Analysis & PDE

Volume 11 No. 8 2018

msp

Analysis & PDE

msp.org/apde

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

The subscription price for 2018 is US $275/year for the electronic version, and $480/year (+$55, 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. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

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. 11, No. 8, 2018
dx.doi.org/10.2140/apde.2018.11.1841

INVARIANT MEASURE AND LONG TIME BEHAVIOR OF REGULAR SOLUTIONS OF THE BENJAMIN–ONO EQUATION

MOUHAMADOU SY

The Benjamin–Ono equation describes the propagation of internal waves in a stratified fluid. In the present work, we study large time dynamics of its regular solutions via some probabilistic point of view. We prove the existence of an invariant measure concentrated on $C^\infty(\mathbb{T})$ and establish some qualitative properties of this measure. We then deduce a recurrence property of regular solutions and other corollaries using ergodic theorems. The approach used in this paper applies to other equations with infinitely many conservation laws, such as the KdV and cubic Schrödinger equations in one dimension. It uses the fluctuation-dissipation-limit approach and relies on a *uniform* smoothing lemma for stationary solutions to the damped-driven Benjamin–Ono equation.

1. Introduction

The problem and statement of the main result. The Benjamin–Ono (BO) equation

$$\partial_t u + H\,\partial_x^2 u + u\,\partial_x u = 0 \tag{1-1}$$

describes the propagation of internal waves in a stratified fluid. The operator H in the equation is the Hilbert transform; it can be defined in the Fourier setting as the multiplier given by $-\mathrm{i}\,\mathrm{sgn}$ (see the Appendix). We assume that $u(t,x)$ is a real-valued function, $t \in \mathbb{R}_+$ and x belongs to the torus $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$. In this setting, existence and uniqueness of solution hold in any Sobolev space H^s for $s \geq 0$ (see, e.g., [Molinet 2008; Molinet and Pilod 2012] for its global well-posedness in $L^2(\mathbb{T})$). In the present paper, we use only the well-posedness of the problem in Sobolev spaces $H^s(\mathbb{T})$ with $s \geq 2$, so we refer the reader to [Abdelouhab et al. 1989].

In $L^2 := L^2(\mathbb{T})$, the well-posedness of (1-1) generates a topological dynamical system (DS) (L^2, ϕ_t), where ϕ_t is the flow of the equation. We are concerned with the description of the long time behavior of this dynamical system.

Given a Borel measure μ on L^2, we say that μ is invariant for (L^2, ϕ_t) if for any Borel set A of L^2 we have

$$\mu(\phi_t^{-1} A) = \mu(A) \quad \text{for all } t.$$

When such a measure exists, the triple (L^2, ϕ_t, μ) is called a measurable dynamical system (MDS). If in addition μ is finite, then we have very important information on the dynamics. Indeed the Poincaré recurrence theorem states that the dynamics is recurrent; that is, μ-almost every orbit returns in any

MSC2010: 35A09, 35B40, 35Q51, 35R60, 37K10.
Keywords: Benjamin–Ono equation, invariant measure, long time behavior, regular solutions, inviscid limit.

neighborhood of its origin in finite time. The well-known Von Neumann and Birkhoff ergodic theorems also apply to give more information on the long time behavior of the system. Our aim here is to construct such a measure, which will contribute to improving the understanding of the behavior of the solutions of (1-1).

Matsuno [1984] derived (at least formally) infinitely many conservation laws for the BO equation (1-1). They have the form

$$E_n(u) = \|u\|_n^2 + R_n(u), \quad n \in \tfrac{1}{2}\mathbb{N}, \tag{1-2}$$

where $\|\cdot\|_n$ stands for the homogeneous Sobolev norm of order n and R_n is a lower-order term.

In [Tzvetkov and Visciglia 2013; 2014; 2015; Deng 2015; Deng et al. 2015], the authors constructed a sequence of invariant Gaussian-type measures $\{\mu_n\}$ for (L^2, ϕ_t) satisfying

$$\mu_n \text{ is concentrated on } H^s(\mathbb{T}) \text{ for } s < n - \tfrac{1}{2}, \tag{$*$}$$

$$\mu_n(H^{n-\frac{1}{2}}(\mathbb{T})) = 0. \tag{$**$}$$

Formally, μ_n is defined as a renormalization of

$$d\mu_n(u) = e^{-E_n(u)}\, du = e^{-R_n(u)} e^{-\|u\|_n^2}\, du,$$

where $E_n(u)$ and $R_n(u)$ are the quantities given in (1-2). These authors constructed a Gaussian interpretation of the expression $e^{-\|u\|_n^2}\, du$ on the concerned spaces and proved that $e^{-R_n(u)}$ is an integrable density. In view of these results, there is an MDS for (1-1) in any Sobolev space and then its large time dynamics is described keeping in mind the theorems mentioned above. However, these results do not apply to infinitely smooth solutions; indeed by the property $(**)$ we have

$$\mu_n(C^\infty(\mathbb{T})) = 0 \quad \text{for all } n.$$

In the present work, we construct a measurable dynamical system for (1-1) on the space $C^\infty(\mathbb{T})$. Naturally, the Dirac measure at 0 is not the desired measure; although it is invariant under the flow of the BO equation, it gives only trivial information. More generally, to get substantial information on the system we have to also avoid singular measures. Another example of such a measure is the one concentrated on a stationary solution. Notice that measures μ_n discussed above verify the following "consistency" property: every set of full μ_n-measure is dense in $\dot{H}^{(n-\frac{1}{2})^-}$. Concerning the space C^∞, an obstruction to the construction of an invariant Gaussian-type measure is the nonexistence of a conservation law compatible with the regularity of that space. In particular, the approach used in the construction of the measures μ_n above does not seem to apply.

Another method allowing the construction of invariant measures (a priori not of Gaussian type) for PDEs was developed in [Kuksin 2004; Kuksin and Shirikyan 2004] in the context of Euler and Schrödinger equations, respectively. It is based on a fluctuation-dissipation (FD) argument and consists of adding to the equation appropriately normalized damping and stochastic terms, constructing an invariant measure for the resulting problem, and passing to the limit. But, a priori, the obstruction encountered in the Gaussian-type measure approach still remains in the FD approach because the underlying regularization is of Sobolev order and not C^∞. The idea in the present work is to exploit the regularization inherent in this approach with the use of an infinite subsequence of the Benjamin–Ono conservation laws to reach the C^∞-regularity.

In order to bring out a key preliminary result, we give the following stochastic set up: consider the diffusion problem (also called the stochastic Benjamin–Ono–Burgers (BOB) equation)

$$\partial_t u + H\,\partial_x^2 u + u\,\partial_x u = \alpha\,\partial_x^2 u + \sqrt{\alpha}\eta, \quad t > 0,\ x \in \mathbb{T}, \tag{1-3}$$

where η is a stochastic force and $\alpha \in (0, 1)$ is a viscosity parameter. In fact the problem (1-1) is the limit as $\alpha \to 0$ of (1-3). A probabilistic global well-posedness for (1-3) is proved in Section 3. Moreover, in Section 4, we establish the existence of stationary solutions[1] for this equation. We present now the following smoothing property for stationary solutions:

Lemma 1.1. *Suppose that the noise η is sufficiently regular in space. Let u_α be a stationary solution to (1-3) such that*

$$\mathbb{E}\|u_\alpha(t)\|^p < \infty \quad \textit{for all } p \geq 2. \tag{1-4}$$

Then

$$\mathbb{E}\|u_\alpha(t)\|_n^2 < \infty \quad \textit{for all } n \geq 1. \tag{1-5}$$

Moreover, if (1-4) holds uniformly in α then so does (1-5).

The proof of this lemma relies on a combination of deterministic and probabilistic estimations based on the conservation laws of (1-1).

We prove in Section 4 that any stationary solution to (1-3) satisfies (1-4) uniformly in α. Then, from (1-5) we conclude that stationary solutions to (1-3) are concentrated on C^∞. Passing to the limit as the viscosity goes to 0, we find the main result of this paper (Theorems 5.3, 6.1, 6.3 and 6.6):

Theorem 1.2. *There is a probability measure μ invariant under the flow of the BO equation (1-1) defined on $H^3(\mathbb{T})$ and such that*

$$\mu(C^\infty(\mathbb{T})) = 1.$$

Moreover, μ satisfies the following properties:

(1) *For any integer n, we have*

$$0 < \int_{H^3} \|u\|_n^2\,\mu(du) < \infty.$$

(2) *There are constants $\sigma, C > 0$ such that for any $R > 0$*

$$\mu(u \in H^3, \|u\| \geq R) \leq Ce^{-\sigma R^2}.$$

(3) *There is an infinite sequence of conservation laws of the form (1-2) whose laws under μ are absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$.*

(4) *The measure μ is of at least 2-dimensional nature in the sense that any compact set of Hausdorff dimension smaller than 2 has μ-measure 0.*

In fact, we expect infinite-dimensionality of the measure constructed here as in [Kuksin 2008; Kuksin and Shirikyan 2012] concerning the 2-dimensional Euler equations. To show this property in the context of the Benjamin–Ono equation, we have to prove some algebraic independence of the gradients of the

[1] Solutions to (1-3) whose laws are invariant along the time.

conservation laws. In the present work, we face a technical difficulty in establishing such an independence for an arbitrary number of conservation laws. We propose a proof inspired by [Kuksin 2008; Kuksin and Shirikyan 2012] which works for the (at least) 2-dimensionality. Then the infinite-dimensionality of μ remains an open question.

We deduce the following result by applying the Poincaré recurrence theorem.

Corollary 1.3. *For μ-almost all w in $C^\infty(\mathbb{T})$, there is a sequence $\{t_k\}$ increasing to infinity such that*

$$\lim_{k\to\infty} \|S_{t_k} w - w\|_n = 0 \quad \text{for any } n \geq 0.$$

Here S_t denotes the flow of the Benjamin–Ono equation (1-1) on $H^3(\mathbb{T})$.

In the construction of such a measure, we use the control of Sobolev norms provided by the infinite sequence of conservation laws. The KdV and cubic 1-dimensional NLS equations have infinitely many conservation laws whose structure is similar to (1-2) and our approach applies to these equations. Notice that an infinite sequence of invariant Gaussian-type measures of increasing regularity was constructed for KdV and cubic 1-dimensional NLS equations in [Zhidkov 2001a; 2001b]; we give then a kind of extension of this work to the $C^\infty(\mathbb{T})$ space. However, the Benjamin–Ono equation is more difficult than these equations because of the weakness of its dispersion compared to KdV and the presence of a derivative in its nonlinearity compared to NLS. Then, here, we confine ourselves to the study of the BO equation, which is less understood.

Let us briefly discuss an equation having infinitely many conservation laws but which is not admissible to the approach developed here. Consider the nonviscous Burgers' equation

$$\partial_t u + u\,\partial_x u = 0. \tag{1-6}$$

It is easy to check that an infinite sequence of conservation laws is given by the quantities

$$L_p(u) = \int u^p, \quad p \geq 1.$$

Our approach does not apply to (1-6). This is due to its lack of dispersion which breaks the control of Sobolev's norms.

Notation. • Let A and B be two positive quantities, we write

$$A \lesssim B$$

if there is a universal constant $\lambda \geq 0$ such that $A \leq \lambda B$.

• For a real number r, we denote by r^+ (resp. r^-) the quantity $r + \epsilon$ (resp. $r - \epsilon$), where ϵ is a positive number close enough to 0, while $r_+ := \max(r, 0)$.

• $\mathbb{Z}$ denotes the set of nonzero integers.

• $\dot{H}(\mathbb{T}) = \{u \in L^2(\mathbb{T}) \mid \int_{\mathbb{T}} u(x)\,dx = 0\}$.

• $\dot{H}^s(\mathbb{T}) = \{u \in \dot{H}(\mathbb{T}) \mid D^s u \in \dot{H}(\mathbb{T})\}$, and D^s is the s-th derivative of u, where $s \geq 0$.

• The $\dot{H}^s$-norm is denoted by $\|\cdot\|_s$ when $s > 0$ and the L^2-norm is denoted by $\|\cdot\|$.

• For a functional $A(u)$, we denote the first and second derivatives of A by $A'(u,v) := \partial_u A(u,v) = \partial A|_u(v)$ and $A''(u,v) := \partial_u^2 A(u,v) = \partial^2 A|_u(v,v)$.

• The sequence $\{e_n \mid n \in \mathbb{Z}\}$ is given by

$$e_n(x) = \begin{cases} \sin(nx)/\sqrt{\pi} & \text{for } n > 0, \\ \cos(nx)/\sqrt{\pi} & \text{for } n < 0 \end{cases}$$

and forms an orthonormal basis of $\dot{H}(\mathbb{T})$.

• $(\Omega, \mathscr{F}, \mathbb{P})$ is a complete probability space and $\mathscr{F}_t$ is a right-continuous filtration augmented with respect to $(\mathscr{F}, \mathbb{P})$. Given a sequence of real numbers $\{\lambda_n\}$ and a sequence of independent real standard Brownian motions $\{\beta_n(t)\}$ adapted to $\mathscr{F}_t$, we set

$$\zeta(t,x) = \sum_{n\in\mathbb{Z}} \lambda_n \beta_n(t) e_n(x), \tag{1-7}$$

$$\eta(t,x) = \frac{d}{dt}\zeta(t,x), \tag{1-8}$$

$$A_s = \sum_{n\in\mathbb{Z}} \lambda_n^2 n^{2s}. \tag{1-9}$$

Some stochastic results. The theorem and lemma below are useful ingredients in our work; we refer to [Karatzas and Shreve 1991] for their proofs.

Theorem 1.4 (Doob's optional theorem). *Let x_t be a continuous $\mathscr{F}_t$-martingale and $\tau \le \sigma$ be two $\mathscr{F}_t$-stopping times which are almost surely finite. Then*

$$\mathbb{E}x_\tau = \mathbb{E}x_\sigma = \mathbb{E}x_0. \tag{1-10}$$

Lemma 1.5. *Let x_t be a continuous random process which is adapted to $\mathscr{F}_t$. Then $x_t(\omega)$ is adapted to $\mathscr{F}_t$.*

Stochastic convolution. Let B be an operator on a separable Hilbert space H with which we endow a Hilbert basis $\{e_m\}_{m\in\mathbb{Z}}$. Suppose that $\{e_m\}$ are eigenvectors of B whose associated eigenvalues are $\{b_m\} \subset \mathbb{C}$, and moreover $|b_m| \to \infty$ as $m \to \infty$. Suppose that

$$V_t(B) := \sum_{m\in\mathbb{Z}} \lambda_m^2 \frac{|1 - e^{2tb_m}|}{2|b_m|} < \infty \quad \text{for all } t \ge 0; \tag{1-11}$$

then the quantity (which is called stochastic convolution)

$$\Theta_t(B) := \int_0^t e^{(t-s)B}\, d\zeta(s,x) := \sum_{m\in\mathbb{Z}} \lambda_m \left(\int_0^t e^{(t-s)b_m}\, d\beta_m(s) \right) e_m(x), \quad t \ge 0, \tag{1-12}$$

is well-defined in H. In fact $\Theta_t(B)$ is a continuous Gaussian process in H: for all $t \ge 0$, we have $\Theta_t(B) \sim \mathscr{N}_H(0, V_t(B))$.

Remark 1.6. If $\mathrm{Re}(b_m) < 0$, then the sequence $\{|1 - e^{2tb_m}|/|2b_m|\}$ is bounded (even uniformly in t); therefore $\Theta_t(B)$ is well-defined in H as soon as $\sum_m \lambda_m^2 < \infty$.

The concrete case that is studied in this paper is the Hilbert space $L^2(\mathbb{T})$ and an operator of type $-\partial_x^2$ (more exactly $B = -(H - \alpha)\,\partial_x^2$). With the use of the Itô isometry, we have

$$\mathbb{E}\|\Theta_t(B)\|_s^2 = \sum_{m\in\mathbb{Z}} m^{2s}\lambda_m^2 \frac{|1 - e^{2tb_m}|}{2|b_m|}, \quad b_m = -(\mathrm{i}\,\mathrm{sgn}(m) + \alpha)m^2. \tag{1-13}$$

Then $\mathrm{Re}(b_m) < 0$ for any $m \in \mathbb{Z}$. Therefore $\Theta_t(B) \in H^s$ almost surely as soon as

$$\sum_{m\in\mathbb{Z}} m^{2s}\lambda_m^2 < \infty. \tag{1-14}$$

In that case, as a Gaussian random variable in H^s, the stochastic convolution Θ_t verifies the Fernique theorem; that is, there is a constant c_s such that

$$\mathbb{E}e^{c_s\|\Theta_t\|_{H^s}^2} < \infty \quad \text{for all } t \geq 0. \tag{1-15}$$

Stochastic well-posedness and the Itô property relative to a Gelfand triple. Let us consider the following stochastic PDE:

$$du_t = (Lu + f(u))\,dt + d\zeta, \tag{1-16}$$

where L is a differential operator, f is a function possibly nonlinear in u and ζ is a Brownian motion defined as in (1-7).

Definition 1.7. Let $s \in \mathbb{R}$. Equation (1-16) is said to be stochastically (globally) well-posed in H^s if for all $T > 0$ the following properties hold:

(1) For any random variable u_0 in H^s which is independent of $\mathscr{F}_t$, we have, for almost all $\omega \in \Omega$,
   (a) (existence) There exists $u := u^\omega \in \Lambda_T(s) := C(0, T; H^s) \cap L^2(0, T; H^{s+1})$ satisfying the relation
   $$u(t) = u_0 + \int_0^t (Lu_s + f(u_s))\,ds + \zeta(t) \quad \text{for all } t \in [0, T] \tag{1-17}$$
   in H^{s-1}. We denote this solution by $u(t, u_0) := u^\omega(t, u_0)$.
   (b) (uniqueness) If $u_1, u_2 \in \Lambda_T(s)$ are two solutions in the sense of (1-17), then $u_1 \equiv u_2$ on $[0, T]$.

(2) (continuity with respect to initial data) For almost all ω, we have
$$\lim_{u_0 \to u_0'} u(\cdot, u_0) = u(\cdot, u_0') \quad \text{in } \Lambda_T(s), \tag{1-18}$$
where u_0 and u_0' are deterministic data in H^s.

(3) The process $(\omega, t) \mapsto u^\omega(t)$ is adapted to the filtration $\sigma(u_0, \mathscr{F}_t)$.

Remark 1.8. In what follows we call (H^{s-1}, H^s, H^{s+1}) a Gelfand triple. The process u_t described in Definition 1.7 satisfies the following properties:

- Considered as a process in H^s, it is progressively measurable with respect to $\sigma(u_0, \mathscr{F}_t)$; this follows from the continuity of u_t and Lemma 1.5.
- It satisfies the Feller property, being continuous in t and with respect to initial data.
- It is a Markov process: Set

$$P_t(w, \Gamma) := \mathbb{P}(u(t, w) \in \Gamma \mid u(0) = w);$$

then P_t satisfies the so-called Chapman–Kolmogorov relation. Let us write down the corresponding Markov semigroups:

$$\mathfrak{P}_t f(v) = \int_{H^s} f(w) P_t(v, dw), \quad C_b(H^s) \to C_b(H^s), \tag{1-19}$$

$$\mathfrak{P}_t^* \mu(\Gamma) = \int_{H^s} \mu(dw) P_t(w, \Gamma), \quad \mathfrak{p}(H^s) \to \mathfrak{p}(H^s). \tag{1-20}$$

Here, $C_b(H^s)$ is the space of bounded continuous functions on H^s, and $\mathfrak{p}(H^s)$ is the set of probability measures on H^s. These maps satisfy the duality relation

$$(\mathfrak{P}_t f, \mu) = (f, \mathfrak{P}_t^* \mu). \tag{1-21}$$

Now, let us introduce the following definition:

Definition 1.9. We say that (1-16) has the Itô property on the Gelfand triple (H^{s-1}, H^s, H^{s+1}) if

(1) it is stochastically well-posed on H^s;

(2) the process $h := Lu + f(u)$ is $\mathscr{F}_t$-adapted and

$$\mathbb{P}\left(\int_0^t (\|u(r)\|_{s+1}^2 + \|h(r)\|_{s-1}^2)\, dr < \infty \;\middle|\; \text{for all } t > 0 \right) = 1, \quad \sum_{m \in \mathbb{Z}} m^{2s} \lambda_m^2 < \infty. \tag{1-22}$$

Remark 1.10. Our definition of the Itô property is different from what we find in some literature. But the interest of our choice is that part (2) gathers "good" properties of a process allowing us to apply a version of the Itô formula proved in Section A.7 (Theorem A.7.5 and Corollary A.7.6) of [Kuksin and Shirikyan 2012]. Below, we present that formula.

Theorem 1.11 [Kuksin and Shirikyan 2012, Section A.7]. *Let $F \in C^2(H^s, \mathbb{R})$ be a functional which is uniformly continuous, together with its first two derivatives, on any ball of H^s. Suppose that F satisfies the following conditions*:

(1) *There is a function $K : \mathbb{R}_+ \to \mathbb{R}_+$ such that*

$$|\nabla_u F(u, v)| \le K(\|u\|_s) \|u\|_{s+1} \|v\|_{s-1}, \quad u \in H^{s+1},\ v \in H^{s-1}. \tag{1-23}$$

(2) *For any sequence $\{w_k\} \subset H^{s+1}$ converging toward $w \in H^{s+1}$ and any $v \in H^{s-1}$, we have*

$$\nabla_u F(w_k, v) \to \nabla_u F(w, v) \quad \text{as } k \to \infty. \tag{1-24}$$

(3)
$$\sum_{m\in\mathbb{Z}} a_m^2 \mathbb{E}\int_0^t |\nabla_u F(u,e_m)|^2\,ds<\infty \quad \text{for all } t>0. \tag{1-25}$$

Then we have

$$F(u(t)) = F(u(0)) + \int_0^t \left(\partial_u F(u(s),f(s)) + \frac{1}{2}\sum_{m\in\mathbb{Z}} \partial_u^2 F(u(s),g_m)\right) ds + \sum_{m\in\mathbb{Z}} \int_0^t \partial_u F(u(s),g_m)\,d\beta_m(s). \tag{1-26}$$

In particular,

$$\mathbb{E}F(u(t)) = \mathbb{E}F(u(0)) + \int_0^t \mathbb{E}\left(\partial_u F(u(s),f(s)) + \frac{1}{2}\sum_{m\in\mathbb{Z}} \partial_u^2 F(u(s),g_m)\right) ds. \tag{1-27}$$

If one omits (1-25), then we have the formula (1-26) in which t is replaced by the stopping time $t\wedge\tau_n$, where

$$\tau_n = \inf\{t\ge 0 \mid \|u(t)\|_s > n\}, \quad n\in\mathbb{N}, \tag{1-28}$$

with the convention $\inf\varnothing = +\infty$.

2. Deterministic estimates

Conservation laws. Following [Tzvetkov and Visciglia 2014], we define the following subsets of $C^\infty(\mathbb{T})$:

$$\mathscr{P}_1 = \{\partial_x^\alpha u \mid \partial_x^\alpha H u,\ \alpha\in\mathbb{N}\},$$
$$\mathscr{P}_2 = \left\{(\partial_x^{\alpha_1} Z_1 u)(\partial_x^{\alpha_2} Z_2 u) \mid \alpha_i\in\mathbb{N},\ Z_i\in\{\mathrm{Id}, H\}\right\}.$$

Let us define in a generic manner the sets $\mathscr{P}_n$, $n\ge 3$, containing the functions of the form

$$p_n(u) = \prod_{i=1}^k Z_i(p_{j_i}(u)), \quad \text{where} \quad Z_i\in\{\mathrm{Id}, H\}, \quad \sum_1^k j_i = n, \quad p_{j_i}\in\mathscr{P}_{j_i}, \quad 2\le k\le n, \quad j_i<n. \tag{2-1}$$

To a function $p_n(u)$ of the form (2-1) we associate the function

$$\tilde{p}_n(u) = \prod_{i=1}^k p_{j_i}(u), \tag{2-2}$$

and we set the quantities

$$S(p(u)) = \sum_{i=1}^n \alpha_i, \quad M(p(u)) = \max_{1\le i\le n} \alpha_i.$$

The following is a description given in [Tzvetkov and Visciglia 2014] for the integer-order remainder terms:

$$R_n(u) = \sum_{\substack{p(u)\in\mathscr{P}_3\\ \tilde{p}(u)=u\,\partial_x^{n-1}u\,\partial_x^n u}} c_n(p)\int p(u) + \sum_{\substack{p(u)\in\mathscr{P}_j, j=3,\dots,2n+2\\ S(p(u))=2n-j+2\\ M(p(u))\le n-1}} c_n(p)\int p(u), \tag{2-3}$$

where $c_n(p)$ are some constants. The first three integer-order conservation laws are

$$E_0(u) = \int u^2,$$

$$E_1(u) = \int (\partial_x u)^2 + \frac{3}{4}\int u^2 H\,\partial_x u + \frac{1}{8}\int u^4,$$

$$E_2(u) = \int (\partial_x^2 u)^2 - \frac{5}{4}\int \Big((\partial_x u)^2 H\,\partial_x u + 2\,\partial_x^2 u H\,\partial_x u\Big) + \frac{5}{16}\int \Big(5u^2(\partial_x u)^2 + u^2(H\,\partial_x u)^2 + 2uH(\partial_x u)H(u\,\partial_x u)\Big) + \int \Big(\frac{5}{32}u^4 H(\partial_x u) + \frac{5}{24}u^3 H(u\,\partial_x u)\Big) + \frac{1}{48}\int u^6.$$

Estimates. Let us give some properties for the integer-order conservation laws of the Benjamin–Ono equation.

Lemma 2.1. *For any integer $n \geq 1$, there are $c_n^-, c_n^+ > 0$ such that for all u in $H^n(\mathbb{T})$*

$$\tfrac{1}{2}\|u\|_n^2 - c_n^-\|u\|^{2n+2} \leq E_n(u) \leq 2\|u\|_n^2 + c_n^+\|u\|^{2n+2}. \tag{2-4}$$

Lemma 2.2. *For all $\epsilon > 0$, there is $C_\epsilon > 0$ such that for all u in $H^{n+1}(\mathbb{T})$*

$$E_n'(u, \partial_x^2 u) \leq (-2+\epsilon)\|u\|_{n+1}^2 + C_\epsilon\|u\|(1+\|u\|)^{b_n},$$

where b_n depends only on n.

Remark 2.3. Since the L^2-norm is preserved by (1-1) we can deduce from (2-4) and the arguments of the proof of Lemma 2.2, by adding appropriate polynomials of $\|u\|$, new conservation laws $E_n^*(u)$ and $\widetilde{E}_n(u)$ satisfying

$$0 \leq \|u\|_n^2 \leq E_n^*(u), \quad 0 \leq \|u\|_n^2 \leq \widetilde{E}_n'(u, u).$$

Inequalities (2-4) can be established using arguments similar to those of the proof of Lemma 2.2.

Proof of Lemma 2.2. Taking into account of the properties of the Hilbert transform such as continuity on H^s and L^p ($s \geq 0$, $p \in]1, \infty[$), we can neglect its effect for our purpose and just consider the functions

$$R_n^1(u) = \int u\,\partial_x^{n-1} u\,\partial_x^n u,$$

$$R_n^{2,j}(u) = \int \prod_{i=1}^{j} \partial_x^{\alpha_i} u, \quad j = 3, \ldots, 2n+2, \ \sum_{i=1}^{j} \alpha_i = 2n+2-j.$$

Here $R_n^1(u)$ corresponds to the first term of (2-3) and the second term of (2-3) can be estimated considering the quantities $R_n^{2,j}(u)$. Set

$$R_n^0 = \|u\|_n^2.$$

Estimates concerning R_n^0.

$$\partial_u R_n^0(u, \partial_x^2 u) = -2\|u\|_{n+1}^2. \tag{2-5}$$

Estimates concerning R_n^1.

$$\begin{aligned}\partial_u R_n^1(u,\partial_x^2 u) &= \int \partial_x^2 u\,\partial_x^{n-1}u\,\partial_x^n u + \int u\,\partial_x^{n+1}u\,\partial_x^n u + \int u\,\partial_x^{n-1}u\,\partial_x^{n+2}u\\ &= \int \partial_x^2 u\,\partial_x^{n-1}u\,\partial_x^n u - \int \partial_x u\,\partial_x^{n-1}u\,\partial_x^{n+1}u\\ &= 2\int \partial_x^2 u\,\partial_x^{n-1}u\,\partial_x^n u + \int \partial_x u(\partial_x^n u)^2 = I + II.\end{aligned}$$

Let $\gamma_i,\ i=1,2,3$, be three positive numbers satisfying $\sum_{i=1}^3 \frac{1}{\gamma_i}=1$. We apply the generalized Hölder formula with them to find

$$|I| \le \|\partial_x^2 u\|_{L^{\gamma_1}}\,\|\partial_x^{n-1}u\|_{L^{\gamma_2}}\,\|\partial_x^n u\|_{L^{\gamma_3}}.$$

By the embedding inequality $\|\cdot\|_{L^{\gamma_i}} \lesssim \|\cdot\|_{\frac12-\frac{1}{\gamma_i}}$, we get

$$|I| \lesssim \|u\|_{\frac52-\frac{1}{\gamma_i}}\,\|u\|_{-\frac12-\frac{1}{\gamma_i}+n}\,\|u\|_{\frac12-\frac{1}{\gamma_i}+n}.$$

Now interpolate between L^2 and H^{n+1} to find

$$|I| \le C_1\|u\|_{n+1}^{d_1}\,\|u\|^{3-d_1},$$

where

$$d_1 = \frac{2n+3}{2(n+1)} < 2.$$

One can establish the same control (with the same d_1) for $|II|$ by remarking that

$$|II| \lesssim \|u\|_1\,\|\partial_x^n u\|_{L^4}^2 \lesssim \|u\|_1\,\|u\|_{n+\frac14}^2 \lesssim \|u\|_{n+1}^{\frac{(n+1-1)+2(n+1-n-1/4)}{n+1}}\,\|u\|^c.$$

Then for suitable b_1

$$|\partial_u R_n^1(u,\partial_x^2 u)| \le \epsilon\|u\|_{n+1}^2 + C_\epsilon^1\|u\|^{b_1}. \tag{2-6}$$

Estimates concerning $R_n^{2,j}$.

$$\partial_u R_n^{2,j}(u,\partial_x^2 u) = \int \prod_{i=1}^{j}\partial_x^{\alpha_i}u, \quad j=3,\dots,2n+2,$$

where $\sum_{i=1}^j \alpha_i = 2n-j+4$ and $\max_{1\le i\le j}\alpha_i \le n+1$.

