (2021), 909–936. MR Zbl

[Chu et al. 2018] J. Chu, V. Tosatti, and B. Weinkove, "$C^{1,1}$ regularity for degenerate complex Monge–Ampère equations and geodesic rays", *Comm. Partial Differential Equations* **43**:2 (2018), 292–312. MR Zbl

[De Giorgi 1961] E. De Giorgi, *Frontiere orientate di misura minima*, Editrice Tecnico Sci., Pisa, 1961. MR Zbl

[Demailly et al. 2014] J.-P. Demailly, S. Dinew, V. Guedj, H. H. Pham, S. Kołodziej, and A. Zeriahi, "Hölder continuous solutions to Monge–Ampère equations", *J. Eur. Math. Soc.* **16**:4 (2014), 619–647. MR Zbl

[Dinew and Kołodziej 2014] S. Dinew and S. Kołodziej, "A priori estimates for complex Hessian equations", *Anal. PDE* **7**:1 (2014), 227–244. MR Zbl

[Dinew and Kołodziej 2017] S. Dinew and S. Kołodziej, "Liouville and Calabi–Yau type theorems for complex Hessian equations", *Amer. J. Math.* **139**:2 (2017), 403–415. MR Zbl

[Fu et al. 2020] X. Fu, B. Guo, and J. Song, "Geometric estimates for complex Monge–Ampère equations", *J. Reine Angew. Math.* **765** (2020), 69–99. MR Zbl

[Guedj and Lu 2021] V. Guedj and C. H. Lu, "Quasi-plurisubharmonic envelopes, I: Uniform estimates on Kähler manifolds", preprint, 2021. arXiv 2106.04273

[Guedj and Lu 2023] V. Guedj and C. H. Lu, "Quasi-plurisubharmonic envelopes, III: Solving Monge–Ampère equations on Hermitian manifolds", *J. Reine Angew. Math.* **800** (2023), 259–298. MR Zbl

[Guo et al. 2023a] B. Guo, D. H. Phong, and F. Tong, "On L^∞ estimates for complex Monge–Ampère equations", *Ann. of Math.* (2) **198**:1 (2023), 393–418. MR Zbl

[Guo et al. 2023b] B. Guo, D. H. Phong, and F. Tong, "Stability estimates for the complex Monge–Ampère and Hessian equations", *Calc. Var. Partial Differential Equations* **62**:1 (2023), art. id. 7. MR Zbl

[Hörmander 1966] L. Hörmander, *An introduction to complex analysis in several variables*, Van Nostrand, Princeton, NJ, 1966. MR Zbl

[Kołodziej 1998] S. Kołodziej, "The complex Monge–Ampère equation", *Acta Math.* **180**:1 (1998), 69–117. MR Zbl

[Tian 1987] G. Tian, "On Kähler–Einstein metrics on certain Kähler manifolds with $C_1(M) > 0$", *Invent. Math.* **89**:2 (1987), 225–246. MR Zbl

[Yau 1978] S. T. Yau, "On the Ricci curvature of a compact Kähler manifold and the complex Monge–Ampère equation, I", *Comm. Pure Appl. Math.* **31**:3 (1978), 339–411. MR Zbl

Received 3 Dec 2021. Revised 27 Jun 2022. Accepted 26 Jul 2022.

BIN GUO: bguo@rutgers.edu
Department of Mathematics & Computer Science, Rutgers University, Newark, NJ, United States

DUONG H. PHONG: phong@math.columbia.edu
Department of Mathematics, Columbia University, New York, NY, United States

FREID TONG: ftong@cmsa.fas.harvard.edu
Center for Mathematical Sciences and Applications, Harvard University, Cambridge, MA, United States

CHUWEN WANG: wang.chuwen@columbia.edu
Department of Mathematics, Columbia University, New York, NY, United States

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at msp.org/apde.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in APDE are usually in English, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with details about how your graphics were generated.

White space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

ANALYSIS & PDE

Volume 17 No. 2 2024