We follow two complementary cases:

Case 1: $\max_{1\le i\le j}\alpha_i \le n$. Let (γ_i) be j real numbers such that $\sum_{i=1}^j \frac{1}{\gamma_i}=1$. Then the generalized Hölder formula combined with usual interpolation inequalities shows

$$|\partial_u R_n^{2,j}(u,\partial_x^2 u)| \le C\prod_{i=1}^{j}\|u\|_{\kappa_i},$$

where $\kappa_i = \frac12 - \frac{1}{\gamma_i} + \alpha_i$. Then

$$|\partial_u R_n^{2,j}(u, \partial_x^2 u)| \le C \prod_{i=1}^{j} \|u\|^{\frac{n+1-\kappa_i}{n+1}} \|u\|_{n+1}^{\frac{\kappa_i}{n+1}}.$$

We remark now that

$$\sum_{i=1}^{j} \kappa_i = \sum_{1}^{j} \Big(\frac{1}{2} - \frac{1}{\gamma_i} + \alpha_i \Big) = 2n + 3 - \frac{j}{2}.$$

Then

$$\sum_{i=1}^{j} \frac{\kappa_i}{n+1} = \frac{2n + 3 - \frac{j}{2}}{n+1} < 2.$$

Thus for suitable b_2,

$$|\partial_u R_n^{2,j}(u, \partial_x^2 u)| \le \epsilon \|u\|_{n+1}^2 + C_\epsilon^2 \|u\|^{b_2}.$$

Case 2: $\alpha_1 = n+1$. Then $\sum_{i=2}^{j} \alpha_i = n - j + 3 \le n$, and we have

$$|\partial_u R_n^{2,j}(u, \partial_x^2 u)| \le \|u\|_{n+1} \Big(\int \prod_{i=2}^{j} |\partial_x^{\alpha_i} u|^2 \Big)^{\frac{1}{2}}.$$

Take again (γ_i) such that $\sum_{i=2}^{j} \frac{1}{\gamma_i} = 1$. Then

$$\begin{aligned} |\partial_u R_n^{2,j}(u, \partial_x^2 u)| &\le \|u\|_{n+1} \prod_{i=2}^{j} \|\partial_x^{\alpha_i} u\|_{L^{2\gamma_i}} \\ &\le \|u\|_{n+1} \prod_{i=2}^{j} \|u\|_{\kappa_i}, \quad \kappa_i = \frac{1}{2} - \frac{1}{2\gamma_i} + \alpha_i, \\ &\le \|u\|_{n+1} \prod_{i=2}^{j} \|u\|^{\frac{n+1-\kappa_i}{n+1}} \|u\|_{n+1}^{\frac{\kappa_i}{n+1}}. \end{aligned}$$

Since $\sum_{i=2}^{j} \kappa_i = n + 2 - \frac{j}{2} \le n + \frac{1}{2}$, we have $\frac{1}{n+1} \sum_{i=2}^{j} \kappa_i < 1$ and the existence of a suitable b_3 such that

$$|\partial_u R_n^{2,j}(u, \partial_x^2 u)| \le \epsilon \|u\|_{n+1}^2 + C_\epsilon^3 \|u\|^{b_3}. \tag{2-7}$$

Combining (2-5), (2-6) and (2-7) with a good choice of ϵ, we have the claim. □

3. IVP of the stochastic BOB equation

Consider the initial value problem concerning the stochastic BOB equation (1-3)

$$\begin{cases} \partial_t u + H\, \partial_x^2 u + u\, \partial_x u = \alpha\, \partial_x^2 u + \sqrt{\alpha}\eta, \quad t > 0, \\ u|_{t=0} = u_0. \end{cases} \tag{3-1}$$

Recall that, for $s \ge 0$,

$$A_s = \sum_{m \in \mathbb{Z}} m^{2s} \lambda_m^2.$$

These quantities measure the regularity in space of the noise. Namely,

$$A_s < +\infty \iff \eta(t,\cdot) \in \dot{H}^s.$$

Stochastic well-posedness, well-structuredness.

Proposition 3.1. *Let $s \geq 2$ be an integer. Suppose A_s is finite. Then the problem (3-1) is stochastically globally well-posed in $\dot{H}^s(\mathbb{T})$ in the sense of Definition 1.7.*

In order to prove the existence result in Proposition 3.1, we split the problem (3-1) as follows:

- A linear stochastic problem:

$$\begin{cases} \partial_t z_\alpha + H\,\partial_x^2 z_\alpha = \alpha\,\partial_x^2 z_\alpha + \sqrt{\alpha}\eta, \quad t > 0, \\ z_\alpha|_{t=0} = 0. \end{cases} \tag{3-2}$$

- A nonlinear deterministic problem:

$$\begin{cases} \partial_t v + H\,\partial_x^2 v + (v + z_\alpha)\,\partial_x(v + z_\alpha) = \alpha\,\partial_x^2 v, \quad t > 0, \\ v|_{t=0} = u_0. \end{cases} \tag{3-3}$$

  Here z_α is a realization of a solution of (3-2).

For z_α and v respective solutions of (3-2) and (3-3), it is easy to see that $u = v + z_\alpha$ is a solution of (3-1). The linear problem (3-2) is solved by the stochastic convolution (see the subsection on page 1845)

$$z_\alpha(t) = \sqrt{\alpha}\int_0^t e^{-(t-s)(H-\alpha)\,\partial_x^2}\,d\zeta(s) =: \sqrt{\alpha}z(t). \tag{3-4}$$

Remark that, as defined, the function z still depends on α. But all its Sobolev norms are uniformly controlled with respect to α; this justifies that abuse of notation.

If, for some $s \geq 0$, A_s is finite, then we have for all $T > 0$

$$z \in \Lambda_T(s) := C([0,T], \dot{H}^s(\mathbb{T})) \cap L^2([0,T], \dot{H}^{s+1}(\mathbb{T})) \quad \text{for } \mathbb{P}\text{-a.e. } \omega \in \Omega. \tag{3-5}$$

Uniqueness of solution for the problem (3-2) is obtained by standard arguments. Moreover, if we suppose A_n finite, we can apply the Itô formula to the $\dot{H}^n$-norms (which are preserved by the linear Benjamin–Ono equation) to find that

$$\mathbb{E}\|z_\alpha\|_n^2 + 2\alpha\int_0^t \mathbb{E}\|z_\alpha\|_{n+1}^2\,ds = \alpha A_n t. \tag{3-6}$$

Denoting by z^m the projection (z, e_m), we have that

$$z^m(t) = \lambda_m \int_0^t e^{m^2(t-s)(\mathrm{i}\,\mathrm{sgn}(m)-\alpha)}\,d\beta_m(s).$$

Since the function $s \to e^{m^2(t-s)(\mathrm{i}\,\mathrm{sgn}(m)-\alpha)}$ is C^1, we employ a usual (stochastic) integration by parts formula to obtain

$$z^m(t) = \lambda_m\beta_m(t) + m^2(\mathrm{i}\,\mathrm{sgn}(m) - \alpha)\lambda_m \int_0^t e^{m^2(t-s)(\mathrm{i}\,\mathrm{sgn}(m)-\alpha)}\beta(s)s\,ds.$$

Then we arrive at

$$\sup_{t\in[0,T]}|z^m(t)|^2 \le 2\lambda_m^2[1+(1-\alpha)^2m^4T^2]\sup_{t\in[0,T]}|\beta_m(t)|^2 \le 2\lambda_m^2[1+m^4T^2]\sup_{t\in[0,T]}|\beta_m(t)|^2.$$

After summing in m, we arrive at

$$\sup_{t\in[0,T]}\|z(t)\|^2 \lesssim_T \sup_{t\in[0,T]}\|\zeta(t)\|_2^2.$$

More generally, for any m such that A_{m+2} is finite, we have

$$\sup_{t\in[0,T]}\|z(t)\|_m^2 \lesssim_T \sup_{t\in[0,T]}\|\zeta(t)\|_{m+2}^2,$$

and finally

$$\sup_{t\in[0,T]}\|z_\alpha(t)\|_m^2 \lesssim_T \alpha \sup_{t\in[0,T]}\|\zeta(t)\|_{m+2}^2. \tag{3-7}$$

Proposition 3.2. *Let $s\ge 2$ be an integer, and suppose $A_s<\infty$. Let u_0 be a random variable in $\dot H^s(\mathbb{T})$ independent of $\mathscr{F}_t$. Then for any $T>0$, for a.e. ω, the nonlinear problem (3-3) associated to u_0 admits a solution in $\Lambda_T(s)$. Moreover the process solution is adapted to $\sigma(u_0,\mathscr{F}_t)$.*

Proposition 3.2 is proved combining the two paragraphs below:

A priori estimates. The following lemma is proved using the first three integer-order (modified) conservation laws $E_n^*(u)$ of the Remark 2.3, its proof is presented in the Appendix.

Lemma 3.3. *For any $T>0$, for almost any realization of z we have the following a priori estimates for the nonlinear problem (3-3):*

$$\sup_{t\in[0,T]}\|v(t)\|_i^2+\alpha\int_0^T\|v(t)\|_{i+1}^2\,dt \le C(T,\|u_0\|_i,\|z\|_{L^\infty(0,T;H^i)}),\quad i=0,1,2, \tag{3-8}$$

where C does not depend on $\alpha\in(0,1)$.

Since $H^2(\mathbb{T})$ is continuously embedded in $C^1(\mathbb{T})$, we infer:

Corollary 3.4. *For any $T>0$, for almost any realization of z, and for any initial datum $u_0\in H^2$, a solution v to (3-3) satisfies*

$$\sup_{t\in[0,T]}\|\partial_x v(t)\|_{L^\infty} \le C(T,\|u_0\|_2,\|z\|_{L^\infty(0,T;H^2)}), \tag{3-9}$$

where C does not depend on $\alpha\in(0,1)$.

Lemma 3.5. *For any $T>0$, for any integer $s>2$, and for almost any realization of z we have the higher-order a priori estimate for (3-3)*

$$\sup_{t\in[0,T]}\|v(t)\|_s^2+\alpha\int_0^T\|v(t)\|_{s+1}^2\,dt \le C(T,\|u_0\|_s,\|z\|_{L^\infty(0,T;H^s)}), \tag{3-10}$$

where C does not depend on $\alpha\in(0,1)$.

Before giving the proof of the estimate (3-10), let us prove the following commutator estimate:

Lemma 3.6. *Let $s \geq 3$ be an integer and v be in H^{s+1}. We have*

$$\|[\partial_x^s, v]\partial_x v\| \lesssim \|v\|_2 \, \|v\|_s, \tag{3-11}$$

where $[\partial_x^s, v]\partial_x v = \partial_x^s(v\,\partial_x v) - v\,\partial_x^s(\partial_x v)$.

Proof. By the Leibniz rule we have

$$[\partial_x^s, v]\partial_x v = \sum_{k=1}^{s} \binom{s}{k} \partial_x^k v\, \partial_x^{s+1-k} v.$$

We separate the above sum into three general terms:

(1) We have $k \in \{1, s\}$ if and only if the general term is $\partial_x v \partial_x^s v$. By using the embedding $H^1 \subset L^\infty$, we have the inequality

$$\|\partial_x v\, \partial_x^s v\| \leq \|v\|_2 \, \|v\|_s.$$

(2) We have $k \in \{2, s-1\}$ if and only if the general term is $\partial_x^2 v\, \partial_x^{s-1} v$. We have (always by $H^1 \subset L^\infty$)

$$\|\partial_x^2 v\, \partial_x^{s-1} v\| \leq \|v\|_2 \, \|v\|_s.$$

(3) When $s \geq 5$ we have the last situation, which is $3 \leq k \leq s-2$; we have then $3 \leq s+1-k \leq s-2$ as well. We estimate the corresponding general term as follows:

$$\|\partial_x^k v\, \partial_x^{s+1-k}\| \leq \|v\|_{k+1} \, \|v\|_{s+1-k} \lesssim \|v\|_2^{\frac{s-k-1}{s-2}} \, \|v\|_s^{\frac{k-1}{s-2}} \, \|v\|_2^{\frac{k-1}{s-2}} \, \|v\|_s^{\frac{s-k-1}{s-2}} = \|v\|_2 \, \|v\|_s.$$

We complete the proof after taking a weighted sum of these terms. □

Proof of the estimate (3-10). We recall the nonlinear equation satisfied by v

$$\partial_t v + H\,\partial_x^2 v - \alpha\,\partial_x^2 v = -v\,\partial_x v - \partial_x(v z_\alpha) - \tfrac{1}{2}\partial_x z_\alpha^2.$$

Then for an integer $s > 2$, we have

$$(\partial_x^s v, \partial_x^s \partial_t v) + \alpha(\partial_x^{s+1} v, \partial_x^{s+1} v) = -(\partial_x^s v, \partial_x^s(v\,\partial_x v)) \underbrace{-(\partial_x^s v, \partial_x^{s+1}(v z_\alpha))}_{=+(\partial_x^{s+1} v, \partial_x^s(v z_\alpha))} + \tfrac{1}{2}(\partial_x^{s+1} v, \partial_x^s z_\alpha^2).$$

Therefore

$$\tfrac{1}{2}\partial_t \|v\|_s^2 + \alpha \|v\|_{s+1}^2 = I + II + III.$$

Using the commutator estimate (3-11) and the algebra structure of $H^s(\mathbb{T})$, we have

$$\begin{aligned} |I| &= |(\partial_x^s v, \partial_x^s(v\,\partial_x v) - v\partial_x^s \partial_x v) + (\partial_x^s v, v\,\partial_x^s \partial_x v)| \\ &= |(\partial_x^s v, [\partial_x^s, v]\partial_x v) - \tfrac{1}{2}(\partial_x v, |\partial_x^s v|^2)| \\ &\lesssim \|v\|_s^2 \, \|v\|_2. \end{aligned}$$

By Cauchy–Schwarz and the algebra structure of H^s, we have

$$|II| + |III| \leq \tfrac{1}{2}\alpha \|v\|_{s+1}^2 + C_1 \|v\|_s^2 \, \|z\|_s^2 + C_2 \alpha \|z\|_s^4,$$

where C_1 and C_2 depend only on s. It remains to combine the Gronwall lemma with (3-8) to get the claim. □

Local and global existence for the nonlinear problem (3-3). Let $s \geq 2$. For a positive T the space $\Lambda_T(s)$ is endowed with the norm defined by

$$\|u\|_{\Lambda_T(s)} = \sup_{t\in[0,T]} \left(e^{-\frac{t}{T}} \left\{ \|u(t)\|_s^2 + \alpha \int_0^t \|u(r)\|_{s+1}^2 \, dr \right\} \right)^{\frac{1}{2}}. \tag{3-12}$$

Let $R > 0$; denote by B_R the ball in H^s of center 0 and radius R.

Remark 3.7. The factor $e^{-\frac{t}{T}}$ in (3-12) is introduced just for convenience in the computations. The norm defined in (3-12) is actually equivalent to the one without that factor.

Proposition 3.8. *Let $s \geq 2$ and $\alpha \in (0,1)$. For all $R > 0$, there is $T_R > 0$ such that for any u_0 in $B_{\frac{R}{2}}$, the nonlinear problem (3-3) has a unique solution in $\Lambda_{T_R}(s)$.*

Remark 3.9. We combine the local existence of Proposition 3.8, Lemma 3.3, and estimate (3-10) to get the global existence for (3-3).

Proof of Proposition 3.8. Let us look for a fixed point of the map

$$\mathfrak{F}v = e^{-t(H-\alpha)\partial_x{}^2} u_0 - \int_0^t e^{-(t-s)(H-\alpha)\partial_x{}^2} (z_\alpha + v)\, \partial_x (z_\alpha + v)\, ds.$$

We proceed as follows:

Step 1: We prove that for any $R > 0$, there is $T > 0$ such that the ball $B_{T,s}$ of $\Lambda_T(s)$ centered at 0 and of radius R satisfies $\mathfrak{F}(B_{T,s}) \subset B_{T,s}$ if $\|u_0\|_s \leq \frac{1}{2}R$:

$$\begin{aligned}
-\frac{1}{2}\frac{d}{dt}\|\mathfrak{F}v\|_s^2 &= -(\partial_t D^s \mathfrak{F}v, D^s \mathfrak{F}(v)) \\
&= -((H-\alpha)D^{s+1}\mathfrak{F}(v), D^{s+1}\mathfrak{F}(v)) + \frac{1}{2}(D^s(z_\alpha+v)^2, D^{s+1}\mathfrak{F}(v)) \\
&\geq \alpha\|\mathfrak{F}(v)\|_{s+1}^2 - \frac{1}{2}\|z_\alpha + v\|_s^2 \, \|\mathfrak{F}(v)\|_{s+1} \\
&\geq \alpha\|\mathfrak{F}(v)\|_{s+1}^2 - \frac{\alpha}{2}\|\mathfrak{F}(v)\|_{s+1}^2 - \frac{C}{\alpha}(\|z_\alpha\|_s^4 + \|v\|_s^4).
\end{aligned}$$

Then there is an universal constant $c > 0$ such that

$$\frac{d}{dt}\|\mathfrak{F}(v)\|_s^2 + \alpha\|\mathfrak{F}(v)\|_{s+1}^2 \leq \frac{c}{\alpha} e^{\frac{2t}{T}} (R^4 + \|z_\alpha\|_{\Lambda_T(s)}^4).$$

Thus, after integration with respect to t, we find

$$\|\mathfrak{F}(v)\|_s^2 + \alpha\int_0^t \|\mathfrak{F}(v)\|_{s+1}^2 \, ds \leq \|u_0\|_s^2 + \frac{\tilde{c}T}{\alpha} e^{\frac{t}{T}} (R^4 + \|z_\alpha\|_{\Lambda_T(s)}^4).$$

Multiplying the last relation by $e^{-\frac{t}{T}}$, it remains to choose T small enough so that we obtain the claimed result.

Step 2: We now prove that $\mathfrak{F}$ is a contraction on the ball constructed above. We have

$$\partial_t \mathfrak{F}v = -\{(v+z_\alpha)\,\partial_x(v+z_\alpha) + (H-\alpha)\,\partial_x{}^2\mathfrak{F}v\}.$$

Then for v_1 and v_2 in $\Lambda_T(s)$, we have

$$\begin{aligned}-\frac{1}{2}\frac{d}{dt}\|\mathfrak{F}v_1-\mathfrak{F}v_2\|_s^2 &= -(\partial_t D^s(\mathfrak{F}v_1-\mathfrak{F}v_2), D^s(\mathfrak{F}v_1-\mathfrak{F}v_2))\\ &= (D^s(F_z(v_1)-F_z(v_2)), D^{s+1}(\mathfrak{F}v_1-\mathfrak{F}v_2)) + \alpha\|\mathfrak{F}v_1-\mathfrak{F}v_2\|_{s+1}^2,\end{aligned}$$

where

$$F_z(v) = \tfrac{1}{2}(z_\alpha+v)^2.$$

We show easily that

$$\|D^s(F_z(v_1)-F_z(v_2))\|^2 \le C(s)\|v_1-v_2\|_s^2(\|v_1+v_2\|_s^2+\|z_\alpha\|_s^2).$$

This allows us to get that

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\|\mathfrak{F}v_1-\mathfrak{F}v_2\|_s^2 + \frac{\alpha}{2}\|\mathfrak{F}v_1-\mathfrak{F}v_2\|_{s+1}^2 &\le \frac{C(s)}{\alpha}\|v_1-v_2\|_s^2(\|v_1+v_2\|_s^2+\|z_\alpha\|_s^2)\\ &\le e^{\frac{t}{T}}\frac{C(s)(4R^2+\|z\|_{\Lambda_T(s)}^2)}{\alpha}\|v_1-v_2\|_{\Lambda_T(s)}^2.\end{aligned}$$

After integration in t, we find

$$\|\mathfrak{F}v_1-\mathfrak{F}v_2\|_s^2 + \alpha\int_0^t \|\mathfrak{F}v_1-\mathfrak{F}v_2\|_{s+1}^2\,ds \le Te^{\frac{t}{T}}\frac{C(s)(4R^2+\|z_\alpha\|_{\Lambda_T(s)}^2)}{\alpha}\|v_1-v_2\|_{\Lambda_T(s)}^2.$$

We multiply this inequality by $e^{-\frac{t}{T}}$; the T found in the first step can be decreased if necessary to give a contraction.

We conclude by using the fixed point theorem. □

Remark 3.10. By definition, v is $\sigma(u_0,\mathscr{F}_t)$-adapted. Then the process $u=v+z_\alpha$ is continuous and $\sigma(u_0,\mathscr{F}_t)$-adapted. Thanks to Lemma 1.5, the process u is progressively measurable with respect to that filtration.

End of the proof of Proposition 3.1, the well-posedness of (1-3). Let u_1 and u_2 be two solutions of (1-3) starting respectively at $u_{1,0}$ and $u_{2,0}$, and set $w=u_1-u_2$; then the problem solved by w is

$$\begin{cases}\partial_t w + (H-\alpha)\,\partial_x^2 w + w\,\partial_x w + \partial_x(wu_2) = 0,\\ w|_{t=0} = u_{1,0}-u_{2,0} =: w_0.\end{cases}$$

Using the arguments of the proof of (3-10), we show

$$\sup_{t\in[0,T]}\|w(t)\|_s^2 + \alpha\int_0^T \|w(r)\|_{s+1}^2\,dr \le C\big(\alpha,T,\|\partial_x w\|_{L^\infty(0,T;L^\infty)},\|u_2\|_{L^\infty(0,T;H^s)}\big)\|w_0\|_s^2.$$

Hence follow the uniqueness and the continuity with respect to initial data. □

The stochastic well-posedness that we just established combined with the estimates (3-8) and the equation (3-10) implies the following:

Proposition 3.11. *Let $j \geq 2$. Suppose A_j finite. Then (1-3) is well-structured on the Gelfand triple (H^{j-1}, H^j, H^{j+1}) in the sense of Definition 1.9.*

Probabilistic estimates and proof of Lemma 1.1.

Exponential control of the L^2-norm.

Proposition 3.12. *Let $p \geq 1$. Then the functional $E_0^p(u) = \|u\|^{2p}$ satisfies the conditions of Theorem 1.11 on the Gelfand triple (H^{-1}, L^2, H^1).*

Proof. Thanks to the polynomial nature of $E_0^p(u)$ on L^2, the uniform continuity on bounded sets and the conditions (1-23) and (1-24) follow easily. We confine ourself to the proof of (1-25). The argument we use, to this end, is the following: As we have already shown, the solution of (1-3) can be represented as the sum of a linear part and a nonlinear part. Now we will show that the nonlinear part can be controlled by the initial datum and an "exponential of the averaged linear part". On the other hand, we show that the linear part is exponentially controlled; then we get the needed control on the initial solution u.

Control of the nonlinear part v. In this part we prove that for all $r, \epsilon > 0$ and $p \geq 1$

$$\|v(r)\|^{2p} \leq e^{f(r,\epsilon,p)} e^{\frac{2\epsilon p}{r}\int_0^r \|\partial_x z_\alpha\|_{L^\infty}^2 \, ds} \left(\|u_0\|^2 + \int_0^r \|z_\alpha\|_1^4 \, ds \right)^p, \tag{3-13}$$

where $f(r, \epsilon, p) = \frac{p}{4}\left(2r + \frac{r^2}{\epsilon}\right)$. Indeed, multiplying (3-3) by v and integrating in x, one obtains

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt}\|v\|^2 + \alpha\|v\|_1^2 &= -(v, \partial_x(vz_\alpha)) - (v, z_\alpha \, \partial_x z_\alpha) \\
&= \frac{1}{2}[(v, v\partial_x z_\alpha) + (v, \partial_x z_\alpha^2)] \\
&\leq \frac{1}{2}[\|v\| \, \|vz_\alpha\| + \|v\| \, \|z_\alpha\|_1^2] \\
&\leq \frac{r}{8\epsilon}\|v\|^2 + \frac{\epsilon}{r}\|v\|^2 \, \|\partial_x z_\alpha\|_{L^\infty}^2 + \frac{1}{4}\|v\|^2 + \frac{1}{4}\|z_\alpha\|_1^4.
\end{aligned}$$

Then we use the Gronwall lemma, choose $t = r$ and take the resulting inequality to the power p to arrive at the claim.

Exponential control of the linear part. Now, the linear part of the solution satisfies the estimate

$$\mathbb{E} e^{\frac{\epsilon}{t}\int_0^t \|z_\alpha\|_2^2 \, ds} \leq 3, \tag{3-14}$$

where $\epsilon > 0$ is small enough. Indeed, by applying the Itô formula to $\|z\|_2^{2p}$ for $p \geq 1$ we have

$$\mathbb{E}\|z_\alpha\|_2^{2p} \leq \frac{A_1^p p^p}{\kappa^p}. \tag{3-15}$$

Integrating in t, we find

$$\mathbb{E}\left(\frac{1}{t}\int_0^t \|z_\alpha\|_2^{2p} \, ds \right) \leq \frac{A_1^p p^p}{\kappa^p}.$$

Thanks to Jensen's inequality, we infer

$$\mathbb{E}\left(\frac{1}{t}\int_0^t \|z_\alpha\|_2^2\,ds\right)^p \le \frac{A_1^p p^p}{\kappa^p}.$$

Now, let $0<\epsilon\le \kappa/(2A_1e)$; then we have

$$\mathbb{E}\frac{\left(\frac{\epsilon}{t}\int_0^t \|z_\alpha\|_2^2\,ds\right)^p}{p!} \le \frac{p^p}{2^p e^p\, p!}.$$

We recall that for any integer $p>0$, we have that $p!\ge \left(\frac{p}{e}\right)^p$; then we arrive at the claimed result.

Control of the quadratic variation of $E_0^p(u)$. We have that

$$\begin{aligned}\sum_{m\ge 0} a_m^2\mathbb{E}\int_0^t |\partial_u(E_0^p)(u,e_m)|^2\,ds &\lesssim_p \sum_{m\in\mathbb{Z}} a_m^2\mathbb{E}\int_0^t \|u\|^{4(p-1)}|(u,e_m)|^2\,ds\\ &\lesssim_p \mathbb{E}\int_0^t \|u\|^{4p-2}\,ds \lesssim_p \mathbb{E}\int_0^t (\|v\|^{4p-2}+\|z_\alpha\|^{4p-2})\,ds.\end{aligned}$$

Set $q=4p-2$; one sees, with the use of the estimate (3-15) (or just by invoking the Fernique theorem), that

$$\mathbb{E}\int_0^t \|z\|_2^q\,ds<\infty \quad \text{for any } t\ge 0.$$

Now we use the estimate (3-13); then, for any $\epsilon>0$,

$$\mathbb{E}\int_0^t \|v_s\|^q\,ds \le \int_0^t e^{f(s,\epsilon,q)}\mathbb{E}\left[e^{\frac{\epsilon q}{s}\int_0^s \|\partial_x z_\alpha\|_{L^\infty}^2 dr}\left(\|u_0\|^2+\int_0^s \|z_\alpha\|_{L1}^4\,dr\right)^q\right]ds.$$

Then for any $\delta>0$, we use the Young inequality to find

$$\mathbb{E}\int_0^t \|u_s\|^q\,ds \lesssim \int_0^t e^{f(s,\epsilon,q)}\mathbb{E}\left[e^{\frac{q(1+\delta)\epsilon}{\delta s}\int_0^s \|\partial_x z_\alpha\|_{L^\infty}^2 dr}+\underbrace{\left(\|u_0\|^2+\int_0^s \|z_\alpha\|_1^4\,dr\right)^{q(1+\delta)}}_{R_{q,\delta(s)}}\right]ds.$$

One uses the estimate (3-15) to bound $\mathbb{E}R_{q,\delta}(s)$ by $C_{q,\delta}(1+s^{q(1+\delta)})$. On the other hand, for any $\delta>0$ we choose $\epsilon>0$ small enough so that one can use the estimate (3-14) and the embedding $H^2\subset L^\infty$ to get the bound

$$\mathbb{E}e^{\frac{2p(1+\delta)\epsilon}{\delta s}\int_0^s \|\partial_x z_\alpha\|_{L^\infty}^2 dr}\le 3.$$

Then we get

$$\mathbb{E}\int_0^t \|v_s\|^{2p}\,ds \lesssim \int_0^t e^{f(s,\epsilon,p)}(1+s^{q(1+\delta)})\,ds<\infty \quad \text{for all } t\ge 0. \qquad \square$$

Proposition 3.13. *Let u be the solution of* (3-1):

(1) *Suppose that $\mathbb{E}E_0(u_0)<\infty$; then*

$$\mathbb{E}E_0(u)+2\alpha\int_0^t \mathbb{E}\|u(s)\|_1^2\,ds=\mathbb{E}E_0(u_0)+\alpha A_0 t. \tag{3-16}$$

(2) *Let $p > 1$. Suppose that $\mathbb{E}E_0^p(u_0) < \infty$; then*

$$\mathbb{E}E_0^p(u) \le e^{-p\alpha t}\mathbb{E}E_0^p(u_0) + p^p A_0^p. \tag{3-17}$$

Proof. The identity (3-16) is easily proven by applying the Itô formula to the conservation law $E_0(u)$. Let us prove (3-17):

For $p > 1$, we apply the Itô formula to $E_0^p(u)$ to find

$$dE_0^p(u) = pE_0^{p-1}(u)dE_0(u) + \frac{\alpha p(p-1)}{2}E_0^{p-2}(u)\sum_{m\in\mathbb{Z}}\lambda_m^2|E_0'(u,e_m)|^2dt.$$

Taking the expectation, we get

$$\mathbb{E}E_0^p(u) + \mathbb{E}\int_0^t f_\alpha(u(s))\,ds = \mathbb{E}E_0^p(u_0),$$

where

$$f_\alpha(u) = 2p\alpha E_0^{p-1}(u)\|u\|_1^2 - \alpha pE_0^{p-1}(u)A_0 - \frac{\alpha p(p-1)}{2}E_0^{p-2}(u)\sum_{m\in\mathbb{Z}}\lambda_m^2|E_0'(u,e_m)|^2.$$

Let us set

$$Q = pE_0^{p-1}(u)A_0 + \frac{p(p-1)}{2}E_0^{p-2}(u)\sum_{m\in\mathbb{Z}}\lambda_m^2|E_0'(u,e_m)|^2.$$

Remarking that

$$\sum_{m\in\mathbb{Z}}\lambda_m^2|E_0'(u,e_m)|^2 \le 2A_0E_0(u),$$

we get, with the use of the Young inequality, the estimate

$$Q \le \epsilon E_0^p(u) + \frac{p^{2p}}{\epsilon^{p-1}}A_0^p.$$

On the other hand

$$p\alpha E_0^{p-1}(u)\|u\|_1^2 \ge p\alpha E_0^p(u).$$

Choosing $\epsilon = p$, we see that

$$\mathbb{E}f_\alpha(u) \ge p\alpha\mathbb{E}E_0^p(u) - p^{p+1}A_0^p\alpha.$$

Then

$$\mathbb{E}E_0^p(u) + p\alpha\int_0^t \mathbb{E}E_0^p(u(s))\,ds \le \mathbb{E}E_0^p(u_0) + p^{p+1}A_0^p\alpha t.$$

Gronwall's lemma gives the claimed result. □

Control of higher-order Sobolev norms. The polynomial nature of the Benjamin–Ono conservation laws E_j allows to establish the following result:

Proposition 3.14. *Let $j \ge 1$, then the functional E_j satisfies the conditions (1-23) and (1-24) of Theorem 1.11 on the triple (H^{j-1}, H^j, H^{j+1}).*

In view of this result the "stopping time" version of the Itô formula (1-26) applies to the functionals E_j.

Theorem 3.15. *Let $j \geq 1$ be an integer. Suppose A_j is finite. There are $\theta_j > 0$, $\gamma_j > 0$ such that for any solution u of (3-1) in H^2 issued from $u_0 \in H^2$ which satisfies $\mathbb{E}E_j(u_0) < \infty$, we have*

$$\mathbb{E}E_j(u) + \alpha \int_0^t \mathbb{E}\|u\|_{j+1}^2 \, ds \leq \mathbb{E}E_j(u_0) + \alpha A_j \left(t + c_j \int_0^t \mathbb{E}\|u\|_j^2 \, ds + \gamma_j \int_0^t \mathbb{E}\|u\|(1+\|u\|)^{\theta_j} \, ds \right), \quad (3\text{-}18)$$

where c_j depends only on j.

Proof. The fact that $E_j(u)$ is preserved by the BO equation translates into

$$\partial_u E_j(u, -H\,\partial_x^2 u - u\,\partial_x u) = 0.$$

Setting the Markov time $\tau_n = \inf\{t \geq 0, \|u(t)\|_j > n\}$ and applying the Itô formula (1-26), we get

$$E_j(u(t\wedge\tau_n)) = E_j(u_0) + \alpha \int_0^{t\wedge\tau_n} \left(\partial_u E_j(u, \partial_x^2 u) + \frac{1}{2} \sum_{m\in\mathbb{Z}} \lambda_m^2 \, \partial_u^2 E_j(u, e_m) \right) ds + \sum_{m\in\mathbb{Z}} \lambda_m \int_0^{t\wedge\tau_n} \partial_u E_j(u, e_m) \, d\beta_m(s).$$

Then by the Doob optional stopping theorem, Theorem 1.4, we have

$$\mathbb{E}E_j(u(t\wedge\tau_n)) = \mathbb{E}E_j(u_0) + \alpha\mathbb{E} \int_0^{t\wedge\tau_n} \left(\partial_u E_j(u, \partial_x^2 u) + \frac{1}{2} \sum_{m\in\mathbb{Z}} \lambda_m^2 \, \partial_u^2 E_j(u, e_m) \right) ds.$$

Using the monotone convergence theorem, we arrive at

$$\mathbb{E}E_j(u(t)) = \mathbb{E}E_j(u_0) + \alpha\mathbb{E} \int_0^{t} \left(\partial_u E_j(u, \partial_x^2 u) + \frac{1}{2} \sum_{m\in\mathbb{Z}} \lambda_m^2 \, \partial_u^2 E_j(u, e_m) \right) ds.$$

By Lemma 2.2, we have

$$\partial_u E_j(u, \partial_x^2 u) \leq -\|u\|_{j+1}^2 + P_j(\|u\|), \quad (3\text{-}19)$$

where P_j is the polynomial of Lemma 2.2. Following the arguments of the proof of Lemma 2.2, we establish that

$$|\partial_u^2 E_j(u, e_m)| \leq c_j m^{2j} (\|u\|_j^2 + Q_j(\|u\|)), \quad (3\text{-}20)$$

where $Q_j(r) = q_j r(1+r)^{k_j}$, q_j and k_j depend only on j. Then take the expectation and combine (3-19) with (3-20) to get the claim. □

Now we are able to give the proof of Lemma 1.1.

Proof of Lemma 1.1. Let u be a stationary solution to (1-3) which satisfies the integrability assumption (1-4), and suppose that A_j is finite for any j. Recall the estimate

$$\mathbb{E}E_j(u) \leq \mathbb{E}\|u\|_j^2 + c_n^+ \mathbb{E}\|u\|^{2j+2}. \quad (3\text{-}21)$$

Then using the integrability assumption (1-4), we see that $\mathbb{E}E_j(u)$ is finite as soon as $\mathbb{E}\|u\|_j^2 < \infty$.

Note that, by the stationarity of u, the estimates (3-18) become (under the assumption that $\mathbb{E}E_j(u)$ is finite)

$$\mathbb{E}\|u\|_{j+1}^2 \le A_j[1 + c_j\mathbb{E}\|u\|_j^2 + \gamma_j\mathbb{E}\|u\|(1+\|u\|)^{\theta_j}] \tag{3-22}$$

since the distribution do not depend on t. We are going to argue by induction. Note that the needed induction property is given by the combination of (3-22) and (3-21) because they give at the same time the finiteness of $\mathbb{E}E_j(u)$ and the control of $\mathbb{E}\|u\|_{j+1}^2$ as soon as $\mathbb{E}\|u\|_j^2$ is finite. Moreover if (1-4) holds uniformly in α then so does $\mathbb{E}\|u\|_{j+1}^2$ once the control on $\mathbb{E}\|u\|_j^2$ is uniform in α. It remains to prove the initial step, namely $\mathbb{E}\|u\|_1^2$ is finite and does not depend on α. But using again the integrability assumption at the order $p = 2$, the stationarity of u combined with the estimate (3-16) gives

$$\mathbb{E}\|u\|_1^2 = \frac{A_0}{2}. \qquad \square$$

4. Stationary measures for the viscous problem

Consider the stochastic BOB problem (1-3) posed on $\dot{H}^2(\mathbb{T})$. By the estimates (3-16), (3-17) and Theorem 3.15, we have

$$\mathbb{E}E_0(u) + 2\alpha\int_0^t \mathbb{E}\|u\|_1^2\,ds = \mathbb{E}E_0(u_0) + \alpha A_0 t,$$

$$\mathbb{E}E_0^p(u) \le e^{-p\alpha t}\mathbb{E}E_0^p(u_0) + C_p A_0^p,$$

$$\mathbb{E}E_1(u) + \alpha\int_0^t \mathbb{E}\|u\|_2^2\,ds \le \mathbb{E}E_1(u_0) + \alpha\left(A_1 t + c_1\int_0^t \mathbb{E}\|u\|_1^2\,ds + \int_0^t \mathbb{E}W_1(\|u\|)\,ds\right),$$

$$\mathbb{E}E_2(u) + \alpha\int_0^t \mathbb{E}\|u\|_3^2\,ds \le \mathbb{E}E_2(u_0) + \alpha\left(A_2 t + c_2\int_0^t \mathbb{E}\|u\|_2^2\,ds + \int_0^t \mathbb{E}W_2(\|u\|)\,ds\right),$$

where W_1 and W_2 are the polynomials resulting from the estimate (3-18); their expectation is controlled using the second estimate. Now suppose $u_0 = 0$ almost surely; then by an induction argument, we get

$$\mathbb{E}E_2(u) + \alpha\int_0^t \mathbb{E}\|u\|_3^2\,ds \le \alpha C t,$$

where C is universal. Now in view of Remark 2.3, we can suppose $E_n(u) \ge 0$ (indeed, adding $c\|u\|^6$ to $E_2(u)$ we find a similar estimate). Then

$$\frac{1}{t}\int_0^t \mathbb{E}\|u\|_3^2\,ds \le C, \tag{4-1}$$

where C is, in particular, independent of t. Denote by $\lambda_\alpha(t)$ the law of the solution $u(t)$ to (1-3) starting at 0, and consider the time average

$$\bar{\lambda}_\alpha(t) = \frac{1}{t}\int_0^t \lambda_\alpha(s)\,ds.$$

Using the estimate (4-1), we show

$$\int_{H^2} \|u\|_3^2 \, \bar{\lambda}_\alpha(t)(du) \le C. \tag{4-2}$$

Then by the Chebyshev inequality we have

$$\bar{\lambda}_\alpha(t)(\{\|u\|_3 > R\}) \le \frac{C}{R^2} \quad \text{for any } R > 0.$$

Thus the compactness of the embedding $H^3(\mathbb{T}) \subset H^2(\mathbb{T})$ combined with the Prokhorov theorem implies that the family $\{\lambda_\alpha(t) \mid t > 0\}$ is compact with respect to the weak topology of H^2. Then for any α we denote by μ_α an accumulation point at infinity of the above family. The classical Bogoliubov–Krylov argument implies that μ_α is a stationary measure for (1-3). Passing to the limit $t \to \infty$ in (4-2) (using an approximation argument), we see that $\mu_\alpha(H^3) = 1$ for any α. We summarize these results in the following statement:

Proposition 4.1. *For any $\alpha \in (0,1)$, the stochastic BOB equation (1-3) posed in $H^2(\mathbb{T})$ has a stationary measure μ_α concentrated on $H^3(\mathbb{T})$.*

Theorem 4.2. *Let $\alpha \in (0,1)$. Suppose that A_n is finite for any n. Then any stationary measure μ_α of the problem (1-3) posed in $\dot{H}^2(\mathbb{T})$ satisfies*

$$\int_{H^2(\mathbb{T})} \|u\|_1^2 \, \mu_\alpha(du) = \frac{A_0}{2}, \tag{4-3}$$

$$\int_{H^2(\mathbb{T})} \|u\|^{2p} \, \mu_\alpha(du) \le p^p A_0^p \quad \textit{for any } 1 \le p < \infty, \tag{4-4}$$

$$\int_{H^2(\mathbb{T})} \|u\|_n^2 \, \mu_\alpha(du) \le D_n \quad \textit{for any } n \ge 2, \tag{4-5}$$

where, for any n, D_n does not depend on (t, α).

Proof. It suffices to prove (4-4) since then the estimate (4-5) follows from Lemma 1.1. We combine (3-16) and the stationarity of u to get (4-3). Let us prove (4-4).

To this end, let $R > 0$. Consider a C^∞-function χ_R satisfying

$$\chi_R(u) = \begin{cases} 1 & \text{if } \|u\|_2 \le R, \\ 0 & \text{if } \|u\|_2 > R + 1. \end{cases}$$

Let $p \ge 1$; we have

$$\int_{H^2} E_0^p(u) \, \chi_R(u) \, \mu_\alpha(du) = \int_{H^2} \mathbb{E}\{E_0^p(u(t,v)) \, \chi_R(u(t,v))\} \mu_\alpha(dv), \tag{4-6}$$

where $u(\cdot, v)$ is the solution of (1-3) starting at v. We pass to the limit $t \to \infty$ in the right-hand side of (4-6), and using (3-17) (u is in the ball of size R) and the stationarity of μ_α, we find

$$\int_{H^2} E_0^p(u) \, \chi_R(u) \, \mu_\alpha(du) \le p^p A_0^p.$$

Now Fatou's lemma allows to conclude. □

Corollary 4.3. *Let $\alpha \in (0,1)$. Suppose $A_n < \infty$ for any n. Then any stationary measure μ_α for the stochastic BOB problem (1-3) posed in $\dot{H}^2(\mathbb{T})$ is concentrated on $C^\infty(\mathbb{T})$.*

Proof. Let $n > 2$. Combining the estimate (4-5) and the Chebyshev inequality we find

$$\mu_\alpha(\{u \in H^2 \mid \|u\|_n \geq R\}) \leq \frac{D_n}{R^2}.$$

Setting $B_n(0,R)$ to be the ball in H^n of center 0 and radius R, we have

$$\int_{H^2} \mathbb{1}_{B_n(0,R)}(u)\, \mu_\alpha(du) = \mu_\alpha(B_n(0,R)) \geq 1 - \frac{D_n}{R^2}.$$

Passing to the limit on R (with the use of the Lebesgue convergence theorem), we get

$$\mu_\alpha(H^n(\mathbb{T})) = 1.$$

Thus

$$1 = \mu_\alpha\left(\bigcap_{n>2} H^n(\mathbb{T})\right) = \mu_\alpha(C^\infty(\mathbb{T})). \qquad \square$$

5. Invariant measure for the BO equation

In this section, $S_t : H^3(\mathbb{T}) \to H^3(\mathbb{T})$, $t \geq 0$, denotes the flow of the Benjamin–Ono equation (1-1). The map $S_{t,\alpha} : H^3 \to H^3$ denotes the one of the stochastic Benjamin–Ono–Burgers equation (1-3). We denote by ϕ_t, ϕ_t^*, $\phi_{t,\alpha}$, $\phi_{t,\alpha}^*$ the associated Markov semigroups, respectively. We suppose in what follows that $A_n < \infty$ for any $n > 0$.

Some convergence results of the stochastic BOB equation to the BO equation.

Lemma 5.1. *For any $T > 0$. For any $w \in H^3(\mathbb{T})$, we have, $\mathbb{P}$-almost surely,*

$$\sup_{t\in[0,T]} \|S_{t,\alpha} w - S_t w\|_2 \to 0 \quad \text{as } \alpha \to 0.$$

Proof. We write

$$\|S_{t,\alpha} w - S_t w\|_2 = \|v + z_\alpha - S_t w\|_2 \leq \|v - S_t w\|_2 + \|z_\alpha\|_2,$$

where

$$z_\alpha(t) = \sqrt{\alpha} \int_0^t e^{-(t-s)(H-\alpha)\,\partial_x^2}\, d\zeta(s) = \sqrt{\alpha} z(t)$$

and v is the solution of

$$\partial_t v + H\,\partial_x^2 v + (v + z_\alpha)\,\partial_x(v + z_\alpha) = \alpha\,\partial_x^2 v, \tag{5-1}$$

$$v_{t=0} = w. \tag{5-2}$$

Thanks to the estimate (3-7), we have that $\sup_{t\in[0,T]} \|z_\alpha\|_2 = \sqrt{\alpha} \sup_{t\in[0,T]} \|z\|_2$, where the quantity $\sup_{t\in[0,T]} \|z\|_2$ does not depend on α. Setting $h = v - S_t w$, we have

$$\sup_{t\in[0,T]} \|S_{t,\alpha} w - S_t w\|_2 \leq \sup_{t\in[0,T]} \|h\|_2 + \sqrt{\alpha} \sup_{t\in[0,T]} \|z\|_2.$$

We claim that $\sup_{t\in[0,T]}\|h\|_2 = O(\sqrt{\alpha})$. Indeed using the estimate (3-10) and the H^3-conservation law, we show that

$$\|h\|_2^3 \le c\|h\|\,\|h\|_3^2 \le C(T, \|w\|_{L^\infty(0,T;H^3)}, \|z\|_{H^3})\|h\|.$$

Taking the difference between (5-1) and the BO equation (1-1), we see that h satisfies

$$\partial_t h + H\,\partial_x^2 h + h\,\partial_x h = -\partial_x(hS_t w) - \partial_x(vz_\alpha) - z_\alpha\,\partial_x z_\alpha.$$

We multiply the above equation by h and we integrate on $\mathbb{T}$ to get

$$\partial_t\|h\|^2 = \tfrac{1}{2}(h^2, \partial_x S_t w) - (h, \partial_x(vz_\alpha)) - \tfrac{1}{2}(h, \partial_x z_\alpha^2).$$

By the Cauchy–Schwarz inequality and the algebra structure of H^1 we find

$$\begin{aligned}\partial_t\|h\|^2 &\le \tfrac{1}{2}\|h\|^2\,\|\partial_x S_t w\|_{L^\infty} + \tfrac{1}{2}\|h\|^2 + C\,\|v\|_1^2\,\|z_\alpha\|_1^2 + \tfrac{1}{4}\|h\|^2 + \tfrac{1}{4}\|z_\alpha\|_1^4\\ &\le \tfrac{1}{2}\|h\|^2\big(\|\partial_x S_t w\|_{L^\infty} + \tfrac{3}{2}\big) + C\alpha \sup_{t\in[0,T]}\|v\|_1^2 \sup_{t\in[0,T]}\|z\|_1^2 + \tfrac{1}{4}\alpha^2 \sup_{t\in[0,T]}\|z\|_1^4.\end{aligned}$$

Using the H^2-conservation law, we control $\|S_t w\|_{L^\infty(0,T;H^{3/2+})}$ (which does not depend on α) and $\|v\|_{L^\infty(0,T;H^1)}$ (see the estimate (3-8)). It remains to apply the Gronwall lemma to get the claim. □

Lemma 5.2. *For all $T, R, r > 0$, we have*

$$\sup_{w\in B(0,R)}\ \sup_{t\in[0,T]}\ \mathbb{E}\big[\|S_{t,\alpha}w - S_t w\|_2 \mathbb{1}_{\{\|z\|_{L^\infty(0,T;H^2)}\le r\}}\big] = O_{R,r,T}(\sqrt{\alpha}).$$

Here $B(0,R)$ is the ball in $H^3(\mathbb{T})$ of center 0 and radius R.

Proof.

$$\begin{aligned}\mathbb{E}\big[\|S_{t,\alpha}w - S_t w\|_2 \mathbb{1}_{\{\|z\|_{L^\infty(0,T;H^2)}\le r\}}\big] &= \int_\Omega \|S_{t,\alpha}w - S_t w\|_2 \mathbb{1}_{\{\|z\|_{L^\infty(0,T;H^2)}\le r\}}(\omega)\,d\mathbb{P}(\omega)\\ &\le \int_\Omega [\|h\|_2 + r\sqrt{\alpha}]\mathbb{1}_{\{\|z\|_{L^\infty(0,T;H^2)}\le r\}}(\omega)\,d\mathbb{P}(\omega),\end{aligned}$$

where $h = v - S_t w$ as before. The arguments of the proof of Lemma 5.1 allow to see that $\sup_{t\in[0,T]}\|h\|_2 \le C_{R,r,T}\sqrt{\alpha}$. This gives the claimed result. □

An accumulation point for the viscous stationary measures. In what follows we denote by $M(H^3)$ the space of probability measures on H^3.

Theorem 5.3. *For any sequence $(\alpha_k)_{k\in\mathbb{N}} \subset (0,1)$ converging to 0 as $k\to\infty$, there is a subsequence $\alpha_{r(k)}$ and $\mu \in M(H^3)$ such that*

- $\lim_{k\to\infty}\mu_{\alpha_{r(k)}} = \mu$ *in the weak topology of H^3,*
- *μ is invariant under the flow of the Benjamin–Ono equation in $H^3(\mathbb{T})$,*
- *μ is concentrated on $C^\infty(\mathbb{T})$,*

- *μ satisfies*

$$\int_{H^3(\mathbb{T})} \|u\|_1^2 \, \mu(du) = \frac{A_0}{2}, \tag{5-3}$$

$$\int_{H^3(\mathbb{T})} \|u\|^{2p} \, \mu(du) \le p^p A_0^p \quad \textit{for any } 1 \le p < \infty, \tag{5-4}$$

$$\int_{H^3(\mathbb{T})} \|u\|_n^2 \, \mu(du) < \infty \quad \textit{for } n \ge 2. \tag{5-5}$$

Proof. The proof consists in the following four steps:

(1) Existence of an accumulation point μ. The estimate (4-5) with $n = 4$ implies the tightness of the sequence of measures (μ_α) in $H^3(\mathbb{T})$ and, by the Prokhorov theorem, the existence of the claimed accumulation point μ on $H^3(\mathbb{T})$.

(2) Invariance of μ under the Benjamin–Ono flow. Denote by $(\mu_{\alpha_k})_{k\in\mathbb{N}}$ a subsequence of (μ_α) converging to μ (with $\lim_{k\to\infty} \alpha_k = 0$); to simplify the notations we write μ_k instead. The corresponding flow and Markov semigroup will be denoted $S_{t,k}$ and $\phi_{t,k}$.

The following diagram represents the idea of proof of the invariance of μ:

$$\begin{array}{ccc} \phi_{t,k}^* \mu_k & \overset{(I)}{=\!=} & \mu_k \\ {\scriptstyle (III)}\downarrow & & \downarrow{\scriptstyle (II)} \\ \phi_t^* \mu & \overset{(IV)}{=\!=} & \mu \end{array}$$

The equality (I) is the invariance of μ_k by $\phi_{t,k}$, and (II) is proved above. Then (IV) is proved once (III) is checked.

Let f be a real bounded Lipschitz function on $H^2(\mathbb{T})$. Without loss of generality assume that f is bounded by 1. Then

$$\begin{aligned} (\phi_{k,t}^* \mu_k, f) - (\phi_t^* \mu, f) &= (\mu_k, \phi_{t,k} f) - (\mu, \phi_t f) \\ &= (\mu_k, \phi_{t,k} f - \phi_t f) - (\mu - \mu_k, \phi_t f) \\ &= A - B. \end{aligned}$$

The term B converges to 0 as $k \to \infty$ by the weak convergence of (μ_k) to μ. And for any $R > 0$

$$\begin{aligned} |A| &\le \int_{H^3} \mathbb{E}|f(S_{t,k}w) - f(S_t w)| \, \mu_k(dw) \\ &= \int_{B(0,R)} \mathbb{E}|f(S_{t,k}w) - f(S_t w)| \, \mu_k(dw) + \int_{H^3 \setminus B(0,R)} \mathbb{E}|f(S_{t,k}w) - f(S_t w)| \, \mu_k(dw) \\ &= A_1 + A_2. \end{aligned}$$

Recalling that f is bounded by 1, we get by the Chebyshev inequality

$$A_2 \le 2\mu_k(H^3 \setminus B(0, R)) \le \frac{C}{R^2}, \tag{5-6}$$

where C is finite and does not depend on k (estimate (4-5)). Denote by $L_t^\infty H_x^2$ the space $L^\infty(0,T;H^2)$. Let $r>0$. We have

$$A_1=\int_{B(0,R)}\mathbb{E}\big[|f(S_{t,k}w)-f(S_tw)|\mathbb{1}_{\{\|z\|_{L_t^\infty H_x^2}\le r\}}\big]\,\mu_k(dw)$$
$$+\int_{B(0,R)}\mathbb{E}\big[|f(S_{t,k}w)-f(S_tw)|\mathbb{1}_{\{\|z\|_{L_t^\infty H_x^2}> r\}}\big]\,\mu_k(dw)=A_{1,1}+A_{1,2}.$$

As before, since f is bounded by 1, we use (3-6) and Chebyshev's inequality to get

$$A_{1,2}\le\frac{C_T}{r^2}.$$

On the other hand, since f is Lipschitz on H^2, we have

$$A_{1,1}\le C_f\int_{B(0,R)}\mathbb{E}\big[\|S_{t,k}w-S_tw\|_2\mathbb{1}_{\{\|z\|_{L_t^\infty H_x^2}\le r\}}\big]\,\mu_k(dw)$$
$$\le C_f\sup_{w\in B(0,R)}\mathbb{E}\big[\|S_{t,k}w-S_tw\|_2\mathbb{1}_{\{\|z\|_{L_t^\infty H_x^2}\le r\}}\big],$$

where C_f is the Lipschitz constant of f.

According to Lemma 5.2, we find

$$A_{1,1}\le C_{f,R,r,T}\sqrt{\alpha_k}.$$

Finally, we arrive at

$$|A|\le C_{f,R,r,T}\sqrt{\alpha_k}+\mathrm{Const}(T)\left(\frac{1}{r^2}+\frac{1}{R^2}\right),$$

where Const does not depend on k. We get the desired result after passing to the limits in the order

$$k\to\infty,\quad R,r\to\infty.$$

(3) The estimates for the measure μ. Denoting by χ_R a bump function on the ball $B(0,R)$ of $H^3(\mathbb{T})$, by (4-3) we have

$$\int_{H^3}\chi_R(v)\|v\|_1^2\,\mu_k(dv)\le\frac{A_0}{2}.$$

Passing to the limit $k\to\infty$ we find

$$\int_{H^3}\chi_R(v)\|v\|_1^2\,\mu(dv)\le\frac{A_0}{2}.$$

Then Fatou's lemma gives

$$\mathbb{E}\|u\|_1^2=\int_{H^3}\|v\|_1^2\,\mu(dv)\le\frac{A_0}{2}.\tag{5-7}$$

We proceed similarly to show (5-4) and (5-5).

Now we write

$$\frac{A_0}{2}=\int_{B(0,R)}\|v\|_1^2\,\mu_k(dv)+\int_{H^3\setminus B(0,R)}\|v\|_1^2\,\mu_k(dv).$$

We use the Cauchy–Schwarz and Chebyshev inequalities to show that

$$\int_{H^3\setminus B(0,R)} \|u\|_1^2\,\mu_k(du) = \int_{H^3} \|u\|_1^2 \mathbb{1}_{\|u\|_3>R}(u)\,\mu_k(du)$$
$$\leq \left(\int_{H^3} \|u\|_1^4\,\mu_k(du)\right)^{\frac{1}{2}} (\mu_k(\|u\|_3 > R))^{\frac{1}{2}} \leq \frac{\sqrt{\mathbb{E}[\|u\|_3^2]\mathbb{E}[\|u\|_1^4]}}{R}.$$

We can control $\mathbb{E}[\|u\|_1^4]$ and $\mathbb{E}[\|u\|_3^2]$ uniformly in k combining interpolation inequalities and the estimates (5-4) and (5-5). Then there is a constant $C > 0$ independent of k such that

$$\frac{A_0}{2} - \frac{C}{R} \leq \int_{H^3} \chi_R(v)\|v\|_1^2\,\mu_k(dv).$$

We find (5-3) after passing to the limits in the order

$$k \to \infty, \quad R \to \infty,$$

and combining this with (5-7).

(4) The measure μ is concentrated on $C^\infty(\mathbb{T})$. This immediately follows from the estimates (5-5) with use of the arguments of the proof of Corollary 4.3. □

6. Qualitative properties of the measure

Absolute continuity of some observables with respect to the Lebesgue measure. The following result is inspired by [Shirikyan 2011; Kuksin and Shirikyan 2012], where the local time concept is used to deduce nondegeneracy properties of measures constructed for the nonlinear Schrödinger and Euler equations.

Theorem 6.1. *Suppose that $\lambda_m \neq 0$ for all m. Then for any integer $n \geq 1$, there are constants b_n and c_n such that the distribution of the observable $\widetilde{E}_n(u) := E_n(u) + c_n\|u\|^2(1+\|u\|^2)^{b_n}$ under μ has a density with respect to the Lebesgue measure on $\mathbb{R}$.*

For the proof of Theorem 6.2 below, we refer the reader to [Shirikyan 2011] and the proof of Theorem 5.2.12 of [Kuksin and Shirikyan 2012], where the authors prove similar results in the case of the nonlinear Schrödinger and Euler equations respectively.

Theorem 6.2. *The measure μ constructed in Theorem 5.3 satisfies the following nondegeneracy properties*:

(1) *Let $\lambda_m \neq 0$ for at least two indices. Then μ has no atom at 0 and*

$$\mu(\{u \in C^\infty \mid \|u\| \leq \delta\}) \leq C\sqrt{A_0}\gamma^{-1}\delta \quad \text{for all } \delta > 0, \tag{6-1}$$

where $\gamma = \inf\{A_0 - \lambda_m^2 \mid m \in \mathbb{Z}\}$ and C is a universal constant.

(2) *Let $\lambda_m \neq 0$ for all indices. Then there is an increasing continuous function $h(r)$ vanishing at $r = 0$ such that*

$$\mu(\{u \in C^\infty(\mathbb{T}) \mid \|u\| \in \Gamma\}) \leq h(\ell(\Gamma)) \tag{6-2}$$

for any Borel set $\Gamma \subset \mathbb{R}$, where ℓ stands for the Lebesgue measure on $\mathbb{R}$.

Proof of Theorem 6.1. We prove the claim for the stationary measures in the case $\alpha > 0$, with uniform bounds in α. Then we can pass to the limit $\alpha \to 0$ to obtain the desired result using the Portmanteau theorem. First we apply the Itô formula to $\widetilde{E}_n(u)$:

$$\widetilde{E}_n(u(t)) = \widetilde{E}_n(u(0)) + \alpha \int_0^t A(s)\, ds + \sqrt{\alpha} \sum_{m\in\mathbb{Z}} \lambda_m \int_0^t \widetilde{E}_n'(u, e_m)\, d\beta_m(s),$$

where

$$A(s) = \partial_u \widetilde{E}_n(u, \partial_x^2 u) + \frac{1}{2} \sum_{m\in\mathbb{Z}} \lambda_m^2\, \partial_u^2 \widetilde{E}_n(u, e_m).$$

Denote by $\Lambda_t(a, \omega)$ its local time which reads (see the identity (A.45) of [Kuksin and Shirikyan 2012])

$$\begin{aligned}\Lambda_t(a,\omega) = (\widetilde{E}_n(u(t)) - a)_+ - (\widetilde{E}_n(u(0)) - a)_+ - \alpha \int_0^t A(s) \mathbb{1}_{(a,+\infty)}(\widetilde{E}_n(u))\, ds \\ - \sqrt{\alpha} \sum_{m\in\mathbb{Z}} \lambda_m \int_0^t \mathbb{1}_{(a,+\infty)}(\widetilde{E}_n(u)) \widetilde{E}_n'(u, e_m)\, d\beta_m(s).\end{aligned}$$

Using the stationarity of u, we infer that

$$\mathbb{E}\Lambda_t(a) = -\alpha t \mathbb{E}[A(0) \mathbb{1}_{(a,+\infty)}(\widetilde{E}_n(u))]. \tag{6-3}$$

Now using the (local time) identity (A.44) of [Kuksin and Shirikyan 2012] with the function $\mathbb{1}_\Gamma$, we get

$$2\int_\Gamma \Lambda_t(a)\, da = \alpha \sum_{m\in\mathbb{Z}} \lambda_m^2 \int_0^t \mathbb{1}_\Gamma(\widetilde{E}_n(u)) \widetilde{E}_n'(u, e_m)^2\, ds.$$

The stationarity of u gives again

$$2\int_\Gamma \mathbb{E}\Lambda_t(a)\, da = \alpha t \sum_{m\in\mathbb{Z}} \lambda_m^2 \mathbb{E}[\mathbb{1}_\Gamma(\widetilde{E}_n(u)) \widetilde{E}_n'(u, e_m)^2]. \tag{6-4}$$

Comparing (6-3) and (6-4), we find

$$\sum_{m\in\mathbb{Z}} \lambda_m^2 \mathbb{E}[\mathbb{1}_\Gamma(\widetilde{E}_n(u)) \widetilde{E}_n'(u, e_m)^2] \le 2\lambda(\Gamma)\mathbb{E}|A(0)| \le C\ell(\Gamma). \tag{6-5}$$

Recall now the form of $\widetilde{E}_n(u)$:

$$\widetilde{E}_n(u) = \|u\|_n^2 + R_n(u) + P_n(\|u\|^2),$$

where

$$P_n(r) = c_n r(1 + r)^{b_n}.$$

Then

$$\widetilde{E}_n'(u, v) = 2(D^n u, D^n v) + R_n'(u, v) + 2(u, v) P_n'(\|u\|^2).$$

Recalling Remark 2.3, we have

$$\widetilde{E}_n'(u, u) \ge \|u\|_n^2. \tag{6-6}$$

Now we define the operator H_n so that

$$\widetilde{E}_n'(u,v) = (H_n u, v).$$

Therefore

$$\begin{aligned}(H_n u, u) &= \sum_{m\in\mathbb{Z}} u_m (H_n u, e_m) = \sum_{|m|\le N} u_m (H_n u, e_m) + \sum_{|m|>N} u_m (H_n u, e_m)\\ &\le \frac{\|u\|}{\underline{\lambda}_N}\Big(\sum_{|m|\le N} \lambda_m^2 (H_n u, e_m)^2\Big)^{\frac{1}{2}} + \|H_n u\|\Big(\sum_{|m|>N} u_m^2\Big)^{\frac{1}{2}}\\ &\le \frac{\|u\|_1}{\underline{\lambda}_N}\Big(\sum_{m\in\mathbb{Z}} \lambda_m^2 \widetilde{E}_n'(u,e_m)^2\Big)^{\frac{1}{2}} + \|H_n u\|\frac{\|u\|_1}{N},\end{aligned}$$

where $\underline{\lambda}_N = \min\{\lambda_m \mid |m|\le N\} > 0$ for any $N > 0$. We take into account (6-6) and consider u belonging to

$$K_\epsilon = \Big\{v \;\Big|\; \|v\| \ge \epsilon,\ \|H_n v\| \le \frac{1}{\epsilon}\Big\}.$$

We get

$$\sum_{m\in\mathbb{Z}} \lambda_m^2 \widetilde{E}_n'(u,e_m)^2 \ge \underline{\lambda}_N^2 \Big(\epsilon - \frac{1}{N\epsilon}\Big)^2.$$

The integer N can be chosen to depend on ϵ so that we have

$$\alpha(\epsilon) := \underline{\lambda}_N^2 \Big(\epsilon - \frac{1}{N\epsilon}\Big)^2 > 0.$$

Then, by (6-5)

$$\mu(\{u \mid \widetilde{E}_n(u) \in \Gamma\} \cap K_\epsilon) \le \frac{C}{\alpha(\epsilon)}\ell(\Gamma).$$

Consider now the complementary set

$$K_\epsilon^c = \Big\{u \;\Big|\; \|u\| < \epsilon \text{ or } \|H_n u\| > \frac{1}{\epsilon}\Big\}$$

Since

$$\mathbb{E}\|H_n u\| \le \text{Const.}$$

Using the Chebyshev inequality, we find

$$\mu_\alpha\Big(\Big\{u \;\Big|\; \|H_n u\| > \frac{1}{\epsilon}\Big\}\Big) \le \text{Const}\ \epsilon.$$

By Theorem 6.2, we have that

$$\mu_\alpha(\{u \mid \|u\| < \epsilon\}) \le C\epsilon.$$

Finally we write

$$\mu_\alpha(\{u \mid \widetilde{E}_n(u) \in \Gamma\}) \le \mu(\{u \mid \widetilde{E}_n(u) \in \Gamma\} \cap K_\epsilon) + \mu(K_\epsilon^c) \le \frac{C_1}{\alpha(\epsilon)}\ell(\Gamma) + C_2\epsilon.$$

This, combined with the Portmanteau theorem, proves the absolute continuity of $\widetilde{E}_n(u)$ under μ with respect to the Lebesgue measure on $\mathbb{R}$. □

About the dimension of the measure μ. This subsection is inspired by [Kuksin 2008; Kuksin and Shirikyan 2012], where it was proved that the invariant measures constructed for the Euler equation are not concentrated on a countable union of finite-dimensional compact sets. The proof relies on a Krylov estimate (see Section A.9 of [Kuksin and Shirikyan 2012]) for Itô processes. Roughly speaking, this estimate provides an inequality of the type (6-5) for multidimensional processes. Namely, for a d-dimensional stationary Itô process

$$y_t = y_0 + \int_0^t x_s\, ds + \sum_{j=1}^{\infty} \int_0^t \theta_j(s)\, d\beta_j(s),$$

define the nonnegative $d \times d$-matrix σ with entries

$$\sigma_{m,n} = \sum_{j=1}^{\infty} \theta_j^m \theta_j^n,$$

where θ_j^i is the i-th component of the d-vector θ_j. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a bounded measurable function. Then the Krylov estimate is

$$\mathbb{E} \int_0^1 f(y_t)(\det \sigma_t)^{\frac{1}{d}}\, dt \le C_d |f|_d \mathbb{E} \int_0^1 |x_t|\, dt, \tag{6-7}$$

where $|\cdot|_d$ stands for the L^d-norm and C_d is a constant that only depends on d.

In our context the independence needed to make the Krylov estimate successful leads to solving nonlinear differential equations with order increasing with the size of the underlying vector (process). This is due to the structure of the BO conservation laws and represents a technical difficulty as discussed in the Introduction, while in the Euler case the components of this vector can be chosen to satisfy this independence. We bypass the equation mentioned above in the 2-dimensional case by splitting suitably the phase space.

Theorem 6.3. *The measure μ is of at least 2-dimensional nature in the sense that any compact set of Hausdorff dimension smaller than 2 has μ-measure 0.*

Before proving Theorem 6.3, we describe the general framework.

We use the following splitting of $H^2(\mathbb{T})$:

$$H^2(\mathbb{T}) = O \cup O^c,$$

where

$$O := \left\{ u \;\middle|\; \int u^2 H\, \partial_x^2 u = 0 \right\}. \tag{6-8}$$

Consider the functionals on $\dot{H}^1(\mathbb{T})$ defined by

$$F_j(u) = \frac{1}{j+1} \int u^{j+1}, \quad j = 1, 2.$$

Remark that F_1 is preserved by the BO equation. Now for u a solution of (1-1), we have that

$$\partial_t F_2(u) = 0 \quad \text{on } O.$$

Therefore the vector $F(u) = (F_1(u), F_2(u))$ is constant on O for any solution u of the BO equation.

On the other hand, consider the following BO conservation laws

$$E_0(u) = \int u^2, \quad E_{\frac{1}{2}}(u) = \int uH\,\partial_x u + \frac{1}{3}\int u^3.$$

Set the following preserved vector

$$E(u) = (E_0(u), E_{\frac{1}{2}}(u)).$$

$E(u)$ is in particular constant on O^c for the solutions of (1-1).

Let μ_1 and μ_2 be two measures. We write $\mu_1 \lhd \mu_2$ if there is a continuous increasing function f vanishing at 0 such that

$$\mu_1(\cdot) \le f(\mu_2(\cdot)).$$

This implies the absolute continuity of μ_1 with respect to μ_2. For ν a probability measure on H^2, we define

$$\nu^O(\cdot) = \nu(\cdot \cap O), \quad \nu^{O^c}(\cdot) = \nu(\cdot \cap O^c),$$

where O is the set described before.

Proposition 6.4. *Suppose $\lambda_m \neq 0$ for all $m \in \mathbb{Z}$, then*

(1) $F_*\mu_\alpha^O \lhd \ell_2$, *where* $F = (F_1, F_2)$,

(2) $E_*\mu_\alpha^{O^c} \lhd \ell_2$, *where* $E = (E_0, E_{\frac{1}{2}})$.

The functions describing the absolute continuity do not depend on α, and ℓ_2 is the Lebesgue measure on $\mathbb{R}^2$.

Proof of Theorem 6.3. Let W be an open set of H^2. Clearly

$$W = (W \cap O) \cup (W \backslash O).$$

By Proposition 6.4, we have

$$\mu_\alpha(W) \le f(\ell_2(F(W \cap O))) + g(\ell_2(E(W \backslash O))),$$

where f and g are the functions describing the absolute continuity established in Proposition 6.4. Using the Portmanteau theorem, we get

$$\mu(W) \le f(\ell_2(F(W \cap O))) + g(\ell_2(E(W \backslash O))), \tag{6-9}$$

and by the regularity of μ and ℓ_2 the estimate (6-9) holds for any bounded Borelian set W.

When W is a compact set of Hausdorff dimension $\mathscr{H}(W) < 2$. It is clear that E and F are Lipschitz on any compact set. Since the Lipschitz maps do not increase the Hausdorff dimension, we have the right-hand side of (6-9) is equal to 0, then so is the left-hand side. □

Proof of Proposition 6.4. The proof consists of two steps:

(1) Absolute continuity uniformly in α of μ on the set O. The first and second derivatives of the functionals $F_j(u)$ are

$$F_j'(u, v) = \int u^j v, \quad F_j''(u, v) = j\int u^{j-1}v^2.$$

Then applying the Itô formula to F_j, we find

$$F_j(u) = F_j(u(0)) + \int_0^t A_j(s)\,ds + \sqrt{\alpha}\sum_{m\in\mathbb{Z}}\lambda_m\int_0^t (u^j, e_m)\,d\beta_m(s), \quad j=1,2,$$

where

$$A_j = -(u^j, H\,\partial_x^2 u - \alpha\,\partial_x^2 u) + j\,\frac{\alpha}{2}\sum_{m\in\mathbb{Z}}\lambda_m^2(u^{j-1}, e_m^2).$$

On the set O, we have $(u^j, H\,\partial_x^2 u)=0,\ j=1,2$. Then recalling estimate on $\mathbb{E}\|u\|_2^2$ (Theorem 5.3), we get

$$\mathbb{E}|A_j| \le \alpha\,\text{Const}, \tag{6-10}$$

where Const does not depend on α.

We consider the 2×2-matrix $\sigma(u),\ u\in O$, with entries

$$\sigma_{k,l}(u) = \sum_{m\in\mathbb{Z}}\lambda_m^2(u^k, e_m)(u^l, e_m), \quad k,l=1,2.$$

It is clear that σ is nonnegative. It follows from the Krylov estimate (6-7) with the use of the function $\mathbb{1}_\Gamma$, Γ being a Borel set of $\mathbb{R}^2$, that

$$\mathbb{E}\big[(\det(\sigma(u)))^{\frac{1}{2}}\mathbb{1}_\Gamma(F)\big] \le C\ell_2(\Gamma), \tag{6-11}$$

where ℓ_2 is the Lebesgue measure on $\mathbb{R}^2$ and C does not depend on α.

Now define the map

$$D : \dot{H}^1(\mathbb{T}) \to \mathbb{R}_+,$$
$$u \mapsto \det(\sigma(u)).$$

We remark that D is continuous since it is the composition of continuous maps. We have the following:

Lemma 6.5. *Suppose $\lambda_m \neq 0$ for all $m\in\mathbb{Z}$; then*

$$D(u)=0 \implies u\equiv 0.$$

Proof. Suppose there is a nonzero vector $\gamma=(\gamma_1,\gamma_2)\in\mathbb{R}^2$ such that

$$\gamma\sigma(u)\gamma^T = 0,$$

then

$$\sum_{m\in\mathbb{Z}}\lambda_m^2\Big(\sum_{j=1}^2\gamma_j(u^j, e_m)\Big)^2 = 0.$$

Since $\lambda_m\neq 0$ for all $m\neq 0$, we infer that

$$\sum_{j=1}^2\gamma_j u^j \equiv \text{Const},$$

which is possible only if $u\equiv 0$, taking into account that $\int u = 0$. □

Now define the set

$$J_\epsilon = \Big\{\|u\|_1^2 \ge \epsilon,\ \|u\|_2^2 \le \frac{1}{\epsilon}\Big\} \subset H^2(\mathbb{T}).$$

$J_\epsilon \cap O$ is a compact set in $H^1(\mathbb{T})$ not containing 0; then by the continuity of D, we have $D(J_\epsilon \cap O)$ is a compact set in $\mathbb{R}_+$ not containing 0. Then there is $c_\epsilon > 0$ such that $D(u) \geq c_\epsilon$ for all $u \in J_\epsilon \cap O$. Using the same splitting argument as in the proof of Theorem 6.1, we arrive at the claimed result.

(2) Absolute continuity uniformly in α of μ on the set O^c. We follow the construction above to set a 2×2-matrix M with entries

$$M_{k,l}(u) = \sum_{m \in \mathbb{Z}} \lambda_m^2 B_k(u) B_l(u), \quad k, l = 1, 2,$$

where

$$B_1 = E_0'(u, e_m) \quad \text{and} \quad B_2 = E_{\frac{1}{2}}'(u, e_m).$$

It follows from the Krylov estimate (6-7) that

$$\mathbb{E}\big[(\det(M(u)))^{\frac{1}{2}} \mathbb{1}_\Gamma(E)\big] \leq C \ell_2(\Gamma),$$

where C does not depend on α thanks to the preservation of E_0 and $E_{\frac{1}{2}}$ by the BO flow.

Now $\det M(u) = 0$ only if there is a nonzero vector $(\gamma_1, \gamma_2) \in \mathbb{R}^2$ such that

$$\gamma_1 u + \gamma_2 (2H\, \partial_x u + u^2) \equiv \text{Const}.$$

Note that if $\gamma_2 = 0$, we have that $u \equiv 0$ since $\int u = 0$; therefore $u \in O$. Now we suppose that $\gamma_2 \neq 0$. We differentiate with respect to x to find

$$\gamma_1 \partial_x u + \gamma_2 (2H\, \partial_x^2 u + 2u\, \partial_x u) \equiv 0.$$

Therefore, multiplying by u^p for $p > 0$ and integrating in x, we find

$$\int u^p H\, \partial_x^2 u = 0,$$

and in particular u belongs to the set O. Then on O^c we have $\det(M(u)) \neq 0$. We can follow the same splitting argument with the use of the splitting set J_ϵ defined in the first part to get the result. □

A Gaussian decay property for the measure μ. Here we establish a large deviation bound for the measure μ.

Theorem 6.6. *The measure μ constructed in Theorem 5.3 satisfies*

$$\mathbb{E} e^{\sigma \|u\|^2} < \infty, \tag{6-12}$$

where $\sigma = (aeA_0)^{-1}$ for arbitrary $a > 1$. In particular, for any $r > 0$

$$\mu(\{u \in C^\infty \mid \|u\| > r\}) \leq C e^{-\sigma r^2},$$

where the constant C does not depend on r.

Proof. Recall the estimate (5-4):

$$\mathbb{E}\|u\|^{2p} \le p^p A_0^p.$$

Then

$$\mathbb{E}(\sqrt{\sigma}\|u\|)^{2p} \le \sigma^p p^p A_0^p = \frac{p^p}{a^p e^p}.$$

Now, with use of the Stirling formula, we have

$$\frac{\mathbb{E}(\sqrt{\sigma}\|u\|)^{2p}}{p!} \le \frac{p^p}{p!\,a^p e^p} \sim_{p\to\infty} \frac{1}{a^p\sqrt{2\pi p}}.$$

Since $a > 1$, we have that the series

$$\sum_{p\ge 1} \frac{\mathbb{E}(\sqrt{\sigma}\|u\|)^{2p}}{p!}$$

is convergent, and we are led to (6-12). The other claim is obtained after combining (6-12) with the Chebyshev inequality. □

Remark 6.7. We obtain in the same way the result of Theorem 6.6 for the viscous measures uniformly in α.

Appendix

Proof of Lemma 3.3. Note first that for a solution v of the nonlinear equation (3-3), we have

$$\partial_t E_n(v) = E_n'(v, \partial_t v) = \alpha E_n'(v, \partial_x^2 v) - \sqrt{\alpha} E_n'(v, \partial_x(vz)) - \alpha \tfrac{1}{2} E_n'(v, \partial_x(z^2)), \quad n = 0, 1, 2. \tag{A-1}$$

The E_n are the first three conservation laws of the BO equation.

The case $n = 0$: $E_0'(v, w) = 2\int vw$. Applying (A-1), we get

$$\begin{aligned}
\partial_t E_0(v) + 2\alpha\|v\|_1^2 &= 2\sqrt{\alpha}(v, \partial_x(vz)) + \alpha(v, \partial_x z^2) \\
&= \sqrt{\alpha}(v^2, \partial_x z) + \alpha(v, \partial_x z^2) \\
&\le \sqrt{\alpha}\|z\|_{\frac{3+}{2}}\|v\|^2 + c\alpha\|v\|\,\|z\|_1^2 \\
&\le \sqrt{\alpha}\|z\|_{\frac{3+}{2}}\|v\|^2 + c\alpha(1 + \|v\|^2)\|z\|_1^2.
\end{aligned}$$

Note that $\|z(\cdot)\|_{\frac{3+}{2}}$ is bounded uniformly in α for almost all realizations and in t (on $[0, T]$) by continuity. Then with the use of the Gronwall inequality we get

$$\sup_{t\in[0,T]} \|v(t)\|^2 + 2\alpha\int_0^T \|v(t)\|_1^2\,dt \le C(T, \omega, \|v_0\|). \tag{A-2}$$

The case $n = 1$: Recall that

$$E_1(u) = \int (\partial_x u)^2 + \frac{3}{4}\int u^2 H\,\partial_x u + \frac{1}{8}\int u^4.$$

Then

$$E_1'(v, w) = -2(\partial_x^2 v, w) + \underbrace{\tfrac{3}{2}(vH\partial_x v, w) + \tfrac{3}{4}(v^2, H\partial_x w) + \tfrac{1}{2}(v^3, w)}_{R_1'(v,w)}.$$

It was already shown that (see the more general estimates (2-7) and (2-6))

$$|R_1'(v,\partial_x^2 v)| \le \epsilon\|v\|_2^2 + C_\epsilon\|v\|^c.$$

Then

$$\alpha E_1'(v,\partial_x^2 v) \le -(2-\epsilon)\alpha\|v\|_2^2 + C_\epsilon\alpha\|v\|^c.$$

Taking into account some properties of H, it suffices to treat $(vH\,\partial_x v, w) + (v^3, w)$ instead of $R_1'(v,w)$ for our purpose.

Now

$$\begin{aligned}
\sqrt{\alpha}|(\partial_x^2 v, \partial_x(vz))| &\le C\sqrt{\alpha}\|v\|_2\,\|v\|_1\,\|z\|_1\\
&\le \epsilon\alpha\|v\|_2^2 + C_\epsilon\|v\|_1^2\,\|z\|_1^2 \le \epsilon\alpha\|v\|_2^2 + C_{T,\epsilon,\omega}\|v\|_1^2,\\
\sqrt{\alpha}|(vH\,\partial_x v, \partial_x(vz))| &= \sqrt{\alpha}|(\partial_x(vH\partial_x v), vz)|\\
&\le C\sqrt{\alpha}\|v\|_1\,\|v\|_2\,\|v\|\,\|z\|_{\frac{1+}{2}}\\
&\le \epsilon\alpha\|v\|_2^2 + C_\epsilon\|v\|_1^2\,\|v\|^2\,\|z\|_{\frac{1+}{2}}^2 \le \epsilon\alpha\|v\|_2^2 + C_{T,\epsilon,\omega}\|v\|_1^2,\\
\sqrt{\alpha}|(v^3, \partial_x(vz))| &\le C\sqrt{\alpha}\|v\|_{L^6}^3\,\|v\|_1\,\|z\|_1\\
&\le C\sqrt{\alpha}\|v\|_{\frac{1}{3}}^3\,\|v\|_1\,\|z\|_1\\
&\le C\sqrt{\alpha}\|v\|^2\,\|v\|_1^2\,\|z\|_1 \le \sqrt{\alpha}C_{T,\omega}\|v\|_1^2.
\end{aligned}$$

To summarize, we have

$$\sqrt{\alpha}E_1'(v,\partial_x(vz)) \le \epsilon\|v\|_2^2 + C_{T,\epsilon,\omega}\|v\|_1^2.$$

To estimate the last term, we compute

$$\begin{aligned}
\alpha|(\partial_x^2 v, \partial_x z^2)| &\le C\alpha\|v\|_2\,\|z\|_1^2\\
&\le \epsilon\alpha\|v\|_2^2 + \alpha C_\epsilon\|z\|_1^4,\\
\alpha|(vH\,\partial_x v, \partial_x z^2)| &\le C\alpha\|v\|_2\,\|v\|_1\,\|z\|_{\frac{1}{4}}^2\\
&\le \epsilon\alpha\|v\|_2^2 + \alpha C_{T,\epsilon,\omega}\|v\|_1^2,\\
\alpha|(v^3, \partial_x z^2)| &\le C\alpha\|v\|^2\,\|v\|_1\,\|z\|_1^2\\
&\le \epsilon\alpha\|v\|_1^2 + \alpha C_{T,\epsilon,\omega}.
\end{aligned}$$

To conclude, we can choose ϵ so that

$$E_1(v) + \alpha\int_0^t \|v(r)\|_2^2\,dr \le E_1(v_0) + C_{T,\omega}^1\int_0^t \|v(r)\|_1^2\,dr + C_{T,\omega}^2 t.$$

Recalling the inequality (2-4) and (A-2) we have

$$\|v\|_1^2 + 2\alpha\int_0^t \|v(r)\|_2^2\,dr \le E_1(v_0) + C_{T,\omega}^0 + C_{T,\omega}^1\int_0^t \|v(r)\|_1^2\,dr + C_{T,\omega}^2 t.$$

With the use of the Gronwall lemma, we arrive at

$$\sup_{t\in[0,T]} \|v(t)\|_1^2 + 2\alpha \int_0^T \|v(t)\|_2^2\, dt \le C_{T,\omega}(\|v_0\|_1).$$

The case $n = 2$: Recall that

$$\begin{aligned} E_2(u) = \int (\partial_x^2 u)^2 - \frac{5}{4}\int \big((\partial_x u)^2 H\,\partial_x u + 2\partial_x^2 u H\,\partial_x u\big) \\ + \frac{5}{16}\int \big(5u^2(\partial_x u)^2 + u^2(H\,\partial_x u)^2 + 2uH(\partial_x u)H(u\,\partial_x u)\big) \\ + \int \Big(\frac{5}{32}u^4 H(\partial_x u) + \frac{5}{24}u^3 H(u\,\partial_x u)\Big) + \frac{1}{48}\int u^6. \end{aligned}$$

The form of $E_2(v)$ combined with some properties of H allows us to reduce to the treatment of the quantity

$$R_2(v) = \|v\|_2^2 + \int (\partial_x v)^3 + (\partial_x^2 v, H\,\partial_x v) + (v^2, (\partial_x v)^2) + (v^4, H\,\partial_x v) + \int v^6.$$

Then

$$\begin{aligned} R_2'(v,w) &= 2(\partial_x^2 v, \partial_x^2 w) + 3((\partial_x v)^2, \partial_x w) + 2(\partial_x^2 v, H\,\partial_x w) + 2(vw, (\partial_x v)^2) \\ &\quad + 2(v^2\partial_x v, \partial_x w) + 4(v^3 H\,\partial_x v, w) + (v^4, H\,\partial_x w) + 6(v^5, w) \\ &= 2(\partial_x^2 v, \partial_x^2 w) + R_3'(v,w). \end{aligned}$$

It was already shown in the proof of Lemma 2.2 (see estimates (2-7) and (2-6)) that

$$|R_3'(v, \partial_x^2 v)| \le \epsilon \|v\|_3^2 + C_\epsilon \|v\|^c$$

for some constants $c, C_\epsilon > 0$. Now we have

$$2\alpha(\partial_x^2 v, \partial_x^2(\partial_x^2 v)) = -2\alpha\|v\|_3^2.$$

Then

$$\alpha E_2'(v, \partial_x^2 v) \le -(2-\epsilon)\alpha\|v\|_3^2 + \alpha C_\epsilon \|v\|^c.$$

Now

$$\begin{aligned} \sqrt{\alpha}|(\partial_x^2 v, \partial_x^2(\partial_x(vz)))| &\le C\sqrt{\alpha}\|v\|_3 \|v\|_2 \|z\|_2 \le \epsilon\alpha\|v\|_3^2 + C_{T,\epsilon,\omega}\|v\|_2^2, \\ \sqrt{\alpha}|((\partial_x v)^2, \partial_x^2(vz))| &\le C_{T,\omega}\sqrt{\alpha}\|v\|_{\frac{5}{4}}^2 \|v\|_2 \le C_{T,\omega}\sqrt{\alpha}\|v\|_{\frac{1}{4}} \|v\|_2^2 \le C_{T,\omega}\sqrt{\alpha}\|v\|_2^2, \\ \sqrt{\alpha}|(\partial_x^2(vz), H\,\partial_x^2 v)| &\le \sqrt{\alpha}C\|v\|_2^2 \|z\|_2 \le \sqrt{\alpha}C_{T,\omega}\|v\|_2^2, \\ \sqrt{\alpha}|(v(\partial_x v)^2, \partial_x(vz))| &\le C\sqrt{\alpha}\|v\| \|z\|_{\frac{1}{2}+} \|v\|_2^2 \le C_{T,\omega}\sqrt{\alpha}\|v\|_2^2, \\ \sqrt{\alpha}|(v^2\partial_x v, \partial_x^2(vz))| &\le \sqrt{\alpha}C\|v\|_1^3 \|v\|_2 \|z\|_2 \le C_{T,\omega}\sqrt{\alpha}\|v\|_2^2, \\ \sqrt{\alpha}|(v^3 H\partial_x v, \partial_x(vz))| &\le C\sqrt{\alpha}\|v\|_1^3 \|v\|_2 \|z\|_{\frac{1}{2}+} \|v\| \le \sqrt{\alpha}C_{T,\omega}\|v\|_2^2, \\ \sqrt{\alpha}|(v^4, H\,\partial_x^2(vz))| &\le \sqrt{\alpha}C\|v\|_1^5 \|z\|_1 \le \sqrt{\alpha}C_{T,\omega}, \\ \sqrt{\alpha}|(v^5, \partial_x(vz))| &\le C\sqrt{\alpha}\|v\|_1^5 \|v\| \|z\|_{\frac{1}{2}+} \le \sqrt{\alpha}C_{T,\omega}. \end{aligned}$$

The estimates concerning the term $\partial_x z^2$ are easier because they do not contain v. Finally, using the same argument as before (in the case of $E_1(v)$), we arrive at the claimed result.

The periodic Hilbert transform. We present in this section a definition of the Hilbert transform in the periodic setting and establish some of its elementary properties. Recall that the sequence defined by

$$e_n(x) = \begin{cases} \sin(nx)/\sqrt{\pi} & \text{if } n < 0, \\ \cos(nx)/\sqrt{\pi} & \text{if } n > 0, \end{cases}$$

forms a Hilbertian basis of $\dot{H}(\mathbb{T})$; let us denote this basis by $\mathscr{B}$. We define the Hilbert transform on $\mathscr{B}$ by

$$He_n(x) = \operatorname{sgn}(n)\, e_{-n}(x),$$

where

$$\operatorname{sgn}(p) = \begin{cases} 1 & \text{if } p > 0, \\ 0 & \text{if } p = 0, \\ -1 & \text{if } p < 0. \end{cases}$$

We first remark that H defines an isometry on $\dot{H}$.

Proposition A.1. *Let $f, g \in \dot{H}(\mathbb{T})$. Then*

$$H^2 f = -f, \tag{A-3}$$

$$\int_{\mathbb{T}} Hf = 0, \tag{A-4}$$

$$(g, Hf) = -(Hg, f), \tag{A-5}$$

$$\widehat{HF}_0(p) = -\mathrm{i}\operatorname{sgn}(p)\hat{f}_0(p), \tag{A-6}$$

where $\hat{h}_0$ denotes the complex Fourier coefficient of a function h, defined below.

Define now the Fourier coefficients associated to a function f in $\dot{H}$:

$$\hat{f}_1(n) = \frac{1}{\sqrt{\pi}} \int_{\mathbb{T}} \cos(nx) f(x)\, dx,$$
$$\hat{f}_2(n) = \frac{1}{\sqrt{\pi}} \int_{\mathbb{T}} \sin(nx) f(x)\, dx.$$

The function f is represented in $\mathscr{B}$ as follows:

$$f(x) = \sum_{n>0} \big(\hat{f}_1(n) e_n(x) - \hat{f}_2(n) e_{-n}(x)\big). \tag{A-7}$$

Hence the Hilbert transform of f can be expressed as

$$Hf(x) = \sum_{n>0} \big(\hat{f}_1(n) e_{-n}(x) + \hat{f}_2(n) e_n(x)\big). \tag{A-8}$$

The complex Fourier coefficient is defined by

$$\hat{f}_0(p) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{T}} e^{-ipx} f(x)\, dx. \tag{A-9}$$

The relationship between the three Fourier coefficients of f is

$$\hat{f}_0(p) = \frac{\hat{f}_1(p) - \mathrm{i}\,\mathrm{sgn}(p)\hat{f}_2(p)}{\sqrt{2}}. \tag{A-10}$$

Proof of Proposition A.1. Equation (A-4) follows immediately from (A-8). Now from (A-7) and (A-8), we can easily deduce that

$$H^2 f(x) = -\sum_{n>0} \big(\hat{f}_1(n)e_n(x) - \hat{f}_2(n)e_{-n}(x)\big) = -f(x).$$

and (A-3) is shown.

From (A-8), we infer that

$$\widehat{HF}_1(p) = -\hat{f}_2(p), \quad \widehat{HF}_2(p) = \hat{f}_1(p).$$

Thus using (A-10), we can write

$$\begin{aligned}\widehat{HF}_0(p) &= \frac{-\hat{f}_2(p) - \mathrm{i}\,\mathrm{sgn}(p)\hat{f}_1(p)}{\sqrt{2}} \\ &= \frac{-\mathrm{i}\,\mathrm{sgn}(p)(\hat{f}_1(p) - \mathrm{i}\,\mathrm{sgn}(p)\hat{f}_2(p))}{\sqrt{2}} = -\mathrm{i}\,\mathrm{sgn}(p)\hat{f}_0(p),\end{aligned}$$

and we have arrived at (A-6).

To prove (A-5), we compute

$$\begin{aligned}(g, Hf) &= \sum_{n>0} \hat{f}_1(n) \int_{\mathbb{T}} g(x)e_{-n}(x)\, dx + \sum_{n>0} \hat{f}_2(n) \int_{\mathbb{T}} g(x)e_n(x)\, dx \\ &= -\sum_{n>0} \hat{f}_1(n)\hat{g}_2(n) + \sum_{n>0} \hat{g}_1(n)\hat{f}_2(n) \\ &= -\sum_{n>0} \hat{g}_2(n) \int_{\mathbb{T}} f(x)e_n(x)\, dx - \sum_{n>0} \hat{g}_1(n) \int_{\mathbb{T}} f(x)e_{-n}(x)\, dx \\ &= -\int_{\mathbb{T}} f(x) \sum_{n>0} (\hat{g}_1(n)e_{-n}(x) + \hat{g}_2(n)e_n(x)) \\ &= -(Hg, f).\end{aligned}$$

□

Acknowledgements

I thank my advisors Armen Shirikyan and Nikolay Tzvetkov for useful discussions and valuable remarks. I am grateful to the referee for valuable remarks that were very useful for improving the text. This research was supported by the program DIM RDMath of “Région Ile-de-France”.

References

[Abdelouhab et al. 1989] L. Abdelouhab, J. L. Bona, M. Felland, and J.-C. Saut, "Nonlocal models for nonlinear, dispersive waves", *Phys. D* **40**:3 (1989), 360–392. MR Zbl

[Deng 2015] Y. Deng, "Invariance of the Gibbs measure for the Benjamin–Ono equation", *J. Eur. Math. Soc.* (*JEMS*) **17**:5 (2015), 1107–1198. MR Zbl

[Deng et al. 2015] Y. Deng, N. Tzvetkov, and N. Visciglia, "Invariant measures and long time behaviour for the Benjamin–Ono equation, III", *Comm. Math. Phys.* **339**:3 (2015), 815–857. MR Zbl

[Karatzas and Shreve 1991] I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, 2nd ed., Graduate Texts in Mathematics **113**, Springer, 1991. MR Zbl

[Kuksin 2004] S. B. Kuksin, "The Eulerian limit for 2D statistical hydrodynamics", *J. Statist. Phys.* **115**:1-2 (2004), 469–492. MR Zbl

[Kuksin 2008] S. B. Kuksin, "On distribution of energy and vorticity for solutions of 2D Navier–Stokes equation with small viscosity", *Comm. Math. Phys.* **284**:2 (2008), 407–424. MR Zbl

[Kuksin and Shirikyan 2004] S. Kuksin and A. Shirikyan, "Randomly forced CGL equation: stationary measures and the inviscid limit", *J. Phys. A* **37**:12 (2004), 3805–3822. MR Zbl

[Kuksin and Shirikyan 2012] S. Kuksin and A. Shirikyan, *Mathematics of two-dimensional turbulence*, Cambridge Tracts in Mathematics **194**, Cambridge University Press, 2012. MR Zbl

[Matsuno 1984] Y. Matsuno, *Bilinear transformation method*, Mathematics in Science and Engineering **174**, Academic Press, Orlando, FL, 1984. MR Zbl

[Molinet 2008] L. Molinet, "Global well-posedness in L^2 for the periodic Benjamin–Ono equation", *Amer. J. Math.* **130**:3 (2008), 635–683. MR Zbl

[Molinet and Pilod 2012] L. Molinet and D. Pilod, "The Cauchy problem for the Benjamin–Ono equation in L^2 revisited", *Anal. PDE* **5**:2 (2012), 365–395. MR Zbl

[Shirikyan 2011] A. Shirikyan, "Local times for solutions of the complex Ginzburg–Landau equation and the inviscid limit", *J. Math. Anal. Appl.* **384**:1 (2011), 130–137. MR Zbl

[Tzvetkov and Visciglia 2013] N. Tzvetkov and N. Visciglia, "Gaussian measures associated to the higher order conservation laws of the Benjamin–Ono equation", *Ann. Sci. Éc. Norm. Supér.* (4) **46**:2 (2013), 249–299. MR Zbl

[Tzvetkov and Visciglia 2014] N. Tzvetkov and N. Visciglia, "Invariant measures and long-time behavior for the Benjamin–Ono equation", *Int. Math. Res. Not.* **2014**:17 (2014), 4679–4714. MR Zbl

[Tzvetkov and Visciglia 2015] N. Tzvetkov and N. Visciglia, "Invariant measures and long time behaviour for the Benjamin–Ono equation, II", *J. Math. Pures Appl.* (9) **103**:1 (2015), 102–141. MR Zbl

[Zhidkov 2001a] P. E. Zhidkov, *Korteweg–de Vries and nonlinear Schrödinger equations: qualitative theory*, Lecture Notes in Mathematics **1756**, Springer, 2001. MR Zbl

[Zhidkov 2001b] P. E. Zhidkov, "On an infinite sequence of invariant measures for the cubic nonlinear Schrödinger equation", *Int. J. Math. Math. Sci.* **28**:7 (2001), 375–394. MR Zbl

Received 14 Nov 2016. Revised 10 Jan 2018. Accepted 14 Feb 2018.

MOUHAMADOU SY: mouhamadou.sy@u-cergy.fr
Laboratoire AGM UMR 8088 CNRS, Université de Cergy-Pontoise, Cergy-Pontoise, France

ANALYSIS AND PDE
Vol. 11, No. 8, 2018
dx.doi.org/10.2140/apde.2018.11.1881

RIGIDITY OF MINIMIZERS IN NONLOCAL PHASE TRANSITIONS

OVIDIU SAVIN

We obtain the classification of certain global bounded solutions for semilinear nonlocal equations of the type

$$\Delta^s u = W'(u) \quad \text{in } \mathbb{R}^n, \quad \text{with } s \in \left(\tfrac{1}{2}, 1\right),$$

where W is a double-well potential.

1. Introduction

We extend to the case of the fractional Laplacian Δ^s with $s \in \left(\frac{1}{2}, 1\right)$ the results from [Savin 2009; 2017] concerning a conjecture of De Giorgi about the classification of certain global bounded solutions for semilinear equations of the type

$$\Delta u = W'(u),$$

where W is a double-well potential.

We consider the Ginzburg–Landau energy functional with nonlocal interactions

$$J(u, \Omega) = \frac{1}{4} \int_{\mathbb{R}^n \times \mathbb{R}^n \setminus (C\Omega \times C\Omega)} \frac{(u(x) - u(y))^2}{|x - y|^{n+2s}} \, dx \, dy + \int_\Omega W(u) \, dx,$$

with $|u| \le 1$. Here W is a double-well potential with minima at 1 and -1 satisfying

$$W \in C^2([-1, 1]), \quad W(-1) = W(1) = 0, \quad W > 0 \text{ on } (-1, 1),$$
$$W'(-1) = W'(1) = 0, \quad W''(-1) > 0, \quad W''(1) > 0.$$

The classical double-well potential W to have in mind is

$$W(s) = \tfrac{1}{4}(1 - s^2)^2.$$

Physically $u \equiv -1$ and $u \equiv 1$ represent the stable "phases". A critical function for the energy J corresponds to a phase transition with nonlocal interaction between these states, and it satisfies the Euler–Lagrange equation

$$\Delta^s u = W'(u),$$

where $\Delta^s u$ is defined as

$$\Delta^s u(x) = \mathrm{PV} \int_{\mathbb{R}^n} \frac{u(y) - u(x)}{|y - x|^{n+2s}} \, dy.$$

The author was partially supported by NSF Grant DMS-1500438.
MSC2010: 35J61.
Keywords: nonlocal phase transitions, De Giorgi conjecture.

Our main result provides the classification of minimizers with asymptotically flat level sets.

Theorem 1.1. *Let u be a global minimizer of J in $\mathbb{R}^n$ with $s \in \left(\frac{1}{2}, 1\right)$. If the 0 level set $\{u = 0\}$ is asymptotically flat at ∞, then u is one-dimensional.*

The hypothesis that $\{u = 0\}$ is asymptotically flat means that there exist sequences of positive numbers θ_k, l_k and unit vectors ξ_k with $l_k \to \infty$, $\theta_k l_k^{-1} \to 0$, such that

$$\{u = 0\} \cap B_{l_k} \subset \{|x \cdot \xi_k| < \theta_k\}.$$

By saying that u is one-dimensional we understand that u depends only on one direction ξ; i.e., $u = g(x \cdot \xi)$.

A more quantitative version of Theorem 1.1 is given in Theorem 6.1.

In a subsequent work [Savin 2018] we will treat also the case $s = \frac{1}{2}$, which requires some modifications of the methods presented in this paper. We remark that Theorem 1.1 when $s \in \left(0, \frac{1}{2}\right)$ was obtained recently by Dipierro, Serra and Valdinoci [2016].

It is known that blowdowns of the level set $\{u = 0\}$ have different behavior depending on the value of s. If $s \geq \frac{1}{2}$, there are sequences $\varepsilon_k\{u = 0\}$ with $\varepsilon_k \to 0$ that converge uniformly on compact sets to a minimal surface and, if $s < \frac{1}{2}$ they converge to an s-nonlocal minimal surface. This follows from a Γ-convergence result together with a uniform density estimate of level sets of minimizers which were obtained by the author and Valdinoci in [Savin and Valdinoci 2012; 2014]; see for example Corollary 1.7 in the latter paper.

From the classification of global minimal surfaces in low dimensions we find that the level sets of minimizers of J are always asymptotically flat at ∞ in dimension $n \leq 7$ if $s \geq \frac{1}{2}$, and we obtain the following corollary of Theorem 1.1.

Theorem 1.2. *A global minimizer of J is one-dimensional in dimension $n \leq 7$ if $s \in \left(\frac{1}{2}, 1\right)$.*

Another consequence of Theorem 1.1 is the following version of De Giorgi's conjecture to the fractional Laplace case.

Theorem 1.3. *Let $u \in C^2(\mathbb{R}^n)$ be a solution of*

$$\Delta^s u = W'(u), \tag{1-1}$$

with $s \in \left(\frac{1}{2}, 1\right)$, such that

$$|u| \leq 1, \quad \partial_n u > 0, \quad \lim_{x_n \to \pm\infty} u(x', x_n) = \pm 1. \tag{1-2}$$

Then u is one-dimensional if $n \leq 8$.

Theorems 1.2 and 1.3 without the limit assumption in (1-2) have been proved in two and three dimensions using stability inequality methods. In dimension $n = 3$ and for $s \geq \frac{1}{2}$ they have been established by Cabre and Cinti [2014], and in dimension $n = 2$ for all $s \in (0, 1)$ by Sire and Valdinoci [2009]; see also [Cabré and Cinti 2010; Cabré and Sire 2015; Cabré and Solà-Morales 2005]. The case $n = 3$ and $s \in \left(0, \frac{1}{2}\right)$ was also addressed recently by S. Dipierro, A. Farina, and E. Valdinoci [Dipierro et al. 2018].

It is not difficult to show that the ± 1 limit assumption implies that u is a global minimizer in $\mathbb{R}^n$; see for example Theorem 1 in [Palatucci et al. 2013]. Since $\{u = 0\}$ is a graph, it is asymptotically flat in dimension $n \leq 8$ and Theorem 1.1 applies.

Similarly we see that if the 0 level set is a graph in the x_n-direction which has a one-sided linear bound at ∞ then the conclusion is true in any dimension.

Theorem 1.4. *If u satisfies (1-1), (1-2),*

$$\{u = 0\} \subset \{x_n < C(1 + |x'|)\},$$

and $s \in \left(\frac{1}{2}, 1\right)$ then u is one-dimensional.

Our proof of Theorem 1.1 follows closely the one for the classical Laplacian given in [Savin 2017]. The main steps consist in (1) finding some appropriate families of radial subsolutions, (2) applying a version of the weak Harnack inequality and (3) a Γ-convergence result. Some new technicalities are present in our setting due to the nonlocal nature of the equation. For example in the improvement-of-flatness property Theorem 6.1, we need to impose a geometric restriction to the level set $\{u = 0\}$ possibly outside the flat cylinder $\mathcal{C}(l, \theta)$.

It turns out that when $s \in \left(\frac{1}{2}, 1\right)$, the level sets of u satisfy a local curvature estimate. For example, at a point of $\{u = 0\}$ which has a large ball of radius R tangent from one side, we can estimate its curvatures in terms of R^{-1} (see Lemma 4.3). In the borderline case $s = \frac{1}{2}$ the curvature bound requires a logarithmic correction and the same methods no longer apply.

We prove Theorem 1.1 by making use of the extension property of the fractional Laplacian of [Caffarelli and Silvestre 2007]. Precisely we consider the extension $U(x, y)$ of $u(x)$ in $\mathbb{R}^{n+1}_+$ such that

$$\operatorname{div}(y^a \nabla U) = 0 \quad \text{in } \mathbb{R}^{n+1}_+, \qquad U(x, 0) = u(x), \qquad a := 1 - 2s \in (-1, 1),$$

and then

$$\Delta^s u(x) = c_{n,s} \lim_{y \to 0^+} y^a U_y(x, y),$$

with $c_{n,s}$ a constant that depends only on n and s. Then global minimizers of $J(u)$ in $\mathbb{R}^n$ with $|u| \leq 1$ correspond to global minimizers of the "extension energy" $\mathcal{J}(U)$ with $|U| \leq 1$, where

$$\mathcal{J}(U) := \frac{c_{n,s}}{2} \int |\nabla U|^2 y^a \, dx \, dy + \int W(u) \, dx.$$

After dividing by a constant and relabeling W, we may fix $c_{n,s}$ to be 1. We obtain an improvement-of-flatness property for the level sets of minimizers of $\mathcal{J}$ which are defined in large balls $\mathcal{B}^+_R$; see Theorem 6.1. We remark that the principal use of the extension is to make the various subsolution computations easier to handle and it is not essential to the method of proof.

The paper is organized as follows. In Sections 2 and 3 we introduce some notation and then construct a family of axial subsolutions. In Section 4 we provide certain "viscosity solution" properties of the level set $\{u = 0\}$. In Section 5 we obtain a Harnack inequality of the 0 level set and in Section 6 we prove Theorem 6.1.

2. Notation and preliminaries

We introduce the following notation:

We denote points in $\mathbb{R}^n$ as $x = (x', x_n)$ with $x' \in \mathbb{R}^{n-1}$. The ball of center z and radius r is denoted by $B_r(z)$,

$$B_r(z) := \{x \in \mathbb{R}^n : |x - z| < r\}, \quad B_r := B_r(0).$$

The cylinder with base l and height θ is denoted by $\mathcal{C}(l, \theta) \subset \mathbb{R}^n$,

$$\mathcal{C}(l, \theta) := \{x : |x'| \le l,\ |x_n| \le \theta\}.$$

Points in the extension variables $\mathbb{R}^{n+1}_+$ are denoted by (x, y) with $y > 0$, and the ball of radius r as $\mathcal{B}^+_r$,

$$\mathcal{B}^+_r := \{(x, y) \in \mathbb{R}^{n+1}_+ : |(x, y)| < r\} \subset \mathbb{R}^{n+1}.$$

Given a function $U(x, y)$, we define u to be its trace on $\{y = 0\}$,

$$u(x) = U(x, 0).$$

Also let

$$a := 1 - 2s \in (-1, 0),$$

and

$$\Delta_a U := \Delta U + a\frac{U_y}{y} = y^{-a} \operatorname{div}(y^a \nabla U),$$

$$\partial_y^{1-a} U(x) := \lim_{y \to 0^+} y^a U_y(x, y) = \frac{1}{1-a} \lim_{y \to 0^+} y^{a-1}(U(x, y) - U(x, 0)).$$

We define the energy $\mathcal{J}$ as

$$\mathcal{J}(U, \mathcal{B}^+_R) := \frac{1}{2} \int_{\mathcal{B}^+_R} |\nabla U|^2 y^a \, dx \, dy + \int_{B_r} W(u) \, dx,$$

and a critical function U for $\mathcal{J}$ satisfies the Euler–Lagrange equation

$$\Delta_a U = 0, \quad \partial_y^{1-a} U = W'(u). \tag{2-1}$$

In [Palatucci et al. 2013, Theorem 2], see also [Cabré and Sire 2014], they proved the existence and uniqueness up to translations of a global minimizer of $\mathcal{J}$ in two dimensions which is increasing in the first variable and which has limits ± 1 at infinity. Precisely there exists a unique $G : \mathbb{R}^2_+ \to (-1, 1)$ that solves (2-1) such that $G(t, y)$ is increasing in the t-variable and its trace $g(t) := G(t, 0)$ satisfies

$$g(0) = 0, \quad \lim_{t \to \pm\infty} g(t) = \pm 1.$$

Moreover, g and g' have the asymptotic behavior

$$1 - |g| \sim \min\{1, |t|^{-2s}\}, \quad g' \sim \min\{1, |t|^{-1-2s}\},$$

and since $a \in (-1, 0)$ we have $\mathcal{J}(G, \mathbb{R}^2_+) < \infty$.

Since $\Delta_a G_t = 0$ and $G_t \geq 0$, we easily conclude that

$$|\nabla G| \leq C \min\{1, r^{-1}\}, \quad G_t \geq c\, r^{-1-2s}, \tag{2-2}$$

where r denotes the distance to the origin in the (t, y)-plane.

In Theorem 6.1 we show that the only global minimizer of $\mathcal{J}$ that has asymptotically flat level sets on $y = 0$ is $G(x_n, y)$ up to translations and rotations.

For simplicity of notation we assume that W is uniformly convex outside the interval $[g(-1), g(1)]$.

Constants that depend on n, s, W, G are called universal constants, and we denote them by C, c. In the course of the proofs, the values of C, c may change from line to line when there is no possibility of confusion. If the constants depend on other parameters, say θ, ρ, then we denote them by $C(\theta, \rho)$ etc.

3. Two-dimensional barriers

We construct two families of comparison functions G_R and Ψ_R which are perturbations of the solution G.

Lemma 3.1 (radial supersolutions). *For all large R, there exist continuous functions $G_R : \mathbb{R}^2 \to (-1, 1]$ and universal constants $\delta > 0$ small, C large such that*

(1) $G_R = 1$ *outside* $\mathcal{B}^+_{R^{1-\delta}} \cup ((-\infty, 0] \times [0, R^{1-\delta}])$,

(2) $G_R(t, y)$ *is nondecreasing in* t, *and* $\partial_t G_R = 0$ *outside* $\mathcal{B}^+_{R^{1-\delta}}$,

(3) $$|G_R - G| \leq \frac{C}{R} \quad \text{in } \mathcal{B}^+_4,$$

(4) $$\Delta_a G_R + \frac{2(n-1)}{R} |\nabla G_R| \leq 0,$$

and on $y = 0$,

$$\partial_y^{1-a} G_R < W'(G_R) \quad \text{if } t \notin [-1, 1].$$

The inequalities in (4) are understood in the viscosity sense.

Notice that by (2-2), property (3) implies

$$G_R(t, y) \leq G\left(t + \frac{C'}{R}, y\right) \quad \text{in } \mathcal{B}^+_4.$$

We remark that property (3) and the inequality above hold in any ball $\mathcal{B}^+_K$, for a fixed large constant K, provided that we replace C/R, C'/R by $C(K)/R$, $C'(K)/R$.

Proof. We begin with the following claim whose proof we provide at the end.

Claim. For each $\alpha \in (1, 1-a)$ there exists H a homogeneous function of degree α such that

$$H \geq r^\alpha, \quad \Delta_a H \leq -r^{\alpha-2}, \quad |\nabla H| \leq C r^{\alpha-1}, \quad \partial_y^{1-a} H \leq C |t|^{\alpha-(1-a)}.$$

Here r denotes the distance to the origin and $C = C(\alpha)$ depends on the universal constants and α.

Fix such an α and define

$$H_R := \min\left\{ G + \frac{C_0}{R}(H + C_1), 1 \right\}, \tag{3-1}$$

with C_0, C_1 large constants to be specified later.

We define G_R as the infimum over all left translations of H_R; i.e.,

$$G_R(t, y) = \inf_{l \geq 0} H_R(t + l, y).$$

Since $|G| < 1$ we have $H_R > -1$, and $H_R = 1$ outside $\mathcal{B}^+_{R^{1-\delta}}$ provided that δ is chosen sufficiently small such that $(1-\delta)\alpha > 1$. Properties (1) and (2) are clearly satisfied.

Notice that H is increasing in a band $[C, \infty) \times [0, 4]$ and we obtain that H_R is increasing in $[-4, \infty) \times [0, 4]$. This gives $G_R = H_R$ in $\mathcal{B}^+_4$ and property (3) is satisfied.

The properties of H and (2-2) imply that in the set $\{H_R < 1\}$ we have

$$|\nabla H_R| \leq C \min\{1, r^{-1}\} + CC_0 R^{-1} r^{\alpha - 1},$$

and

$$\Delta_a H_R \leq -C_0 R^{-1} r^{\alpha - 2}.$$

Then the first inequality in (4) holds for H_R provided that C_0 is chosen sufficiently large, and therefore holds also for G_R as the infimum over translations of H_R.

On $y = 0$ in the set $\{H_R < 1\}$ we have

$$\partial_y^{1-a} H_R = \partial_y^{1-a} G + C_0 R^{-1} \partial_y^{1-a} H \leq W'(G) + CR^{-1}|t|^{\alpha-(1-a)}.$$

From the behavior of g and g' for large t, we see that the minimum of $H_R(t, 0)$ occurs at some $t = q_R \sim -R^{1/(2s+\alpha)} \ll -1$ and

$$\|(H_R - G)(t, 0)\|_{L^\infty([q_R, \infty))} \to 0 \quad \text{as } R \to \infty.$$

Since $W'' \geq c$ outside $[g(-1), g(1)]$ we find that when $t \in [q_R, \infty) \setminus [-1, 1]$ and $\{H_R < 1\}$ we have

$$W'(H_R) - W'(G) \geq \tfrac{1}{2} c (H_R - G) \geq c' R^{-1}(|t|^\alpha + C_1);$$

thus, if C_1 is sufficiently large,

$$\partial_y^{1-a} H_R < W'(H_R) \quad \text{in } [q_R, \infty) \setminus [-1, 1].$$

Now the second inequality of (4) is satisfied by G_R as the infimum of left translations of H_R. □

Proof of Claim. We find H as a perturbation of the function Cy^α near $y = 0$. Notice that y^{1-a} is Δ_a-harmonic; thus y^α is Δ_a-superharmonic for $\alpha < 1 - a$. However, Cy^α does not satisfy the first and last properties given in the claim.

We write H in polar coordinates as $H = r^\alpha h(\theta)$, with h an even function with respect to $\frac{\pi}{2}$, and then

$$r^{2-\alpha} \Delta_a H = h'' + \alpha(\alpha + a)h + a \cot\theta \; h', \tag{3-2}$$

$$\partial_y^{1-a} H = r^{\alpha-(1-a)} \, \partial_\theta^{1-a} h. \tag{3-3}$$

For all small σ, the function

$$h_\sigma = \sigma + \theta^{1-a} - \theta^2$$

gives a negative right-hand side in (3-2) when θ belongs to a small fixed interval $[0, c]$. We choose first M large and then σ small such that the graphs of Mh_σ and $(\sin\theta)^\alpha$ become tangent by above at some point in the interval $[0, c]$. We "glue" parts of the two graphs in a single graph of a $C^{1,1}$ function $\tilde{h}$. Now it is easy to check that all properties hold by taking h to be a large multiple of $\tilde{h}$. □

From the construction of H_R, G_R we see that both of them decrease with R as we increase R.

Next we construct a similar family Ψ_R with a slightly slower decay in R than G_R. This allows us to have more flexibility in the choice of the two-dimensional profiles of explicit supersolutions. In the next lemma we compare two such profiles Ψ_R and $G_{\bar{R}}$ when R and $\bar{R} \gg R$ have different orders of magnitude. This is an important tool in the proof of the key Propositions 4.6 and 4.7 from next section, where two explicit supersolutions need to be compared in a certain region.

Lemma 3.2. *There exist functions G_R and Ψ_R that satisfy the properties (1)–(4) of Lemma 3.1 for some δ, C universal such that*

$$G_R(t + R^{-\sigma}, y) \geq \Psi_{R^{1-\sigma}}(t, y),$$

with $\sigma \in \left(0, \frac{\delta}{3}\right)$ small universal.

Proof. Denote by $G_{R,\alpha}$ the function constructed in Lemma 3.1.

We choose $G_R := G_{R,\alpha}$, $\Psi_R := G_{R,\beta}$ for some fixed α, β such that $1 < \beta < \alpha < 1 - a$. We take

$$\delta = \min\{\delta(\alpha), \delta(\beta)\} \quad \text{and} \quad C = \max\{C(\alpha), C(\beta)\}$$

and then Lemma 3.1 holds for both G_R and Ψ_R with the same constants δ and C.

We show that

$$H_{R,\alpha}(t + R^{-\sigma}, y) \geq H_{R^{1-\sigma},\beta}(t, y),$$

with $H_{R,\alpha}$ defined as in (3-1), and the lemma follows by taking the infimum over the left translations.

In the inequality above it suffices to restrict to the set where $\{H_{R,\alpha} < 1\}$. We have

$$H_R \geq G + R^{-1}(c_1 r^\alpha + c_2)$$

for some constants c_1, c_2 depending on α. After a translation of $R^{-\sigma}$ we obtain, see (2-2),

$$H_R(t + R^{-\sigma}, y) \geq G(t, y) + cR^{-\sigma} \min\{1, r^{-1-2s}\} + \tfrac{1}{2}R^{-1}(c_1 r^\alpha + c_2).$$

When $r \geq 1$ we use the inequality $a + b \geq a^\mu b^{1-\mu}$ for $\mu > 0$ small, and we find

$$H_R(t + R^{-\sigma}, y) \geq G(t, y) + c(\alpha)R^{-\eta}(r^\gamma + 1), \tag{3-4}$$

with

$$\gamma = \alpha(1-\mu) - \mu(1+2s), \quad \eta = 1 - \mu + \sigma\mu$$

(and $\eta > \sigma$). We choose μ small and then σ such that $\gamma > \beta$ and $\eta < 1-\sigma$. Then the right-hand side of (3-4) is greater than

$$G + R^{\sigma-1}(C_1(\beta)r^\beta + C_2(\beta)) \geq H_{R^{1-\sigma},\beta}$$

for all large R, and the lemma is proved. □

Remark 3.3. Using the monotonicity of Ψ_r with respect to r, we have

$$G_R(s + R^{-\sigma}, y) \geq \Psi_r(s, y) \quad \text{for all } r \geq R^{1-\sigma}.$$

4. Estimates for $\{u = 0\}$

We now derive properties of the level sets of solutions to

$$\Delta_a U = 0, \quad \partial_y^{1-a} U = W'(U), \tag{4-1}$$

which are defined in large domains.

In the next lemma we find axial approximations to the two-dimensional solution G.

Lemma 4.1 (axial approximations). *Let $G_R : \mathbb{R}^2_+ \to (-1, 1]$ be the function constructed in Lemma 3.2. Then its axial rotation in $\mathbb{R}^{n+1}$*

$$\Phi_R(x, y) := G_R(|x| - R, y)$$

satisfies

(1) $\Phi_R = 1$ *outside* $\mathcal{B}^+_{R+R^{1-\delta}}$,

(2) $$\Delta_a \Phi_R \leq 0 \quad \text{in } \mathbb{R}^{n+1}_+,$$

and

$$\partial_y^{1-a} \Phi_R < W'(\Phi_R) \quad \text{when } |x| - R \notin [-1, 1].$$

Let $\phi_R(x) = \Phi_R(x, 0)$ denote the trace of Φ_R on $\{y = 0\}$. Notice that ϕ_R is radially increasing, and $\{\phi_R = 0\}$ is a sphere which is in a C/R-neighborhood of the sphere of radius R.

Proof. We have

$$\Delta_a \Phi_R(x, y) = \Delta_a G_R(s, y) + \frac{n-1}{R+s}\, \partial_s G_R(s, y), \quad s = |x| - R,$$

$$\partial_y^{1-a} \Phi_R(x, 0) = \partial_y^{1-a} G_R(s, 0).$$

The conclusion follows from Lemma 3.2 since $\partial_s G_R = 0$ when $|s| \geq R^{1-\delta}$ and $R + s > \frac{1}{2}R$ when $|s| < R^{1-\delta}$. □

Definition 4.2. We denote by $\Phi_{R,z}$ the translation of Φ_R by z; i.e.,

$$\Phi_{R,z}(x, y) := \Phi_R(x - z, y) = G_R(|x - z| - R, y).$$

Similarly we define $\Psi_{R,z}$ to be the axial rotation of the other two-dimensional solution Ψ_R given in Lemma 3.2,

$$\Psi_{R,z}(x, y) := \Psi_R(|x - z| - R, y).$$

Clearly $\Psi_{R,0}$ satisfies properties (1), (2) of Lemma 4.1.

We recall that we use ϕ, ψ to denote the traces of Φ and Ψ.

Sliding the graph of Φ_R. Assume that u is less than ϕ_{R,x_0} in $B_{2R}(x_0)$. By the maximum principle we obtain $U < \Phi_{R,z}$ with $z = x_0$ in $\mathcal{B}_{2R}(x_0, 0)$ (and therefore globally). We translate the function Φ_R above by moving continuously the center z, and let's assume that it touches U by above, say for simplicity when $z = 0$; i.e., the strict inequality becomes equality for some contact point (x^*, y^*). From Lemma 4.1 we know that Φ_R is a strict supersolution away from $\{y = 0\}$, and moreover the contact point must satisfy $y^* = 0$, $|x^*| - R \in [-1, 1]$; that is, it belongs to the annular region $B_{R+1} \setminus B_{R-1}$ in the n-dimensional subspace $\{y = 0\}$.

Lemma 4.3 (estimates near a contact point). *Assume that the graph of Φ_R touches by above the graph of U at a point $(x^*, 0, u(x^*))$ with $x^* \in B_{R+1} \setminus B_{R-1}$. Let $\pi(x^*)$ be the projection of x^* onto the sphere ∂B_R. Then in $\mathcal{B}_1(\pi(x^*), 0)$:*

(1) *$\{u = 0\}$ is a smooth hypersurface in $\mathbb{R}^n$ with curvatures bounded by C/R which stays in a C/R neighborhood of ∂B_R.*

(2) $$|U - G(x \cdot \nu - R, y)| \le \frac{C}{R}, \quad \nu := \frac{\pi(x^*)}{R}.$$

Proof. Assume for simplicity that x^* is on the positive x_n-axis and therefore $\pi(x^*) = Re_n$, $|x^* - Re_n| \le 1$. By Lemma 4.1 we have

$$U \le \Phi_R \le G\left(|x| - R + \frac{C}{R}, y\right) \le G\left(x_n - R + \frac{C'}{R}, y\right) =: V \quad \text{in } \mathcal{B}_3(Re_n).$$

Both U and V solve (4-1), and

$$(V - U)(x^*, 0) \le \frac{C''}{R}.$$

Since $V - U \ge 0$ satisfies

$$\Delta_a(V - U) = 0, \quad \partial_y^{1-a}(V - U) = b(x)(V - U),$$
$$b(x) := \int_0^1 W''(tu(x) + (1-t)v(x))\, dt,$$

we obtain

$$|V - U| \le \frac{C}{R} \quad \text{in } \mathcal{B}_{5/2}(Re_n)$$

from the Harnack inequality with Neumann condition for Δ_a. Moreover since b has bounded Lipschitz norm and $s > \frac{1}{2}$ we obtain $U - V \in C_x^{2,\alpha}$ for some $\alpha > 0$, and

$$\|U - V\|_{C_x^{2,\alpha}(\mathcal{B}_2(Re_n))} \le \frac{C}{R}$$

by local Schauder estimates. This easily implies the lemma. □

Remark 4.4. If instead of $\mathcal{B}_1((\pi(x^*), 0))$ we write the conclusion in $\mathcal{B}_K((\pi(x^*), 0))$ for some large, fixed constant K, then we need to replace C/R by $C(K)/R$. Here $C(K)$ represents a constant which depends also on K.

Next we obtain estimates near a point on $\{u = 0\}$ which admits a one-sided tangent ball of large radius R.

Lemma 4.5. *Assume that U is defined in $\mathcal{B}^+_{2R}$, satisfies (4-1), and that*

(a) $B_R(-Re_n) \subset \{u < 0\}$ *is tangent to* $\{u = 0\}$ *at* 0,

(b) *there is* $x_0 \in B_{R/2}(-Re_n)$ *such that* $u(x_0) \le -1 + c$ *for some* $c > 0$ *small.*

Then:

(1) $\{u = 0\}$ *is smooth in* B_1 *and has curvatures bounded by* C/R.

(2) $|U - G(x_n, y)| \le C/R$ *in* $\mathcal{B}_1$.

Proof. Assume first that $u < \phi_{R/8,z}$ for $z = -Re_n$.

We translate the graph of $\Phi_{R/8,z}$ by moving z continuously upward on the x_n axis. We stop when the translating graph becomes tangent by above to the graph of U for the first time. Denote by $(x^*, 0, u(x^*))$ the contact point and by z^* the final center z and by $\pi(x^*)$ the projection of x^* onto $\partial B_{R/8}(z^*)$.

By Lemma 4.3, $\{u = 0\}$ must be in a C_1/R neighborhood of $\partial B_{R/8}(z^*) \cap B_1(\pi(x^*))$ for some C_1 universal. This implies

$$z^* = te_n \quad \text{with } t \in \left[-\frac{R}{8} - \frac{C_1}{R}, -\frac{R}{8} + \frac{C_1}{R}\right].$$

Moreover, $\pi(x^*) \in B_{C_2}$ for some C_2 large universal, since otherwise $\pi(x^*)$ is at a distance greater than

$$\frac{1}{R}\frac{C_2^2}{8} > \frac{C_1}{R}$$

in the interior of the ball $B_R(-Re_n)$; hence $\{u = 0\}$ must intersect this ball and we reach a contradiction.

Now we apply Lemma 4.3 and Remark 4.4 at $\pi(x^*)$ and obtain the conclusion of the lemma.

It remains to show that $u < \phi_{R/8,-Re_n}$. By hypothesis (b) and the Harnack inequality we see that u is still sufficiently close to -1 in a whole ball $B_{R_0}(x_0)$ for some large universal R_0, and therefore $u < \phi_{R_0/2,x_0}$ provided that c is sufficiently small. Now we deform $\Phi_{R_0/2,x_0}$ by a continuous family of functions $\Phi_{r,z}$ and first we move z continuously from x_0 to $-Re_n$ and then we increase the radius r from R_0 to $\frac{1}{8}R$. By Lemma 4.3, the graphs of these functions cannot touch the graph of U by above and we obtain the desired inequality. With this the lemma is proved. □

In the next proposition we prove a localized version of Lemma 4.5.

Proposition 4.6. *Assume that U satisfies the equation in $\mathcal{B}_{R^{1-\sigma}}$ with σ small, universal as in Lemma 3.2, and*

(a) $B_R(-Re_n) \cap B_{R^{1/2-\sigma}} \subset \{u < 0\}$ *is tangent to* $\{u = 0\}$ *at* 0,

(b) *all balls of radius* $\frac{1}{4}R^{1-\sigma}$ *which are tangent by below to* $\partial B_R(-Re_n)$ *at some point in* $B_{R^{1/2-\sigma}}$ *are included in* $\{u < 0\}$,

(c) *there is* $x_0 \in B_{R^{1-\sigma}/4}\left(-\frac{1}{2}R^{1-\sigma}e_n\right)$ *such that* $u(x_0) \le -1 + c$.

Then in B_1 we have that $\{u = 0\}$ is smooth and has curvatures bounded by C/R.

Proof. As in Lemma 4.5, we slide the graph of $\Phi_{R/8,z}$ in the e_n-direction until it touches the graph of U, except that now we restrict only to the region

$$\mathcal{C}_R := \{|x'| \le \tfrac{1}{2}R^{1/2-\sigma}, |x_n| \le \tfrac{1}{2}R^{1-\sigma}, |y| \le \tfrac{1}{2}R^{1-\sigma}\}. \tag{4-2}$$

In order to repeat the argument above we need to show that the first contact point is an interior point and it occurs in $\mathcal{C}_{R/2}$. For this it suffices to prove that

$$U < \Phi_{R/8,z_0} \quad \text{in } \mathcal{C}_R \setminus \mathcal{C}_{R/2}, \qquad z_0 := \left(-\frac{R}{8} + \frac{C_1}{R}\right)e_n. \tag{4-3}$$

We estimate U by using the functions $\Psi_{R,z}$ given in Definition 4.2. Notice that Lemma 4.3 holds if we replace Φ_R by Ψ_R.

Now we slide the graphs $\Psi_{r,z}$, with $r := \frac{1}{4}R^{1-\sigma}$ and $|z'| \le R^{1/2-\sigma}$, $z_n = -2r$, upward in the e_n-direction. We use hypotheses (b), (c) and as in the proof of Lemma 4.5 we find $\Psi_{r,z} > U$ as long as $B_r(z)$ is at distance greater than Cr^{-1} from $\partial B_R(-Re_n)$. We obtain

$$U(x) < \Psi_r(d_1(x) + Cr^{-1}, y), \tag{4-4}$$

where $d_1(x)$ is the signed distance to $\partial B_R(-Re_n)$. From Remark 3.3 we have

$$\Psi_r(s, y) \le G_{R/8}\big(s + \big(\tfrac{1}{8}R\big)^{-3\sigma}, y\big).$$

We obtain

$$U(x, y) < G_{R/8}(d_1(x) + 2R^{-3\sigma}, y). \tag{4-5}$$

Let $d_2(x)$ represent the distance to $\partial B_{R/8}(z_0)$. Then in the region $\mathcal{C}_R \setminus \mathcal{C}_{R/2}$ we have either

(i) $|x'| \ge \frac{1}{2}\big(\frac{1}{2}R\big)^{1/2-\sigma}$ and then

$$d_2(x) - d_1(x) \ge -\frac{C_1}{R} + \frac{1}{R}|x'|^2 \ge 2R^{-3\sigma}, \tag{4-6}$$

or

(ii) $\min\{|x_n|, y\} \ge \frac{1}{8}R^{1-\sigma}$ and then both $(d_2(x), y)$ and $(d_1(x) + 2R^{-\sigma}, y)$ are outside $B^+_{1-\delta} \subset \mathbb{R}^2$; thus $G_{R/8}$ has the same value at these two points.

From (4-5) we find

$$U(x, y) < G_{R/8}(d_2(x), y) \quad \text{in } \mathcal{C}_R \setminus \mathcal{C}_{R/2}, \tag{4-7}$$

and (4-3) is proved. □

Next we consider the case in which the 0 level set of u is tangent by above at the origin to the graph of a quadratic polynomial.

Proposition 4.7. *Let U satisfy the equation in $\mathcal{B}_{R^{1-\sigma}}$ and hypothesis (c) of Proposition 4.6. Assume the surface*

$$\Gamma := \left\{x_n = \sum_1^{n-1} \frac{a_i}{2}x_i^2 + b' \cdot x'\right\} \cap B_{R^{1/2-\sigma}} \quad \textit{with } |b'| \le \varepsilon, \ \ |a_i| \le \varepsilon^{-2}R^{-1},$$

is tangent to $\{u = 0\}$ *at* 0 *for some small* ε *that satisfies* $\varepsilon \geq R^{-\sigma/2}$, *and assume further that all balls of radius* $\frac{1}{2}R^{1-\sigma}$ *which are tangent to* Γ *by below are included in* $\{u < 0\}$. *Then*

$$\sum_1^{n-1} a_i \leq CR^{-1}.$$

Proposition 4.7 states that the blowdown of $\{u = 0\}$ satisfies the minimal surface equation in some viscosity sense. Indeed, if we take $\varepsilon = R^{-\sigma/2}$, then the set $R^{\sigma-1}\{u = 0\}$ cannot be touched at 0 in an $R^{-1/2}$ neighborhood of the origin by a surface with curvatures bounded by $\frac{1}{2}$ and mean curvature greater than $CR^{-\sigma}$.

Proof. We argue as in the proof of Proposition 4.6 except that now we replace $\partial B_R(-Re_n)$ by Γ and $\partial B_{R/8}(z_0)$ by

$$\Gamma_2 := \left\{ x_n = \sum_1^{n-1} \frac{a_i}{2} x_i^2 + b' \cdot x' + \frac{C_1}{R} - \frac{1}{R}|x'|^2 \right\}.$$

We claim that

$$U(x, y) < G_{R/8}(d_2(x), y) \quad \text{in } \mathcal{C}_R \setminus \mathcal{C}_{R/2}, \tag{4-8}$$

where d_2 represents the signed distance to the Γ_2 surface and $\mathcal{C}_R$ is defined in (4-2). Using the surfaces $\Psi_{r,z}$ as comparison functions we obtain as in (4-4), (4-5) above that

$$U(x, y) < G_{R/8}(d_1(x) + C'r^{-1}, y) \quad \text{in } \mathcal{C}_R,$$

with $d_1(x)$ representing the signed distance to Γ. Notice that (4-6) is valid in our setting. Now we argue as in (4-7) and obtain the desired claim (4-8).

Next we show that $G_{R/8}(d_2(x), y)$ is a supersolution away from the set $\{|d_2| \leq 1, y = 0\}$ provided that

$$\sum_1^{n-1} a_i \geq MR^{-1}$$

for some M large, universal to be made precise later. The boundary inequality on $\{y = 0\}$ is clearly satisfied and on $\{y > 0\}$ we have

$$\Delta_a G_{R/8}(d_2(x), y) = \Delta_a G_{R/8}(s, y) + H(x)\, \partial_s G_{R/8}(s, y), \quad s := d_2(x), \tag{4-9}$$

where $H(x)$ represents the mean curvature at x of the parallel surface to Γ_2, and Δ_a on the right-hand side is with respect to the variables (s, y). If $|s| > R^{1-\delta}$ then $\partial_s G_{R/8} = 0$, and if $|s| \leq R^{1-\delta}$ we show below that $H < 0$, and in both cases we obtain $\Delta_a G_{R/8} \leq 0$.

Let κ_i, $i = 1, \ldots, n-1$, be the principal curvatures of Γ_2 at the projection of x onto Γ_2. Notice that at this point the slope of the tangent plane to Γ_2 is less than 4ε; hence we have

$$|\kappa_i| \leq 2\varepsilon^{-2}R^{-1} \leq 2R^{\sigma-1}, \quad \sum \kappa_i \leq -\sum a_i + C\varepsilon^2 \max|a_i| \leq -\tfrac{1}{2}MR^{-1}.$$

When $|d_2| \leq R^{1-\delta}$, we obtain $d_2\kappa_i = o(1)$, $d_2\kappa_i^2 = o(R^{-1})$ since $\sigma < \frac{\delta}{3}$; hence

$$H(x) = \sum \frac{\kappa_i}{1 - d_2\kappa_i} = \sum \left(\kappa_i + \frac{d_2\kappa_i^2}{1 - d_2\kappa_i} \right) \leq -\tfrac{1}{4}MR^{-1}. \tag{4-10}$$

Now we translate the graph of $G_{R/8}(d_2, y)$ along the e_n-direction until it touches the graph of U by above. Precisely, we consider the graphs of $G_R(d_2(x - te_n), y)$ with $t \le 0$ and start with t negative so that the function is identically 1 in $\mathcal{C}_R$. Then we increase t continuously until this graph becomes tangent by above to the graph of U in $\mathcal{C}_R$. Since $u(0) = 0$, a contact point must occur for some $t \le 0$ and, by (4-8), this point is interior to $\mathcal{C}_{R/2}$ and lies on $y = 0$. Let $(x^*, 0, u(x^*))$ be the first contact point where a translate $G_{R/8}(d_2(x - t^*e_n), y)$ touches U by above. We show that we reach a contradiction if M is chosen sufficiently large.

Define V as

$$V(x, y) := G\left(d_2(x - t^*e_n) + \frac{C}{R}, y\right) \ge G_{R/8}(d_2(x - t^*e_n), y) \ge U(x, y).$$

Notice that

$$\partial_y^{1-a} V = W'(V), \quad (V - U)(x^*, 0) \le \frac{C}{R}.$$

In $\mathcal{B}_1(x^*)$ we use the computation (4-9) above for V together with (4-10) and obtain

$$\Delta_a V \le -cMR^{-1} \quad \text{in } \mathcal{B}_1(x^*).$$

The function $Q := (V - U)/(cMR^{-1}) \ge 0$ satisfies in $\mathcal{B}_1(x^*)$

$$\Delta_a Q \le -1, \quad |\partial_y^{1-a} Q| \le CQ, \quad Q(x^*, 0) \le C'M^{-1}.$$

By the maximum principle

$$Q(x, y) \ge \mu^2 + \mu\, y^{1-a} - \frac{1}{2(n+1)}(|x - x^*|^2 + y^2)$$

for some μ small universal, and we reach a contradiction at $(x^*, 0)$ if M is sufficiently large. □

5. Harnack inequality

We use Proposition 4.6 to prove a Harnack-inequality property for flat level sets; see Theorem 5.1 below. The key step in the proof is to control the x_n-coordinate of the level set $\{u = 0\}$ in a set of large measure in the x'-variables.

Notation. We denote by $\mathcal{C}(l, \theta)$ the cylinder

$$\mathcal{C}(l, \theta) := \{|x'| \le l, |x_n| \le \theta\}.$$

Theorem 5.1 (Harnack inequality for minimizers). *Let U be a minimizer of J in $\mathcal{B}_q$ and assume that*

$$0 \in \{u = 0\} \cap \mathcal{C}(l, l) \subset \mathcal{C}(l, \theta),$$

and that all balls of radius $q := (l^2\theta^{-1})^{1-\sigma/2}$ which are tangent to $\mathcal{C}(l, \theta)$ by below and above are included in $\{u < 0\}$ and $\{u > 0\}$ respectively.

Given $\theta_0 > 0$ there exist $\omega > 0$ small depending on n, W, and $\varepsilon_0(\theta_0) > 0$ depending on n, W and θ_0 such that if

$$\theta l^{-1} \le \varepsilon_0(\theta_0), \quad \theta_0 \le \theta,$$

then

$$\{u=0\}\cap \mathcal{C}(\bar{l},\bar{l})\subset \mathcal{C}(\bar{l},\bar{\theta}),\quad \bar{l}:=\tfrac{l}{4},\ \bar{\theta}:=(1-\omega)\theta,$$

and all balls of radius $\bar{q}:=(\bar{l}^2\bar{\theta}^{-1})^{1-\sigma/2}$ *which are tangent to* $\mathcal{C}(\bar{l},\bar{\theta})$ *by below or above do not intersect* $\{u=0\}$.

The fact that u is a minimizer of J is only used in a final step of the proof. This hypothesis can be replaced by x_n-monotonicity for u, or more generally by the monotonicity of u in a given direction which is not perpendicular to e_n.

Definition 5.2. For a small $a>0$, we denote by $\mathcal{D}_a$ the set of points on

$$\{u=0\}\cap \mathcal{C}\left(\tfrac{3}{4}l,\theta\right)$$

which have a paraboloid of opening $-a$ and vertex $y=(y',y_n)$

$$P_{a,y}:=\left\{x_n=-\tfrac{a}{2}|x'-y'|^2+y_n\right\}$$

tangent by below in $\mathcal{C}(l,\theta)$, and with $P_{a,y}$ below the lateral boundary of $\mathcal{C}(l,\theta)$. In other words we allow only those polynomials $P_{a,y}$ which exit $\mathcal{C}(l,\theta)$ through the "bottom".

We denote by $D_a\subset\mathbb{R}^{n-1}$ the projection of $\mathcal{D}_a$ into $\mathbb{R}^{n-1}$ along the e_n-direction.

By Proposition 4.6 we see that as long as

$$l^{-1}\geq a\geq l^{-2-\eta}\quad\text{and}\quad l\geq C(\theta_0) \tag{5-1}$$

for some η small universal (depending on σ), $\{u=0\}$ has the following property (P):

(P) In a neighborhood of any point of $\mathcal{D}_a$, the set $\{u=0\}$ is a graph in the e_n-direction of a C^2 function with second derivatives bounded by Λa with Λ a universal constant.

Indeed, since $a\leq l^{-1}$, at a point $z\in\mathcal{D}_a$ the corresponding paraboloid at z has a tangent ball of radius

$$R:=ca^{-1}\leq l^{2+\eta}$$

by below. Since $|z'|\leq\frac{3}{4}l$ we see that $\{u=0\}\cap B_{l/4}(z)$ has a tangent ball $B_R(x_0)$ by below at z and hypothesis (a) of Proposition 4.6 holds since

$$\tfrac{l}{4}\geq R^{1/2-\sigma}.$$

The assumption that all balls of radius $q\geq c(\theta_0)l^{2-\sigma}\geq R^{1-\sigma}$ tangent by below to $C(l,\theta)$ are included in $\{u<0\}$ gives that all balls tangent to $\partial B_R(x_0)\cap B_{l/4}(z)$ by below are also included in $\{u<0\}$; hence hypothesis (b) of Proposition 4.6 holds.

Since u is a minimizer, in any sufficiently large ball in $\{u<0\}$ we have points that satisfy $u<-1+c$ and hypothesis (c) holds as well. In conclusion Proposition 4.6 applies and property (P) holds.

Since $\{u=0\}$ satisfies property (P), it satisfies a general version of the weak Harnack inequality which we proved in [Savin 2017]. In particular we are in the setting of Propositions 6.2 and 6.4 (see also Remark 6.7) in that paper.

This means that for any $\mu > 0$ small, there exists $M(\mu)$ depending on μ and universal constants such that if

$$\{u = 0\} \cap (B'_{l/2} \times [-\theta, (\omega - 1)\theta]) \neq \varnothing, \quad \omega := (32M)^{-1}, \tag{5-2}$$

then, by Proposition 6.2 in [Savin 2017], we obtain

$$\mathcal{H}^{n-1}(D_a \cap B'_{l/2}) \geq (1-\mu)\mathcal{H}^{n-1}(B'_{l/2}), \quad \text{with } a := M\,\omega\,\theta l^{-2}, \tag{5-3}$$

and

$$\mathcal{D}_a \cap \{|x'| \leq \tfrac{l}{2}\} \subset \{x_n \leq (8M\omega - 1)\theta\} = \{x_n \leq -\tfrac{3}{4}\theta\}. \tag{5-4}$$

We can apply that proposition since the interval I of allowed openings of the paraboloids satisfies, see (5-1),

$$I = [\omega\,\theta l^{-2}, M\omega\,\theta l^{-2}] \subset [l^{-2-\eta}, l^{-1}],$$

provided that $l \geq C(\mu, \theta_0)$ and $\varepsilon_0 \leq c$.

Next we let $\mathcal{D}^*_a$ denote the set of points on

$$\mathcal{D}^*_a := \{u = 0\} \cap \left(\{|x'| \leq \tfrac{l}{2}\} \times \left[-\tfrac{\theta}{2}, \theta\right]\right) \tag{5-5}$$

which admit a tangent paraboloid of opening a by above which exit $\mathcal{C}(l, \theta)$ through the "top". Also we denote by $D^*_a \subset \mathbb{R}^{n-1}$ the projection of $\mathcal{D}^*_a$ along e_n. Then according to Proposition 6.4 in [Savin 2017], (applied "upside down") we have

$$\mathcal{H}^{n-1}(D^*_{\tilde{a}} \cap B'_{l/2}) \geq \mu_0\,\mathcal{H}^{n-1}(B'_{l/2}), \quad \text{with } \tilde{a} = 8\theta l^{-2}, \tag{5-6}$$

for some μ_0 universal.

We choose μ in (5-2)–(5-4) universal as

$$\mu := \tfrac{1}{2}\mu_0.$$

According to (5-3), (5-6) this gives

$$\mathcal{H}^{n-1}(D_a \cap D^*_{\tilde{a}}) \geq \tfrac{1}{2}\mu_0 \mathcal{H}^{n-1}(B'_{l/2}). \tag{5-7}$$

Notice that by (5-4), (5-5) the sets $\mathcal{D}_a$ and $\mathcal{D}^*_{\tilde{a}}$ are disjoint.

At this point we would reach a contradiction to (5-2) if $\{u = 0\}$ were assumed to be a graph in the e_n-direction. Instead we use (5-7) and show that U cannot be a minimizer.

Proof of Theorem 5.1. It suffices to show that

$$\{u = 0\} \cap \mathcal{C}\left(\tfrac{l}{2}, \tfrac{l}{2}\right) \subset C\left(\tfrac{l}{2}, (1-\omega)\theta\right).$$

Then the existence of the balls of size $q \ll l^2\theta^{-1}$ (included in $\{u < 0\}$ and $\{u > 0\}$ respectively) tangent to $C\left(\frac{l}{4}, (1-\omega)\theta\right)$ follows easily as we restrict from the cylinder of size $\frac{l}{2}$ to the one of size $\frac{l}{4}$, and the conclusion is satisfied since $\tilde{q} \leq q$.

Assume by contradiction that (5-2) holds, and therefore (5-3), (5-7) hold as well. For each $x \in \mathcal{D}_a$ the set $\{u = 0\}$ has a tangent ball of radius $ca^{-1} \geq cl$ by below. Moreover, the normal to this ball at the

contact points in the e_n-direction makes a small angle which is bounded by $c\theta l^{-1} \le c\varepsilon_0$. According to Lemma 4.5 part (2) and Remark 4.4, we conclude that for any fixed constant K we have

$$\max_{(t,y)\in\mathcal{B}_K^+} |U(x', x_n + t, y) - G(t, y)| \le \rho, \tag{5-8}$$

with $\rho = \rho(K, \varepsilon_0) \to 0$ as $\varepsilon_0 \to 0$.

We denote the two-dimensional half disk of radius r in the (x_n, y)-variables centered at $z \in \mathbb{R}^n$ as

$$\mathcal{B}_{r,z}^+ := \{(z', z_n + t, y) : |(t, y)| \le r,\ y \ge 0\}.$$

From above we find for all $x \in \mathcal{D}_a$, or similarly if $x \in \mathcal{D}_a^*$, we have

$$J(U, \mathcal{B}_{K,x}^+) \ge \mathcal{J}(G, \mathcal{B}_K^+) - \bar\rho, \tag{5-9}$$

with $\bar\rho = \bar\rho(K, \varepsilon_0) \to 0$ as $\varepsilon_0 \to 0$.

If $x' \in D_a \cap D_{\tilde a}^*$ then by (5-4), (5-5) the two points $x^1 = (x', x_n^1) \in \mathcal{D}_a$ and $x^2 = (x', x_n^2) \in \mathcal{D}_{\tilde a}^*$ satisfy $x_n^2 - x_n^1 \ge \frac14\theta \ge \frac14\theta_0$. By (5-8) this means that the two disks $\mathcal{B}_{K,x^i}$ are disjoint provided that ρ is small; thus

$$\mathcal{J}(U, \mathcal{B}_{l/2,(x',0)}^+) \ge 2(\mathcal{J}(G, \mathcal{B}_K^+) - \bar\rho) \quad \text{if } x' \in D_a \cap D_{\tilde a}^*.$$

We integrate in x' and use also (5-3), (5-7), (5-9) to obtain

$$\mathcal{J}(U, A_{l/2}) \ge \big(1 + \tfrac12\mu_0\big)(\mathcal{J}(G, \mathcal{B}_K^+) - \bar\rho)\, \mathcal{H}^{n-1}(B'_{l/2}),$$

with

$$A_{l/2} := \mathcal{C}\big(\tfrac{l}{2}, \tfrac{l}{2}\big) \times \big[0, \tfrac{l}{2}\big].$$

We choose first K large and then ε_0 small such that $\bar\rho$ is sufficiently small so that

$$\mathcal{J}(U, A_{l/2}) \ge \big(1 + \tfrac14\mu_0\big)\, \mathcal{J}(G, \mathbb{R}^2_+)\, \mathcal{H}^{n-1}(B'_{l/2}).$$

This contradicts Lemma 5.3 below provided that ε_0 is taken sufficiently small. □

The next lemma is a Γ-convergence result and it is a consequence of the minimality of U in $A_{l/2}$.

Lemma 5.3.

$$\mathcal{J}(U, A_{l/2}) \le \mathcal{J}(G, \mathbb{R}^2_+)\, \mathcal{H}^{n-1}(B'_{l/2}) + \gamma(\varepsilon_0)\, l^{n-1}, \tag{5-10}$$

with $\gamma(\varepsilon_0) \to 0$ as $\varepsilon_0 \to 0$.

Proof. We interpolate between U and $V(x, y) := G(x_n, y)$ as

$$H = (1 - \varphi)U + \varphi V.$$

Here φ is a cutoff Lipschitz function such that $\varphi = 0$ outside $A_{l/2}$, $\varphi = 1$ in $\mathcal{R}$ and $|\nabla\varphi| \le 8/(1 + y)$ in $A_{l/2} \setminus \mathcal{R}$, where $\mathcal{R}$ is the cone

$$\mathcal{R} := \big\{(x, y) : \max\{|x'|, |x_n|\} \le \tfrac{l}{2} - 1 - 2y\big\}.$$

By the minimality of U we have

$$\mathcal{J}(U, A_{l/2}) \le \mathcal{J}(H, A_{l/2}) = \mathcal{J}(V, \mathcal{R}) + \mathcal{J}(H, A_{l/2} \setminus \mathcal{R}).$$

Since

$$\mathcal{J}(V, \mathcal{R}) \le \mathcal{J}(V, A_{l/2}) \le \mathcal{J}(G, \mathbb{R}^2_+)\, \mathcal{H}^{n-1}(B'_{l/2}),$$

we need to show that

$$\mathcal{J}(H, A_{l/2} \setminus \mathcal{R}) \le \gamma\, l^{n-1}, \tag{5-11}$$

with γ arbitrarily small. We have

$$\begin{aligned}&\mathcal{J}(H, A_{l/2} \setminus \mathcal{R})\\ &\le 4 \int_{A_{l/2}\setminus\mathcal{R}} \big(|\nabla\varphi|^2 (V-U)^2 + |\nabla(V-U)|^2\big) y^a\, dx\, dy + \int_D W(u) + W(v) + C(v-u)^2\, dx,\end{aligned} \tag{5-12}$$

with $D := \mathcal{C}\big(\frac{l}{2}, \frac{l}{2}\big) \setminus \mathcal{C}\big(\frac{l}{2} - 1, \frac{l}{2} - 1\big)$.

We use that $|U|, |V| \le 1$, $|\nabla U|, |\nabla V| \le C/(1+y)$ and we see that in (5-12) the first integral in the region where $y \ge C\gamma^{1/a}$ is bounded by

$$\int_{C\gamma^{1/a}}^{l/2} C_1 (1+y)^{-2} (1+y)\, y^a\, dy \le \frac{\gamma}{4}.$$

Next we notice that u and v are sufficiently close to each other in $\mathcal{C}\big(\frac{l}{2}, \frac{l}{2}\big)$ away from a thin strip around $x_n = 0$. Indeed, we can use barrier functions as in Proposition 4.6, see (4-4), and bound u by above and below in terms of the function $\psi_{l/2}$ and distance to the hyperplanes $x_n = \pm\theta$. This implies

$$W(u), W(v), |v-u| \le \gamma \quad \text{in } \mathcal{C}\big(\tfrac{l}{2}, \tfrac{l}{2}\big) \text{ if } |x_n| \ge C(\gamma) + \theta,$$

with $C(\gamma)$ large, depending on the universal constants and γ. For the extensions U and V, this gives

$$|V-U|, |\nabla(V-U)| \le C_2\gamma \quad \text{in } A_{l/2} \text{ if } |x_n| \ge C'(\gamma) + \theta \text{ and } y \le C\gamma^{1/a},$$

with C_2 universal. Now (5-11) easily follows from (5-12). □

6. Improvement of flatness

We state the improvement-of-flatness property of minimizers.

Theorem 6.1 (improvement of flatness). *Let U be a minimizer of J in $\mathcal{B}_q$ and assume that*

$$0 \in \{u = 0\} \cap \mathcal{C}(l, l) \subset \mathcal{C}(l, \theta),$$

and that all balls of radius $q := (l^2\theta^{-1})^{1-\sigma/2}$ which are tangent to $\mathcal{C}(l, \theta)$ by below and above are included in $\{u < 0\}$ and $\{u > 0\}$ respectively.

Given $\theta_0 > 0$ there exist $\eta > 0$ small depending on n, and $\varepsilon_1(\theta_0) > 0$ depending on n, W and θ_0 such that if

$$\theta l^{-1} \le \varepsilon_1(\theta_0), \quad \theta_0 \le \theta,$$

then

$$\{u=0\}\cap \mathcal{C}_\xi(\bar l,\bar l)\subset C_\xi(\bar l,\bar\theta),\quad \bar l:=\eta l,\ \bar\theta:=\eta^{3/2}\theta,$$

and all balls of radius $\bar q:=(\bar l^2\bar\theta^{-1})^{1-\sigma/2}$ *which are tangent to* $\mathcal{C}_\xi(\bar l,\bar\theta)$ *by below and above are included in* $\{u<0\}$ *and* $\{u>0\}$ *respectively.*

Here $\xi\in\mathbb{R}^n$ *is a unit vector and* $C_\xi(\bar l,\bar\theta)$ *represents the cylinder with axis* ξ, *base* $\bar l$ *and height* $\bar\theta$.

As a consequence of this flatness theorem we obtain our main theorem.

Theorem 6.2. *Let* U *be a global minimizer of* $\mathcal{J}$. *Suppose that the* 0 *level set* $\{u=0\}$ *is asymptotically flat at* ∞. *Then the* 0 *level set is a hyperplane and* u *is one-dimensional.*

Proof. Without loss of generality assume $u(0)=0$. Fix $\theta_0>0$, and $\varepsilon\le\varepsilon_1(\theta_0)$. We choose k sufficiently large such that, after increasing θ_k if necessary we have $\theta_k l_k^{-1}=\varepsilon$. We can apply Theorem 6.1 since $q=(l_k\varepsilon^{-1})^{1-\sigma/2}\ll l_k$, and we obtain that $\{u=0\}$ is trapped in a flatter cylinder. We apply Theorem 6.1 repeatedly until the height of the cylinder becomes less than θ_0. We conclude that $\{u=0\}$ is trapped in a cylinder with flatness less than ε and height θ_0. We let first $\varepsilon\to 0$ and then $\theta_0\to 0$ and obtain the desired conclusion. □

Proof of Theorem 6.1. The proof is by compactness and it follows from Theorem 5.1 and Proposition 4.7. Assume by contradiction that there exist U_k, θ_k, l_k, ξ_k such that u_k is a minimizer of J, $u_k(0)=0$, and the level set $\{u_k=0\}$ stays in the flat cylinder $\mathcal{C}(l_k,\theta_k)$ with $\theta_k\ge\theta_0$, $\theta_k l_k^{-1}\to 0$ as $k\to\infty$ for which the conclusion of Theorem 6.1 doesn't hold.

Let A_k be the rescaling of the 0 level sets given by

$$(x',x_n)\in\{u_k=0\}\mapsto(z',z_n)\in A_k,$$
$$z'=x'l_k^{-1},\quad z_n=x_n\theta_k^{-1}.$$

Claim 1. A_k has a subsequence that converges uniformly on $|z'|\le\frac12$ to a set $A_\infty=\{(z',w(z')),|z'|\le\frac12\}$, where w is a Holder continuous function. In other words, given ε, all but a finite number of the A_k's from the subsequence are in an ε neighborhood of A_∞.

Proof of Claim 1. Fix z_0', $|z_0'|\le\frac12$ and suppose $(z_0',z_k)\in A_k$. We apply Theorem 5.1 for the function u_k in the cylinder

$$\{|x'-l_kz_0'|<\tfrac12 l_k\}\times\{|x_n-\theta_kz_k|<2\theta_k\}$$

in which the set $\{u_k=0\}$ is trapped. Thus, there exists an increasing function $\varepsilon_0(\theta)>0$, $\varepsilon_0(\theta)\to 0$ as $\theta\to 0$, such that $\{u_k=0\}$ is trapped in the cylinder

$$\{|x'-l_kz_0'|<\tfrac18 l_k\}\times\{|x_n-\theta_kz_k|<2(1-\omega)\theta_k\}$$

provided that $4\theta_k l_k^{-1}\le\varepsilon_0(2\theta_k)$. Rescaling back we find that

$$A_k\cap\{|z'-z_0'|\le\tfrac18\}\subset\{|z_n-z_k|\le 2(1-\omega)\}.$$

We apply the Harnack inequality repeatedly and we find that

$$A_k \cap \{|z' - z_0'| \le 2^{-2m-1}\} \subset \{|z_n - z_k| \le 2(1-\omega)^m\} \tag{6-1}$$

provided that

$$\theta_k l_k^{-1} \le 4^{-m-1}\varepsilon_0(2(1-\omega)^m\theta_k).$$

Since these inequalities are satisfied for all k large we conclude that (6-1) holds for all but a finite number of k's. Now the claim follows from Arzelà–Ascoli theorem. □

Claim 2. The function w is harmonic (in the viscosity sense).

Proof of Claim 2. The proof is by contradiction. Fix a quadratic polynomial

$$z_n = P(z') = \tfrac{1}{2} z'^T M z' + \xi \cdot z', \quad \|M\| < \delta^{-1},\ |\xi| < \delta^{-1},$$

such that $\operatorname{tr} M > \delta$ and $P(z') + \delta|z'|^2$ touches the graph of w, say, at 0 for simplicity, and stays below w in $|z'| < 8\delta$ for some small δ. Notice that at all points in the cylinder $|z'| < 2\delta$, the quadratic polynomial above admits a tangent paraboloid by below of opening $-\delta^{-2}$ which is below $z_n = -2$ when $|z'| \ge 6\delta$.

Thus, for all k large we find points (z_k', z_{kn}) close to 0 such that $P(z') + \text{const}$ touches A_k by below at (z_k', z_{kn}) and stays below it in $|z' - z_k'| < \delta$.

This implies that, after eventually a translation, there exists a surface

$$\Gamma := \left\{ x_n = \frac{\theta_k}{l_k^2}\frac{1}{2} x'^T M x' + \frac{\theta_k}{l_k}\xi_k \cdot x' \right\}, \quad |\xi_k| < 2\delta^{-1},$$

that touches $\{u_k = 0\}$ at the origin and stays below it in $\mathcal{C}(\delta l_k, 2\theta_k)$. Moreover in the cylinder $\mathcal{C}\left(\frac{1}{2} l_k, 2\theta_k\right)$ the surface Γ admits at all points with $|x'| \le \delta l$ a tangent ball by below of radius $\delta^2 l_k^2 \theta_k^{-1} \gg q$. In view of our hypothesis we conclude that $\Gamma \cap B_{\delta l_k}$ admits at all its points a tangent ball of radius q by below which is included in $\{u < 0\}$.

We contradict Proposition 4.7 by choosing R as

$$R^{-1} := C^{-1}\,\delta\,\theta_k l_k^{-2},$$

with C the constant from Proposition 4.7 and with $\varepsilon = \delta^2$. Then for all large k we have

$$\theta_k l_k^{-1}|\xi_k| \le \varepsilon, \quad \theta_k l_k^{-2}\|M\| \le \varepsilon^{-2}R^{-1}, \quad \delta l_k \ge R^{1/2-\sigma}, \quad q \ge R^{1-\sigma},$$

and Proposition 4.7 applies. We obtain $\operatorname{tr} M \le \delta$ and we have reached a contradiction. □

Since w is harmonic, there exists $0 < \eta$ small depending only on n such that

$$|w - \xi \cdot z'| < \tfrac{1}{2}\eta^{3/2} \quad \text{for } |z'| < 2\eta\ ,$$

and the parabolas of opening $-C$ tangent by below (and above) to

$$z_n = \xi \cdot z' \pm \tfrac{1}{2}\eta^{3/2}$$

in the cylinder $|z'| < 2\eta$ lie below (or above) to the graph of w.

Rescaling back and using the fact that the A_k's converge uniformly to the graph of w and that $\bar{q} < q$ we easily conclude that u_k satisfies the conclusion of Theorem 6.1 for k large enough, and we have reached a contradiction. □

References

[Cabré and Cinti 2010] X. Cabré and E. Cinti, "Energy estimates and 1-D symmetry for nonlinear equations involving the half-Laplacian", *Discrete Contin. Dyn. Syst.* **28**:3 (2010), 1179–1206. MR Zbl

[Cabré and Cinti 2014] X. Cabré and E. Cinti, "Sharp energy estimates for nonlinear fractional diffusion equations", *Calc. Var. Partial Differential Equations* **49**:1-2 (2014), 233–269. MR Zbl

[Cabré and Sire 2014] X. Cabré and Y. Sire, "Nonlinear equations for fractional Laplacians, I: Regularity, maximum principles, and Hamiltonian estimates", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **31**:1 (2014), 23–53. MR Zbl

[Cabré and Sire 2015] X. Cabré and Y. Sire, "Nonlinear equations for fractional Laplacians, II: Existence, uniqueness, and qualitative properties of solutions", *Trans. Amer. Math. Soc.* **367**:2 (2015), 911–941. MR Zbl

[Cabré and Solà-Morales 2005] X. Cabré and J. Solà-Morales, "Layer solutions in a half-space for boundary reactions", *Comm. Pure Appl. Math.* **58**:12 (2005), 1678–1732. MR Zbl

[Caffarelli and Silvestre 2007] L. Caffarelli and L. Silvestre, "An extension problem related to the fractional Laplacian", *Comm. Partial Differential Equations* **32**:7-9 (2007), 1245–1260. MR Zbl

[Dipierro et al. 2016] S. Dipierro, J. Serra, and E. Valdinoci, "Improvement of flatness for nonlocal phase transitions", preprint, 2016. arXiv

[Dipierro et al. 2018] S. Dipierro, A. Farina, and E. Valdinoci, "A three-dimensional symmetry result for a phase transition equation in the genuinely nonlocal regime", *Calc. Var. Partial Differential Equations* **57**:1 (2018), art. id. 15. MR Zbl

[Palatucci et al. 2013] G. Palatucci, O. Savin, and E. Valdinoci, "Local and global minimizers for a variational energy involving a fractional norm", *Ann. Mat. Pura Appl.* (4) **192**:4 (2013), 673–718. MR Zbl

[Savin 2009] O. Savin, "Regularity of flat level sets in phase transitions", *Ann. of Math.* (2) **169**:1 (2009), 41–78. MR Zbl

[Savin 2017] O. Savin, "Some remarks on the classification of global solutions with asymptotically flat level sets", *Calc. Var. Partial Differential Equations* **56**:5 (2017), art. id. 141. MR Zbl

[Savin 2018] O. Savin, "Rigidity of minimizers in nonlocal phase transitions, II", preprint, 2018. arXiv

[Savin and Valdinoci 2012] O. Savin and E. Valdinoci, "Γ-convergence for nonlocal phase transitions", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **29**:4 (2012), 479–500. MR Zbl

[Savin and Valdinoci 2014] O. Savin and E. Valdinoci, "Density estimates for a variational model driven by the Gagliardo norm", *J. Math. Pures Appl.* (9) **101**:1 (2014), 1–26. MR Zbl

[Sire and Valdinoci 2009] Y. Sire and E. Valdinoci, "Fractional Laplacian phase transitions and boundary reactions: a geometric inequality and a symmetry result", *J. Funct. Anal.* **256**:6 (2009), 1842–1864. MR Zbl

Received 8 Dec 2016. Revised 9 Feb 2018. Accepted 9 Apr 2018.

OVIDIU SAVIN: savin@math.columbia.edu
Department of Mathematics, Columbia University, New York, NY, United States

ANALYSIS AND PDE
Vol. 11, No. 8, 2018
dx.doi.org/10.2140/apde.2018.11.1901

PROPAGATION AND RECOVERY OF SINGULARITIES IN THE INVERSE CONDUCTIVITY PROBLEM

ALLAN GREENLEAF, MATTI LASSAS, MATTEO SANTACESARIA, SAMULI SILTANEN AND GUNTHER UHLMANN

The ill-posedness of Calderón's inverse conductivity problem, responsible for the poor spatial resolution of electrical impedance tomography (EIT), has been an impetus for the development of hybrid imaging techniques, which compensate for this lack of resolution by coupling with a second type of physical wave, typically modeled by a hyperbolic PDE. We show in two dimensions how, using EIT data alone, to use propagation of singularities for complex principal-type PDEs to efficiently detect interior jumps and other singularities of the conductivity. Analysis of variants of the CGO solutions of Astala and Päivärinta (*Ann. Math.* (2) **163**:1 (2006), 265–299) allows us to exploit a complex principal-type geometry underlying the problem and show that the leading term in a Born series is an invertible nonlinear generalized Radon transform of the conductivity. The wave front set of all higher-order terms can be characterized, and, under a prior, some refined descriptions are possible. We present numerics to show that this approach is effective for detecting inclusions within inclusions.

1. Introduction

Electrical impedance tomography (EIT) aims to reconstruct the electric conductivity, σ, inside a body from active current and voltage measurements at the boundary. In many important applications of EIT, such as medical imaging [Assenheimer et al. 2001; Cheney et al. 1999; Isaacson et al. 2006] and geophysical

Greenleaf is partially supported by DMS-1362271 and a Simons Foundation Fellowship, Lassas and Siltanen are partially supported by Academy of Finland, Uhlmann is partially supported by a FiDiPro professorship.
MSC2010: 35R30, 58J40, 65N21.
Keywords: electrical impedance tomography, propagation of singularities, Calderón's problem, tomography, Radon transform.

prospecting, the primary interest is in detecting the location of interfaces between regions of inhomogeneous but relatively smooth conductivity. For example, the conductivity of bone is much lower than that of either skin or brain tissue, so there are jumps in conductivity of opposite signs as one transverses the skull.

In this paper we present a new approach in two dimensions to determining the singularities of a conductivity from EIT data. Analyzing the complex geometrical optics (CGO) solutions, originally introduced by Sylvester and Uhlmann [1987] and in the form required here by Astala and Päivärinta [2006a] and Huhtanen and Perämäki [2012], we transform the boundary values of the CGO solutions, which are determined by the Dirichlet-to-Neumann map [Astala and Päivärinta 2006b], in such a way as to extract the leading singularities of the conductivity, σ.

We show that the leading term of a Born series derived from the boundary data is a nonlinear Radon transform of σ and allows for good reconstruction of the singularities of σ, with the higher-order terms representing multiple scattering. Although one cannot escape the exponential ill-posedness inherent in EIT, the well-posedness of Radon inversion results in a robust method for detecting the leading singularities of σ. In particular, one is able to detect inclusions within inclusions (i.e., nested inclusions) within an unknown inhomogeneous background conductivity; this has been a challenge for other EIT methods. This property is crucial for one of the main applications motivating this study, namely using EIT for classifying strokes as ischemic (caused by an embolism preventing blood flow to part of the brain) or hemorrhagic (caused by bleeding in the brain); see [Holder 1992a; 1992b; Malone et al. 2014].

Our algorithm consists of two steps, the first of which is the reconstruction of the boundary values of the CGO solutions, and this is known to be exponentially ill-posed, i.e., satisfy only logarithmic stability estimates [Knudsen et al. 2009]. The second step begins with a separation of variables and partial Fourier transform in the radial component of the spectral variable. Thus, one instability of our algorithm arises from the exponential instability of the reconstruction of the CGO solutions from the Dirichlet-to-Neumann map. Another instability arises from low-pass filtering in Fourier inversion (similar to those of regularization methods used for CT and other linear inverse problems), and (presumably) the multiple scattering terms in the Born series we work with, which we only control rigorously for low orders and under some prior. Nevertheless, based on both the microlocal analysis and numerical simulations we present, the method appears to allow for robust detection of singularities of σ, in particular the location and signs of jumps. See Section 1A for further discussion of the ill-posedness issues raised by this method.

EIT can be modeled mathematically using the inverse conductivity problem of [Calderón 1980]. Consider a bounded, simply connected domain $\Omega \subset \mathbb{R}^n$ with smooth boundary and a scalar conductivity coefficient $\sigma \in L^\infty(\Omega)$ satisfying $\sigma(x) \geq c > 0$ almost everywhere. Applying a voltage distribution f at the boundary leads to the elliptic boundary-value problem

$$\nabla \cdot \sigma \nabla u = 0 \quad \text{in } \Omega, \qquad u|_{\partial\Omega} = f. \tag{1-1}$$

Infinite-precision boundary measurements are then modeled by the Dirichlet-to-Neumann map

$$\Lambda_\sigma : f \mapsto \sigma \frac{\partial u}{\partial \vec{n}}\bigg|_{\partial\Omega}, \tag{1-2}$$

where $\vec{n}$ is the outward normal vector of $\partial\Omega$.

Astala and Päivärinta [2006b] transformed the construction of the CGO solutions in two dimensions by reducing the conductivity equation to a Beltrami equation. Identify $\mathbb{R}^2$ with $\mathbb{C}$ by setting $z = x_1 + ix_2$ and define the Beltrami coefficient

$$\mu(z) = \frac{1-\sigma(z)}{1+\sigma(z)}.$$

Since $c_1 \le \sigma(z) \le c_2$, we have $|\mu(z)| \le 1-\epsilon$ for some $\epsilon > 0$. Further, if we assume $\sigma \equiv 1$ outside some $\Omega_0 \Subset \Omega$, then $\operatorname{supp}(\mu) \subset \overline{\Omega}_0$. Now consider the unique solution of

$$\bar{\partial}_z f_\pm(z,k) = \pm\mu(z)\overline{\partial_z f_\pm(z,k)}, \qquad e^{-ikz} f_\pm(z,k) = 1 + \omega^\pm(z,k), \tag{1-3}$$

where $ikz = ik(x_1 + ix_2)$ and $\omega^\pm(z,k) = \mathcal{O}(1/|z|)$ as $|z| \to \infty$. Here z is considered as a spatial variable and $k \in \mathbb{C}$ as a spectral parameter. We note that $u = \operatorname{Re} f_+$ satisfies (1-1), and denote $\omega^\pm$ by $\omega^\pm_\mu$ when emphasizing dependence on the Beltrami coefficient μ. Recently, this technique has been generalized also for conductivities that are not in $L^\infty(\Omega)$ but only exponentially integrable [Astala et al. 2016].

The two crucial ideas of the current work are:

(i) To analyze the scattering series, we use the modified construction of Beltrami-CGO solutions of [Huhtanen and Perämäki 2012], which only involves exponentials of modulus 1 and where the solutions are constructed as a limit of an iteration of linear operations. This differs from the original construction of [Astala and Päivärinta 2006b], where the construction of the exponentially growing solutions is based on the Fredholm theorem.

(ii) To transform the CGO solutions, we introduce polar coordinates in the spectral parameter k, followed by a partial Fourier transform in the radial direction.

These ideas are used as follows: Formally one can view the Beltrami equation (1-3) as a scattering equation, where μ is considered as a compactly supported scatterer and the "incident field" is the constant function 1. Using (i), we write the CGO solutions $\omega^\pm$ as a "scattering series",

$$\omega^\pm(z,k) \sim \sum_{n=1}^{\infty} \omega_n^\pm(z,k), \tag{1-4}$$

considered as a formal power series (see Theorem 1.1)

Using (ii), we decompose $k = \tau e^{i\varphi}$ and then, for each n, form the partial Fourier transform of the n-th order scattering term from (1-4) in τ, denoting these by

$$\hat{\omega}_n^\pm(z,t,e^{i\varphi}) := \mathcal{F}_{\tau\to t}(\omega_n^\pm(z,\tau e^{i\varphi})). \tag{1-5}$$

As is shown in Section 5B, singularities in σ can be detected from averaged versions of $\hat{\omega}_1^\pm$, denoted by $\hat{\omega}_1^{a,\pm}$, formed by taking a complex contour integral of $\hat{\omega}_1^\pm(z,t,e^{i\varphi})$ over $z \in \partial\Omega$; see Figure 1.

Recall that the traces of CGO solutions $\omega^\pm$ can be recovered perfectly from infinite-precision data Λ_σ [Astala and Päivärinta 2006a; 2006b]. When σ is close to 1, the single-scattering term $\omega_1^\pm$ is close to $\omega^\pm$. Figure 1 suggests that what we can recover resembles parallel-beam X-ray projection data of the singularities of μ. Indeed, we derive approximate reconstruction formulae for μ (thus mildly nonlinear in σ), analogous to the classical filtered back-projection method of X-ray tomography.

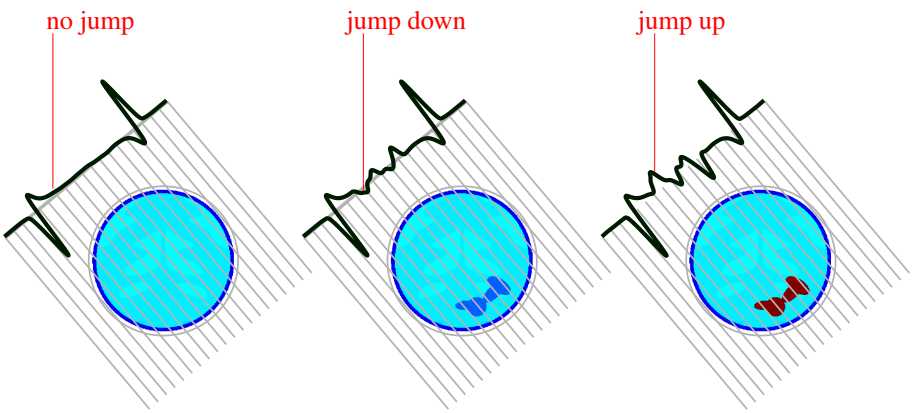

Figure 1. The method provides information about inclusions within inclusions in an unknown inhomogeneous background. Jump singularities in the conductivity show up in the function values much like in parallel-beam X-ray tomography: recording integrals along parallel lines over the coefficient function. This is illustrated using stroke-like computational phantoms. Left: Intact brain. Dark blue ring, with low conductivity, models the skull. Middle: Ischemic stroke, or blood clot preventing blood flow to the dark blue area. The conductivity in the affected area is less than that of the background. Right: Hemorrhagic stroke, or bleeding in the brain. The conductivity in the affected area is greater than the background. The function shown is $T^{a,+}\mu(t/2, e^{i\varphi}) - T^{a,-}\mu(t/2, e^{i\varphi})$, and φ indicates a direction perpendicular to the virtual "X-rays".

The wave front sets of all of the terms $\hat{\omega}_n^{\pm}$ are analyzed in Theorem 7.2. More detailed descriptions of the initial three terms, $\hat{\omega}_1^{\pm}$, $\hat{\omega}_2^{\pm}$ and $\hat{\omega}_3^{\pm}$, identifying the latter two as sums of paired Lagrangian distributions under a prior on the conductivity, are given in Section 5A, 6 and 9, respectively.

Let $X = \{\mu \in L^\infty(\Omega) : \text{ess supp}(\mu) \subset \Omega_0,\ \|\mu\|_{L^\infty(\Omega)} \leq 1 - \epsilon\}$, recalling that $\Omega_0 \Subset \Omega$. The expansion in (1-4) comes from the following:

Theorem 1.1. *For $k \in \mathbb{C}$, define nonlinear operators $W^{\pm}(\,\cdot\,; k) : X \to L^2(\Omega)$ by*

$$W^{\pm}(\mu; k)(z) := \omega_{\mu}^{\pm}(z, k).$$

Then, at any $\mu_0 \in X$, we know $W^{\pm}(\,\cdot\,; k)$ has Fréchet derivatives in μ of all orders $n \in \mathbb{N}$, denoted by $D^n W_k|_{\mu_0}$, and the multiple scattering terms in (1-4) are given by

$$\omega_n^{\pm} = [D^n W_k^{\pm}(\mu, \mu, \ldots, \mu)]|_{\mu=0}. \tag{1-6}$$

The n-th order scattering operators,

$$T_n^{\pm} : \mu \mapsto \hat{\omega}_n^{\pm} := \mathcal{F}_{\tau \to t}(\omega_n^{\pm}(z, \tau e^{i\varphi})), \quad z \in \partial\Omega,\ t \in \mathbb{R},\ e^{i\varphi} \in \mathbb{S}^1, \tag{1-7}$$

which are homogeneous forms of degree n in μ, have associated multilinear operators whose Schwartz kernels K_n have wave front relations which can be explicitly computed. See formulas (5-6) and (5-7) for

the case $n = 1$ and (4-14) for $n \geq 2$. K_1 is a Fourier integral distribution; K_2 is a generalized Fourier integral (or paired Lagrangian) distribution; and for $n \geq 3$, K_n has wave front set contained in a union of a family of 2^{n-1} pairwise cleanly intersecting Lagrangians.

Singularity propagation for the first-order scattering $\hat{\omega}_1^{\pm}$ is described by a Radon-type transform and a filtered back-projection formula; see [Kuchment 2014].

Theorem 1.2. *Define averaged operators $T_n^{a,\pm}$ for $n \in \mathbb{N}$ and $T^{a,\pm}$ by the complex contour integrals*[1]

$$T_n^{a,\pm}\mu(t, e^{i\varphi}) = \frac{1}{2\pi i}\int_{\partial\Omega} \hat{\omega}_n^{\pm}(z, t, e^{i\varphi})\, dz, \tag{1-8}$$

$$T^{a,\pm}\mu(t, e^{i\varphi}) = \frac{1}{2\pi i}\int_{\partial\Omega} \hat{\omega}^{\pm}(z, t, e^{i\varphi})\, dz, \tag{1-9}$$

with $\omega_n^{\pm}$ defined via formulas (1-6)–(1-7) and $\omega^{\pm}$ defined via (1-3). Then we have

$$(-\Delta)^{-1/2}(T_1^{a,\pm})^* T_1^{a,\pm}\mu = \mu. \tag{1-10}$$

Theorem 1.2 suggests an approximate reconstruction algorithm:

- Given Λ_σ, follow [Astala et al. 2011, Section 4.1] to compute both $\omega^+(z, k)$ and $\omega^-(z, k)$ for $z \in \partial\Omega$ by solving the boundary integral equation derived in [Astala and Päivärinta 2006a].
- Introduce polar coordinates in the spectral variable k and compute the partial Fourier transform, $\hat{\omega}^{\pm}(z, t, e^{i\varphi})$.
- Using the operator $T^{a,\pm}$ defined in (1-9), we compute $\tilde{\mu}^+ := \Delta^{-1/2}(T_1^{a,+})^* T^{a,+}\mu$ and $\tilde{\mu}^- := \Delta^{-1/2}(T_1^{a,-})^* T^{a,-}\mu$. Note the difference with (1-10).
- Approximately reconstruct by $\sigma = (\mu - 1)/(\mu + 1) \approx (\tilde{\mu} - 1)/(\tilde{\mu} + 1)$, where $\tilde{\mu} = (\tilde{\mu}^+ - \tilde{\mu}^-)/2$. The approximation comes from using $T^{a,\pm}\mu$ instead of $T_1^{a,\pm}\mu$ in the previous step.

See the middle column of Figure 2 for an example.

One can also use the identity $(T_1^{a,\pm})^* T_1^{a,\pm} = (-\Delta)^{1/2}$ to enhance the singularities in the reconstruction. This is analogous to Λ-tomography in the context of linear X-ray tomography [Faridani et al. 1992; 1997]. See the right-most column in Figure 2 for reconstructions using the operator $(T_1^{a,\pm})^* T^{a,\pm}$.

Our general theorem on singularity propagation is quite technical, and so we illustrate it here using a simple example, postponing the precise statement and proof to Section 7 below.

Assume that the conductivity is of the form $\sigma(z) = \sigma(|z|)$ and smooth except for a jump across the circle $|z| = \rho$. One can describe the singular supports of the $\hat{\omega}_n^{\pm}(z, t, e^{i\varphi})$. For $m \in \mathbb{N}$, define hypersurfaces

$$\Pi_m = \{(z, t, e^{i\varphi}) \in \mathbb{C} \times \mathbb{R} \times \mathbb{S}^1 : \ t = 2\rho m\}.$$

Using the analysis later in the paper, one can see that

$$\big(\text{sing supp}(\hat{\omega}_n^{\pm}) \cap \{(z, t, e^{i\varphi});\ |z| \geq 1\}\big) \subset \bigcup \{\Pi_m : -n \leq m \leq n,\ m \equiv n \mod 2\}.$$

[1] Throughout, dz will denote the element of complex contour integration along a curve, while $d^1\boldsymbol{x}$ is arc length measure. We denote by d^2z two-dimensional Lebesgue measure in $\mathbb{C}$.

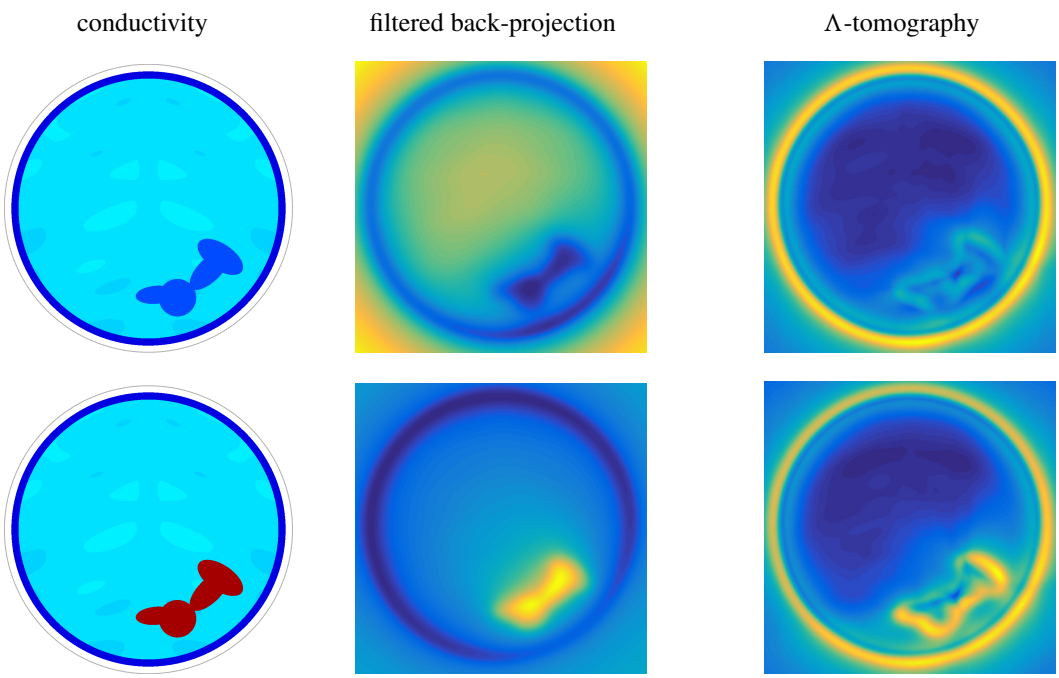

Figure 2. Reconstructions, of computational phantoms modeling ischemic strokes (top row) and hemorrhagic strokes (bottom row), from very high precision simulated EIT data. The results are promising for portable, cost-effective classification of strokes without use of ionizing radiation.

However, it turns out that, by a parity symmetry property described in Section 8 , subtracting $\hat{\omega}^-$ from $\hat{\omega}^+$ eliminates the even terms, $\hat{\omega}^{\pm}_{2n}$, so that their singularities, including a strong one for $\hat{\omega}^{\pm}_2$ at $t = 0$, do not create artifacts in the imaging. See Figure 3 for a diagram of singularity propagation in the case $\rho = 0.2$.

1A. ***Ill-posedness, noise and deconvolution.*** The exponential ill-posedness of the Calderón inverse problem (i.e., it satisfies a stability estimate of only logarithmic type) has important consequences for EIT with realistic data. Calderón inverse problems for elliptic equations were shown to be exponentially ill-posed in [Mandache 2001]. Corresponding to this, in [Knudsen et al. 2009, Lemma 2.4] it was shown that when the Dirichlet-to-Neumann map is given with error ϵ, the boundary values of the CGO solutions, or equivalently, $\omega(z, k)|_{z\in\partial\Omega}$, can be found with accuracy ϵ only for the frequencies $|k| \leq R_\epsilon = c \log(\epsilon^{-1})$.

This exponential instability holds even under the prior that conductivities consist of inclusions [Alessandrini and Di Cristo 2005]. Furthermore, inclusions need to have a minimum size to be detectable [Alessandrini 1988; Isaacson 1986; Cheney and Isaacson 1992], and in order to appear in reconstructions, the deeper inclusions are inside an object, the larger they must be [Nagayasu et al. 2009; Alessandrini and Scapin 2017; Garde and Knudsen 2017]. Finally, the resolution of reconstructions is limited by

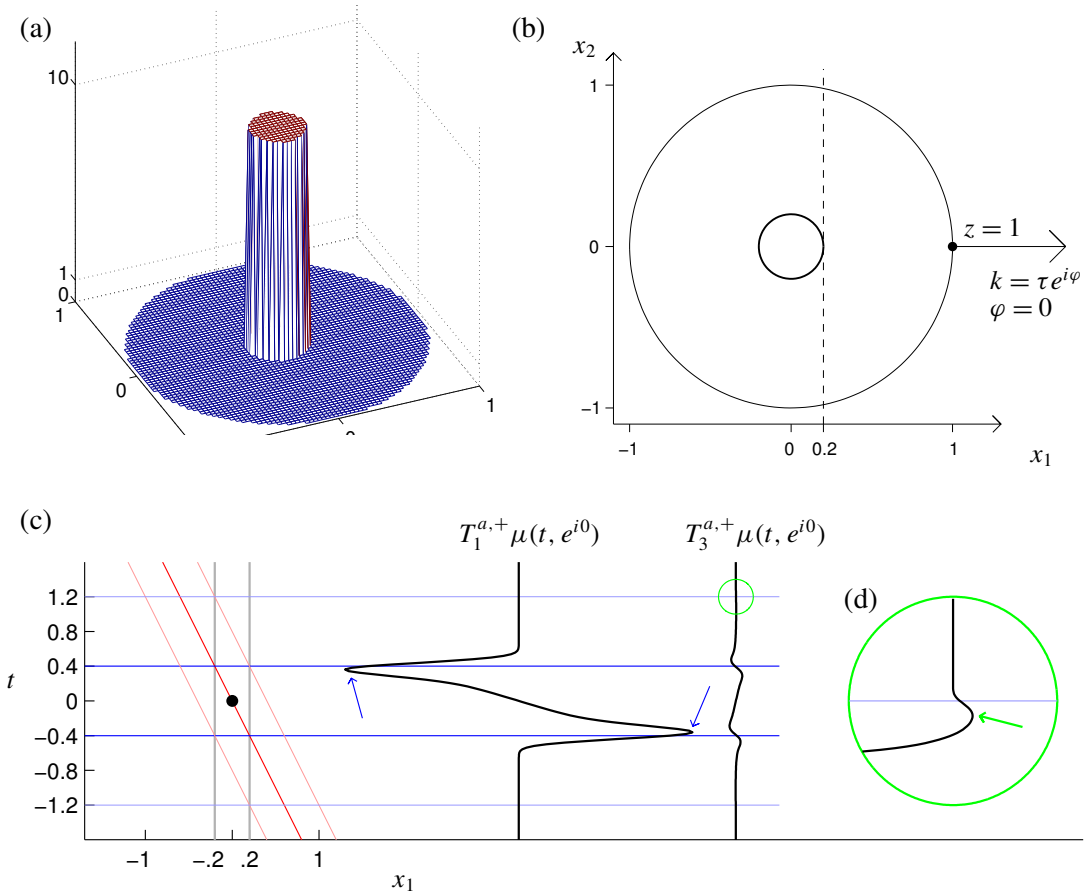

Figure 3. (a) Three-dimensional plot of the conductivity having a jump along the circle with radius $\rho = 0.2$ and center at the origin. (b) Unit disc and singular support of the conductivity in the z-plane, where $z = x_1 + ix_2$. (c) The term $T_1^{a,+}\mu(t, e^{i0})$ has peaks, indicated by blue arrows, at $t = \pm 2\rho$ corresponding to the locations of the main singularities in μ, as expected by Theorem 1.2. The higher-order term $T_3^{a,+}\mu(t, e^{i0})$, smaller than $T_1^{a,+}\mu(t, e^{i0})$ in amplitude, exhibits singularities caused by reflections at both $t = \pm 2\rho$ and $t = \pm 6\rho$. (d) The singularities of the term $T_3^{a,+}\mu(t, e^{i0})$ at $t = \pm 6\rho$ are very small. Shown is a zoom-in near $t = 6\rho$, with amplitude increased by a factor of 70.

noisy data. It is natural to ask how these limitations are reflected in the approach described in this paper.

Our results show that the part of the conductivity's wave front set in the direction specified by φ is seen as specific singularities in the function $\hat{\omega}^{\pm}(z, \cdot, e^{i\varphi})$, defined in (1-5). However, due to algebraic decay of the principal symbol of a Fourier integral operator, the amplitude of the measured singularity is bounded by $C\,\mathrm{dist}(\partial\Omega, z)^{-1}$, making it harder to recover details deep inside the imaging domain.

Furthermore, with realistic and noisy data, we can compute $\omega^{\pm}(z, k)$ only in a disc $|k| \le k_{\max}$ with a measurement apparatus and noise-dependent radius $k_{\max} > 0$; see [Knudsen et al. 2009; Astala et al.

2011; 2014]. With smaller noise we can take a larger $k_{\max}$, whereas large noise forces $k_{\max}$ to be small. This makes it more difficult to locate singularities precisely.

To better understand the difficulty, consider the truncated Fourier transform:

$$\int_{-k_{\max}}^{k_{\max}} e^{-it\tau}\omega^{\pm}(z,\tau e^{i\varphi})\,d\tau = \int_{-\infty}^{\infty} e^{-it\tau}\omega^{\pm}(z,\tau e^{i\varphi})\chi_{k_{\max}}(\tau)\,d\tau, \tag{1-11}$$

where $\chi_{k_{\max}}(\tau)$ is the characteristic function of the interval $[-k_{\max}, k_{\max}]$. Note that

$$\hat{\chi}_{k_{\max}}(t) = C\,\frac{\sin(k_{\max}\,t)}{t} \tag{1-12}$$

with a constant $C \in \mathbb{R}$. Noise forces us to replace the Fourier transform in (1-5) by a truncated integral such as (1-11). Therefore, we need to apply one-dimensional deconvolution in t to recover $\hat{\omega}^{\pm}(z,\cdot\,,e^{i\varphi})$ approximately from $\hat{\omega}^{\pm}(z,\cdot\,,e^{i\varphi}) * \hat{\chi}_{k_{\max}}$. Higher noise level means a smaller $k_{\max}$, which by (1-12) leads to a wider blurring kernel $\hat{\chi}_{k_{\max}}$; due to the Nyquist–Shannon sampling theorem, this results in a more ill-posed deconvolution problem and thus limits the imaging resolution.

In practice it is better to use a smooth windowing function instead of the characteristic function for reducing unwanted oscillations (Gibbs phenomenon), and there are many suitable deconvolution algorithms in the literature [Chen et al. 2001; Candès and Fernandez-Granda 2013; 2014].

It is also natural to ask how the method introduced here compares to previous work in terms of detecting inclusions and jumps.

Many methods have been proposed for regularized edge detection from EIT data. Examples include the *enclosure method* [Ikehata 2000; Ikehata and Siltanen 2000; 2004; Brühl and Hanke 2000; Ide et al. 2007; Uhlmann and Wang 2008], the *factorization method* [Kirsch 1998; Brühl and Hanke 2000; Lechleiter 2006; Lechleiter et al. 2008], the *monotonicity method* [Harrach and Ullrich 2013; 2015]. These methods can only detect the outer boundary of an inclusion in conductivity, whereas the method described here, which exploits the propagation of singularities for complex principal-type operators, can see nested jump curves. Also, the proposed method can deal with inclusions within inclusions, and with conductivities having both positive and negative jumps, even in unknown inhomogeneous smooth background.

One can also attempt edge detection based on EIT algorithms originally designed for reconstructing the full conductivity distribution. There are two main approaches: sharpening blurred EIT images in data-driven postprocessing [Hamilton et al. 2014; 2016], and applying sparsity-promoting inversion methods such as total variation regularization [Dobson and Santosa 1994; Kaipio et al. 2000; Rondi and Santosa 2001; Chan and Tai 2004; Chung et al. 2005; Tanushev and Vese 2007; van den Doel and Ascher 2006; Jin and Maass 2012; Garde and Knudsen 2016; Zhou et al. 2015]. As of now, the former approach does not have rigorous analysis available. Some of the latter kinds of approaches are theoretically capable of detecting nested inclusions; however, in variational regularization there is typically an instability issue, where a large low-contrast inclusion may be represented by a smaller high-contrast feature in the reconstruction. Numerical evidence suggests that method introduced here can accurately and robustly reconstruct jumps, both in terms of location and sign.

2. Complex principal-type structure of CGO solutions

We start by describing the microlocal geometry underlying the exponentially growing, or so-called *complex geometrical optics* (*CGO*), solutions to the conductivity equation on $\mathbb{R}^d$, $d \geq 2$,

$$\nabla \cdot \sigma \nabla u(x) = 0, \quad x \in \mathbb{R}^n, \tag{2-1}$$

originating in [Sylvester and Uhlmann 1987]. For complex frequencies $\zeta = \zeta_R + i\zeta_I \in \mathbb{C}^n$ with $\zeta \cdot \zeta = 0$, one can decompose ζ as $\zeta = \tau\eta$, with $\tau \in \mathbb{R}$ and $\eta = \eta_R + i\eta_I$, $|\eta_R| = |\eta_I| = 1$, $\eta_R \cdot \eta_I = 0$. Now consider solutions to (2-1) of the form

$$u(x) := e^{i\zeta\cdot x} w(x,\tau) = e^{i\tau\eta\cdot x} w(x,\tau).$$

Physically speaking, τ can be considered as a spatial frequency, with the voltage on the boundary $\partial\Omega$ oscillating at length scale τ^{-1}.

The conductivity equation (2-1) becomes

$$\begin{aligned} 0 &= \frac{1}{\sigma}\nabla\cdot\sigma\nabla u(x) = \frac{1}{\sigma}\nabla\cdot\sigma\nabla(e^{i\tau\eta\cdot x} w(x,\tau)) \\ &= \left(\Delta + \left(\frac{1}{\sigma}\nabla\sigma\right)\cdot\nabla\right)(e^{i\tau\eta\cdot x} w(x,\tau)) \\ &= \left(\Delta w(x,\tau) + 2i\tau\eta\cdot\nabla w(x,\tau) + \left(\frac{1}{\sigma}\nabla\sigma\right)\cdot(\nabla + i\tau\eta) w(x,\tau)\right) e^{i\tau\eta\cdot x}. \end{aligned}$$

Hence, we have

$$\Delta w(x,\tau) + 2i\tau\eta\cdot\nabla w(x,\tau) + \left(\frac{1}{\sigma}\nabla\sigma\right)\cdot(\nabla + i\tau\eta) w(x,\tau) = 0.$$

Taking the partial Fourier transform $\hat{w}$ in the τ-variable and denoting the resulting dual variable by t, which can be thought of as a "pseudo-time", one obtains

$$\Delta\hat{w}(x,t) - 2\eta\frac{\partial}{\partial t}\cdot\nabla\hat{w}(x,t) + \left(\frac{1}{\sigma}\nabla\sigma\right)\cdot\left(\nabla - \eta\frac{\partial}{\partial t}\right)\hat{w}(x,t) = 0.$$

The principal part of this equation is given by the operator

$$\widetilde{\Box} = \mathcal{P}_R + i\mathcal{P}_I = \Delta - 2\eta\frac{\partial}{\partial t}\cdot\nabla,$$

where

$$\mathcal{P}_R = \Delta - 2\eta_R\frac{\partial}{\partial t}\cdot\nabla \quad \text{and} \quad \mathcal{P}_I = -2\eta_I\frac{\partial}{\partial t}\cdot\nabla.$$

With ξ the variable dual to x, the full symbols of $\mathcal{P}_R$ and $\mathcal{P}_I$ are

$$p_R(x,t,\xi,\tau) = -\xi^2 + 2\tau\eta_R\cdot\xi, \quad p_I(x,t,\xi,\tau) = 2\tau\eta_I\cdot\xi,$$

and these commute in the sense of Poisson brackets: $\{p_R, p_I\} = 0$. Furthermore, on the characteristic variety

$$\begin{aligned} \Sigma &:= \{(x,t,\xi,\tau) \in \mathbb{R}^{d+1}\times(\mathbb{R}^{d+1}\setminus\{0\}) : p_R(x,t,\xi,\tau) = 0,\ p_I(x,t,\xi,\tau) = 0\} \\ &= \{(x,t,\xi,\tau) \in \mathbb{R}^{d+1}\times(\mathbb{R}^{d+1}\setminus\{0\}) : |\xi|^2 - 2\tau\eta_R\cdot\xi = 0,\ 2\tau\eta_I\cdot\xi = 0\} \\ &= \{(x,t,\xi,\tau) \in \mathbb{R}^2\times\mathbb{R}\times\mathbb{R}^2\times(\mathbb{R}\setminus\{0\}) : \xi = 2\tau\eta_R \text{ or } \xi = 0\}, \end{aligned}$$

the gradients $dp_R = (-2\xi + 2\tau\eta_R, 2\eta_R \cdot \xi)$ and $dp_I = (2\tau\eta_I, 2\eta_I \cdot \xi)$ are linearly independent. Finally, no bicharacteristic leaf (see below) is trapped over a compact set. Thus, $\widetilde{\Box} = \mathcal{P}_R + i\mathcal{P}_I$ is a *complex principal-type operator* in the sense of [Duistermaat and Hörmander 1972].

Recall that for a *real* principal-type operator, such as $\partial/\partial x_1$ in $\mathbb{R}^m$, $m \geq 2$, or the d'Alembertian wave operator, the singularities propagate along curves (the characteristics); for instance, for the wave equation, singularities propagate along light rays. *Complex* principal-type operators, such as $\partial_{x_1} + i\partial_{x_2}$ in $\mathbb{R}^m$, $m \geq 3$, or the operator $\widetilde{\Box}$ above, also propagate singularities, but now along *two*-dimensional surfaces, called leaves, which are the spatial projections of the bicharacteristic surfaces formed by the joint flowout of H_{p_R}, H_{p_I}. For the operator $\widetilde{\Box}$ above, this roughly means that if $\widetilde{\Box}\hat{w}(x,t) = \hat{f}(x,t)$ and $(x_0, t_0, \xi_0, \tau_0) \in \Sigma$ is in the wave front set of $\hat{f}(x,t)$, then the wave front set of $\hat{w}(x,t)$ contains a plane through this point. See [Duistermaat and Hörmander 1972, Section 7.2] for detailed statements.

In the situation relevant for this paper, the x-projection of any bicharacteristic leaf is all of $\mathbb{R}^2$ and thus reaches all points of $\overline{\Omega}$. Thus, complete information about σ in the interior is accessible to boundary measurements made at *any* point on $\partial\Omega$. We will see below that although this is the case, using suitable weighted integrals over the boundary produces far superior imaging; however, this is due to the amplitudes, not the underlying geometry.

For the remainder of the paper, we limit ourselves to the Calderón problem in $\mathbb{R}^2$; we begin by recalling the complex Beltrami equation formalism and CGO solutions of [Astala and Päivärinta 2006b], as well as their modification in [Huhtanen and Perämäki 2012]. The complex analysis in these approaches reflects the complex principal-type structure discussed above, disguised by the fact that we are working in two dimensions.

3. Conductivity equations and CGO solutions

On a domain $\Omega \subset \mathbb{R}^2 = \mathbb{C}$, let $\sigma \in L^\infty(\Omega)$ be a strictly positive conductivity, $\sigma \equiv 1$ near $\partial\Omega$, and extended to be $\equiv 1$ outside of Ω. The complex frequencies $\zeta \in \mathbb{C}^2$ with $\zeta \cdot \zeta = 0$ may be parametrized by $\zeta = (k, ik)$, $k \in \mathbb{C}$; thus, with $z = x_1 + ix_2$, one has $\zeta \cdot x = kz$. Following [Astala and Päivärinta 2006b], consider simultaneously the conductivity equations for the two scalar conductivities σ and σ^{-1},

$$\nabla \cdot \sigma \nabla u_1 = 0, \quad u_1 \sim e^{ikz}, \tag{3-1}$$

$$\nabla \cdot \sigma^{-1} \nabla u_2 = 0, \quad u_2 \sim e^{ikz}. \tag{3-2}$$

The complex geometrical optics (CGO) solutions of [Astala and Päivärinta 2006b], see also [Astala et al. 2010; 2014; Brown and Uhlmann 1997; Caro and Rogers 2016; Greenleaf and Uhlmann 2001; Haberman 2015; Haberman and Tataru 2013; Hamilton et al. 2012; Knudsen 2003; Knudsen et al. 2007; Nachman 1996], are specified by their asymptotics $u_j \sim e^{ikz}$, meaning that for all $k \in \mathbb{C}$,

$$u_j(z,k) = e^{ikz}\left(1 + \mathcal{O}\left(\frac{1}{z}\right)\right) \quad \text{as } |z| \to \infty. \tag{3-3}$$

The CGO solutions are constructed via the Beltrami equation

$$\bar{\partial}_z f_\mu = \mu\, \overline{\partial_z f_\mu}, \tag{3-4}$$

where the Beltrami coefficient μ is defined in terms of σ by

$$\mu := \frac{1-\sigma}{1+\sigma}. \tag{3-5}$$

The Beltrami coefficient μ is a compactly supported, $(-1, 1)$-valued function and, due to the assumption that $0 < c_1 \le \sigma \le c_2 < \infty$, one has $|\mu| \le 1-\epsilon$ for some $\epsilon > 0$. It was shown in [Astala and Päivärinta 2006b] that (3-4) has solutions for coefficients μ and $-\mu$ of the form

$$f_\mu(z,k) = e^{ikz}(1+\omega^+(z,k)) \quad \text{and} \quad f_{-\mu}(z,k) = e^{ikz}(1+\omega^-(z,k)), \tag{3-6}$$

with

$$\omega^\pm(z,k) = \mathcal{O}\left(\frac{1}{|z|}\right) \quad \text{as } |z| \to \infty.$$

The various CGO solutions are then related by the equation

$$2u_1(z,k) = f_\mu(z,k) + f_{-\mu}(z,k) + \overline{f_\mu(z,k)} - \overline{f_{-\mu}(z,k)}, \tag{3-7}$$

which follows from the fact that the real part of $f_\mu(z,k)$ solves (3-1), while the imaginary part solves (3-2).

In this work we will mainly focus on ω^+, henceforth denoted simply by ω; however, we will use ω^- in the symmetry discussion in Section 8. Both of these can be extracted from voltage/current measurements for σ at the boundary, $\partial\Omega$, as encoded in the Dirichlet-to-Neumann (DN) map of (3-1). For the most part we will suppress the superscripts $\pm$, with it being understood in the formulas that for $\omega^\pm$, one uses $\pm\mu$.

Huhtanen and Perämäki [2012] introduced the following modified derivation of ω, which, by avoiding issues caused by the exponential growth in the $k^\perp$-directions, is highly efficient from a computational point of view.

Let $e_k(z) := \exp(i(kz+\bar{k}\bar{z})) = \exp(i2\,\mathrm{Re}(kz))$; note that $|e_z(z)| \equiv 1$ and $\bar{e}_k = e_{-k}$. Define, as in [Astala and Päivärinta 2006a; 2006b],

$$\nu(z,k) := e_{-k}(z)\mu(z) \quad \text{and} \quad \alpha(z,k) := -i\bar{k}e_{-k}(z)\mu(z). \tag{3-8}$$

Both α and ν are compactly supported in Ω; since $\bar\partial\omega = \nu\overline{\partial\omega} + \alpha\bar\omega + \alpha$, we see that $\bar\partial\omega$ is compactly supported as well. For future use, also note that

$$\bar\nu(z,k) = e_k(z)\mu(z) \quad \text{and} \quad \bar\alpha(z,k) = ike_k(z)\mu(z). \tag{3-9}$$

Astala and Päivärinta [2006b, (4.8)] showed that $\omega(z,k)$ satisfies the inhomogeneous Beltrami equation

$$\bar\partial\omega - \nu\overline{\partial\omega} - \alpha\bar\omega = \alpha, \tag{3-10}$$

where the Cauchy–Riemann operator $\bar\partial$ and derivative ∂ are taken with respect to z. Recall the (solid) Cauchy transform P and Beurling transform S, defined by

$$Pf(z) = -\frac{1}{\pi}\int_{\mathbb{C}} \frac{f(z_1)}{z_1 - z}\, d^2z_1, \tag{3-11}$$

$$Sg(z) = -\frac{1}{\pi}\int_{\mathbb{C}} \frac{g(z_1)}{(z_1 - z)^2}\, d^2z_1, \tag{3-12}$$

which satisfy $\bar\partial P = I$, $S = \partial P$ and $S\bar\partial = \partial$ on $C_0^\infty(\mathbb{C})$; see [Astala et al. 2009].

It is shown in [Huhtanen and Perämäki 2012], using the results of [Astala and Päivärinta 2006b], that (3-10) has a unique solution $\omega \in W^{1,p}(\mathbb{C})$ for $2 < p < p_\epsilon := 1 + 1/(1-\epsilon)$, where $\epsilon > 0$ is such that $|\mu| \le 1-\epsilon$. Now define u on Ω by $\bar{u} = -\bar{\partial}\omega$; note that $u \in L^p(\Omega)$, $\omega = -P\bar{u}$ and $\partial\omega = -S\bar{u}$. Rewriting (3-10) in terms of u leads to

$$-\bar{u} - \nu(\overline{-S\bar{u}}) - \alpha(\overline{-P\bar{u}}) = \alpha.$$

Using (3-8), this further simplifies to

$$u + (-\bar{\nu}S - \bar{\alpha}P)\bar{u} = -\bar{\alpha}, \tag{3-13}$$

which then can be expressed as the integral equation

$$(I + A\rho)u = -\bar{\alpha}, \tag{3-14}$$

where $\rho(f) := \bar{f}$ denotes complex conjugation and $A := (-\bar{\alpha}P - \bar{\nu}S)$. As shown in [Astala and Päivärinta 2006b, Huhtanen and Perämäki 2012, Section 2], $I + A$ is invertible on $L^p(\Omega)$. Denote by $U(k, \mu) = u(\cdot\,, k)|_\Omega$ the restriction to $\overline{\Omega}$ of the unique solution to (3-14), and hence (3-13).

4. Fréchet differentiability and the Neumann series

We now come to the key construction of the paper. For $\epsilon > 0$ and any $\Omega_0 \Subset \Omega$, let

$$X = \{\mu \in L^\infty(\Omega) : \text{ess supp}(\mu) \subset \Omega_0,\ \|\mu\|_{L^\infty(\Omega)} \le 1-\epsilon\}.$$

Furthermore, define Y to be the closure of $C^\infty(\overline{\Omega})$ with respect to

$$||u||_Y := \|u\|_{L^2(\Omega)} + \|\,u|_{\partial\Omega}\,\|_{L^\infty(\partial\Omega)}.$$

For $k \in \mathbb{C}$, let U_k be the $\mathbb{R}$-linear map $U_k : X \to L^2(\Omega)$, given by $U_k(\mu) = u_\mu(\cdot\,, k)$, where $u_\mu(z, k)$ is the unique solution $u = u_\mu(\cdot\,, k) \in L^2(\Omega)$ of (3-13). Define $W_k : X \to Y$ by

$$W_k\mu = \omega_\mu(\cdot\,, k) = -P(\overline{u_\mu(\cdot\,, k)}).$$

4A. ***Fréchet differentiability.*** We will show that, for each $k \in \mathbb{C}$, W_k is a C^∞-map $X \to Y$ and analyze its Fréchet derivatives at $\mu_0 = 0$. For each k, one can solve (3-14) by a Neumann series which converges for $\|\mu\|_{L^\infty}$ sufficiently small. We analyze the individual terms of the series by introducing polar coordinates in the k-plane, $k = \tau e^{i\varphi}$, and then taking the partial Fourier transform in τ. The leading term in the Neumann series will be the basis for the edge-detection imaging technique that is the main point of the paper, while the higher-order terms are transformed into multilinear operators acting on μ. The remainder of the paper will then be devoted to understanding the Fourier-transformed terms, using the first derivative for effective edge detection in EIT and obtaining partial control over the higher derivatives.

Theorem 4.1. *The map $U_k : X \to L^2(\Omega)$, $U_k(\mu) := u_\mu(\cdot\,, k)$, is infinitely Fréchet-differentiable with respect to μ, and its Fréchet derivatives are real-analytic functions of $k \in \mathbb{C}$. Moreover, for $p \ge 1$, its p-th order Fréchet derivative at $\mu = 0$ in the direction $(\mu_1, \mu_2, \ldots, \mu_p) \in (L^2(\Omega_0))^p$ satisfies*

$$\left\| \frac{D^p U_k}{D\mu^p}\bigg|_{\mu=0}(\mu_1, \mu_2, \ldots, \mu_p) \right\|_{L^2(\Omega)} \le C_p(1+|k|)^p \|\mu_1\|_{L^2(\Omega)} \cdot \|\mu_2\|_{L^2(\Omega)} \cdots \|\mu_p\|_{L^2(\Omega)} \tag{4-1}$$

for some $C_p > 0$. In particular, the first Fréchet derivative has the form

$$\left.\frac{DU_k}{D\mu}\right|_{\mu=0}(\mu_1) = -P\rho(ike_{-k}\mu_1). \tag{4-2}$$

Moreover, for $k \in \mathbb{C}$ the map $W_k : X \to Y$,

$$W_k(\mu) := \omega_\mu(\cdot\,, k) = -P\rho(u_\mu(\cdot\,, k)),$$

is infinitely Fréchet-differentiable with respect to μ and its Fréchet derivatives are real-analytic functions of $k \in \mathbb{C}$.

Proof. We can rewrite (3-13) for $u = u_\mu(\cdot\,, k) \in L^2(\Omega)$ as

$$(I - e_k\mu S\rho)u + ike_k\mu P\rho u = ike_k\mu. \tag{4-3}$$

On the left-hand side, e_k and μ denote pointwise multiplication operators with the functions $e_k(z)$ and $\mu(z)$, respectively; on the right, $e_k(z)\mu(z)$ is an element of $L^2(\Omega)$.

Since $\|\rho\|_{L^2(\Omega)\to L^2(\Omega)} = 1$, $\|S\|_{L^2(\Omega)\to L^2(\Omega)} = 1$, and $\|\mu\|_{L^\infty(\Omega)} < 1$, the inverse operator $(I - e_k\mu S\rho)^{-1} : L^2(\Omega) \to L^2(\Omega)$ exists and is a C^ω function (i.e., a real analytic function) of k. Thus, (4-3) can be rewritten as

$$(I - B_{\mu,k})u = K_{\mu,k}(ike_k\mu), \tag{4-4}$$

where

$$K_{\mu,k}u = (I - e_k\mu S\rho)^{-1}u, \quad B_{\mu,k}u = K_{\mu,k}(ike_k\mu P\rho u). \tag{4-5}$$

Since $P : L^2(\Omega) \to L^2(\Omega)$ is a compact operator, (4-5) defines a compact operator $B_{\mu,k} : L^2(\Omega) \to L^2(\Omega)$. To find the kernel of $I - B_{\mu,k}$, consider $u^0 \in L^2(\Omega)$ satisfying $(I - B_{\mu,k})u^0 = 0$. Then,

$$(I - e_k\mu S\rho)u^0 + ike_k\mu P\rho u^0 = 0. \tag{4-6}$$

When we consider P, given in (3-11), as an operator $P : L^2(\Omega) \to L^2_{\mathrm{loc}}(\mathbb{C})$, equation (4-6) yields that $f^0(z) = -e^{ikz}(P\bar{u})(z) \in L^2_{\mathrm{loc}}(\mathbb{C})$ satisfies

$$\begin{aligned} \bar\partial_z f^0(z) &= \mu(z)\,\overline{\partial_z f^0(z)}, \quad z \in \mathbb{C}, \\ e^{-ikz} f^0(z) &= \mathcal{O}\left(\frac{1}{|z|}\right) \quad \text{as } |z| \to \infty. \end{aligned} \tag{4-7}$$

By [Astala and Päivärinta 2006b], the solution f^0 of (4-7) has to be zero. Hence,

$$u^0(z) = -\overline{\partial(e^{-ikz} f^0(z))} = 0$$

and the operator $I - B_{\mu,k} : L^2(\Omega) \to L^2(\Omega)$ is one-to-one. Thus the Fredholm equation (4-4) is uniquely solvable and we can write its solutions as $u = u_\mu(\cdot\,, k)$,

$$u_\mu(\cdot\,, k) = (I - B_{\mu,k})^{-1}K_{\mu,k}(ike_k\mu). \tag{4-8}$$

By the analytic Fredholm theorem, the maps $k \mapsto K_{\mu,k}$ and $k \mapsto (I - B_{\mu,k})^{-1}$ are real-analytic, $\mathbb{C} \to \mathcal{L}(L^2(\Omega), L^2(\Omega))$, where $\mathcal{L}(L^2(\Omega), L^2(\Omega))$ is the space of the bounded linear operators $L^2(\Omega) \to L^2(\Omega)$.

Define

$$K^{(p)} = \frac{D^p}{D\mu^p} K_{\mu,k}\Big|_{\mu=0} \quad \text{and} \quad B^{(p)} = \frac{D^p}{D\mu^p} B_{\mu,k}\Big|_{\mu=0}.$$

Since $K_{\mu,k}|_{\mu=0} = I$, we see that

$$K^{(p)}(\mu_1, \mu_2, \ldots, \mu_p) = \sum_{\sigma} (e_k \mu_{\sigma(1)} S\rho) \circ (e_k \mu_{\sigma(2)} S\rho) \circ \cdots \circ (e_k \mu_{\sigma(p)} S\rho),$$

where the sum is taken over permutations $\sigma : \{1, 2, \ldots, p\} \to \{1, 2, \ldots, p\}$. Furthermore, one has

$$\begin{aligned} B^{(p)}(\mu_1, \mu_2, \ldots, \mu_p) &= K^{(p-1)}(\mu_2, \mu_3, \mu_4, \ldots, \mu_p) \circ (ike_k \mu_1 P\rho) \\ &\quad + K^{(p-1)}(\mu_1, \mu_3, \mu_4, \ldots, \mu_p) \circ (ike_k \mu_2 P\rho) \\ &\quad + K^{(p-1)}(\mu_1, \mu_2, \mu_4, \ldots, \mu_p) \circ (ike_k \mu_3 P\rho) \\ &\quad + \cdots + K^{(p-1)}(\mu_1, \mu_2, \ldots, \mu_{p-1}) \circ (ike_k \mu_p P\rho). \end{aligned}$$

We can compute the higher-order derivatives

$$\frac{D^p}{D\mu^p} (I - B_{\mu,k})^{-1}\Big|_{\mu=0},$$

in the direction $(\mu_1, \mu_2, \ldots, \mu_p)$, using the polarization identity for symmetric multilinear functions, if these derivatives are known in the case when $\mu_1 = \mu_2 = \cdots = \mu_p$. In the latter case the derivatives can be computed using Faà di Bruno's formula, which generalizes the chain rule to higher derivatives,

$$\frac{d^p}{dt^p} f(g(t)) = \sum \frac{p!}{m_1! m_2! \cdots m_p!} \cdot f^{(m_1+\cdots+m_n)}(g(t)) \cdot \prod_{j=1}^{n} \left(\frac{g^{(j)}(t)}{j!} \right)^{m_j},$$

where the sum runs over indices $(m_1, m_2, \ldots, m_p) \in \mathbb{N}^p$ satisfying $m_1 + 2m_2 + \cdots + pm_p = p$. Indeed, this formula can be applied with $f(B) = (I - B)^{-1}$ and $g(t) = B_{t\mu_1,k}$. As $g(0) = 0$ and the norm of the p-th derivative of $B_{t\mu_1,k}$ with respect to t is bounded by $c_p(1 + |k|)^p \|\mu_1\|^p$, we obtain estimate (4-1). Moreover, since $k \mapsto ike_k\mu$ is a real analytic map, $\mathbb{C} \to L^2(\Omega)$, we see that the Fréchet derivatives

$$k \mapsto \frac{D^p u_\mu}{D\mu^p}\Big|_{\mu=0}(\cdot, k) \in L^2(\Omega)$$

are real analytic maps of $k \in \mathbb{C}$.

Finally, recall that $\Omega_0 \subset \Omega$ is a relatively compact set. For $\mu \in X$, we have $\operatorname{supp}(\mu) \subset \Omega_0$, and thus the function $u_\mu(\cdot, k) = U_k(\mu)$ is also supported in Ω_0. As P is given in (3-11) we see easily that for $(\mu_1, \mu_2, \ldots, \mu_p) \in (L^2(\Omega_1))^p$ the Fréchet derivatives

$$\frac{D^p W_k}{D\mu^p}\Big|_{\mu=0}(\mu_1, \mu_2, \ldots, \mu_p) = -P\rho \frac{D^p U_k}{D\mu^p}\Big|_{\mu=0}(\mu_1, \mu_2, \ldots, \mu_p)$$

are in Y, and these derivatives are real analytic functions of $k \in \mathbb{C}$. □

4B. ***Neumann series.*** Now consider a Neumann-series-expansion approach to solving (3-14), looking for $u \sim \sum_{n=1}^{\infty} u_n$, with $u_1 := -\bar{\alpha}$ and $u_{n+1} := -A\bar{u}_n,\ n \geq 1$; the resulting ω_n are defined by

$$\omega = -P\bar{u} \sim \sum_{n=1}^{\infty} -P\bar{u}_n =: \sum_{n=1}^{\infty} \omega_n.$$

The first three terms of each expansion are given by

$$u_1 = -\bar{\alpha}, \qquad \omega_1 = P\alpha, \tag{4-9}$$

$$u_2 = A\alpha = -(\bar{\alpha}P + \bar{\nu}S)(\alpha), \qquad \omega_2 = P(\alpha\overline{P\alpha} + \nu\overline{S\alpha}), \tag{4-10}$$

$$u_3 = -(\bar{\alpha}P + \bar{\nu}S)(\alpha\bar{P}\bar{\alpha} + \nu\bar{S}\bar{\alpha}), \quad \omega_3 = P(\alpha\bar{P} + \nu\bar{S})(\bar{\alpha}P\alpha + \bar{\nu}S\alpha). \tag{4-11}$$

By Theorem 4.1, $U_k : X \to L^2(\Omega)$ is C^∞, and hence we have

$$u_n(\cdot\,, k) = \frac{D^n U_k}{D\mu^n}\bigg|_{\mu_0=0}(\mu, \mu, \ldots, \mu), \quad \omega_n(\cdot\,, k) = -P\rho(u_n(\cdot\,, k)). \tag{4-12}$$

Due to the polynomial growth in the estimates (4-1), the functions $u_n(z, k)$ and $\omega_n(z, k)$ are tempered distributions in the k-variable. Hence we can introduce polar coordinates, $k = \tau e^{i\varphi}$, and then take the partial Fourier transform with respect to τ of the tempered distributions $\tau \mapsto u_n(z, k)|_{k=\tau e^{i\varphi}}$ and $\tau \mapsto \omega_n(z, k)|_{k=\tau e^{i\varphi}}$. Later we prove the following theorem concerning the partial Fourier transforms of the Fréchet derivatives:

Theorem 4.2. *Let $\mu \in X$ and consider the partial Fourier transforms of the Fréchet derivatives*

$$\begin{aligned} \hat{\omega}_n^{z_0}(t, e^{i\varphi}) &= \mathcal{F}_{\tau\to t}(\omega_n(z_0, k)|_{k=\tau e^{i\varphi}}), \quad n = 1, 2, \ldots, \\ \omega_n(\cdot\,, k) &= -P\rho\left(\frac{D^{n+1}U_k}{D\mu^{n+1}}\bigg|_{\mu_0=0}(\mu, \mu, \ldots, \mu)\right), \end{aligned} \tag{4-13}$$

which we denote at $z_0 \in \partial\Omega$ by

$$\hat{\omega}_n(z_0, t, e^{i\varphi}) = \hat{\omega}_n^{z_0}(t, e^{i\varphi}).$$

Then we have

$$\hat{\omega}_n^{z_0}(t, e^{i\varphi}) = T_n^{z_0}(\mu \otimes \cdots \otimes \mu),$$

where $T_n^{z_0}$ are n-linear operators given by

$$T_n^{z_0}(\mu_1 \otimes \cdots \otimes \mu_n) := \int_{\mathbb{C}^n} K_n^{z_0}(t, e^{i\varphi}; z_1, \ldots, z_n)\, \mu_1(z_1) \cdots \mu_n(z_n)\, d^2z_1 \cdots d^2z_n.$$

The wave front set of the Schwartz kernel $K_n^{z_0}$ is contained in the union of a collection $\{\Lambda_J : J \in \mathcal{J}\}$ of 2^{n-1} pairwise cleanly intersecting Lagrangian manifolds, indexed by $\mathcal{J}$, the power set of $\{1, \ldots, n-1\}$. For each $J \in \mathcal{J}$, we have Λ_J is the conormal bundle of a smooth submanifold, $L_n^J \subset \mathbb{R} \times \mathbb{S}^1 \times \mathbb{C}^n$, i.e., $\Lambda_J = N^*L_n^J$, with

$$L_n^J := \left\{t + (-1)^{n+1}\, 2\,\mathrm{Re}\left(e^{i\varphi}\sum_{j=1}^{n}(-1)^j z_j\right) = 0\right\} \cap \bigcap_{j\in J}\{z_j - z_{j+1} = 0\}. \tag{4-14}$$

Roughly speaking, Theorem 4.2 implies that the operator $T_n^{z_0}$ transforms singularities of μ to singularities of $\hat{\omega}_n^{z_0}$ so that the singularities of μ propagate along the L_n^J. Further discussion, as well as the proof of the theorem, will be found later in the paper.

The first-order term ω_1 will serve as the basis for stable edge and singularity detection, while the higher-order terms need to be characterized in terms their regularity and the location of their wave front sets. After the partial Fourier transform $\omega \to \hat{\omega}$ described in the next section, the map $T_1 : \mu \to \hat{\omega}_1$ turns out to be essentially a derivative of the Radon transform. Thus, *the leading term of $\hat{\omega}$ is a nonlinear Radon transform* of the conductivity σ, allowing for good reconstruction of the singularities of σ from the singularities of $\hat{\omega}_1$. The higher-order terms $\hat{\omega}_n$ record scattering effects and explain artifacts observed in simulations; these should be filtered out or otherwise taken into account for efficient numerics and accurate reconstruction. We characterize this scattering in detail for $\hat{\omega}_2$ in terms of oscillatory integrals, almost as precisely for $\hat{\omega}_3$, and in terms of the wave front set for $\hat{\omega}_n$, $n \geq 4$.

5. Fourier transform and the virtual variable

We continue the analysis with two elementary transformations of the problem:

(i) First, one introduces polar coordinates in the complex frequency, k, writing $k = \tau e^{i\varphi}$, with $\tau \in \mathbb{R}$ and $e^{i\varphi} \in \mathbb{S}^1$.

(ii) Secondly, one takes a partial Fourier transform in τ, introducing a nonphysical artificial (i.e., *virtual*) variable, t. We show that the introduction of this variable reveals the complex principal-type structure of the problem, as discussed in Section 2. This allows for good propagation of singularities from the interior of Ω to the boundary, allowing singularities of the conductivity in the interior to be robustly detected by voltage-current measurements at the boundary.

By (3-8), $\omega_1 = ikP(e_k\mu)$, see also (4-2), so that

$$\omega_1(z,k) = \frac{ik}{\pi}\int_{\mathbb{C}} \frac{e_k(z_1)\mu(z_1)}{z-z_1}\, d^2z_1. \tag{5-1}$$

Write the complex frequency as $k = \tau e^{i\varphi}$ with $\tau \in \mathbb{R}$, $\varphi \in [0, 2\pi)$ (which we usually identify with $e^{i\varphi} \in \mathbb{S}^1$). Taking the partial Fourier transform in τ then yields

$$\begin{aligned}
\hat{\omega}_1(z,t,e^{i\varphi}) &:= \int_{\mathbb{R}} e^{-it\tau}\omega_1(z,\tau e^{i\varphi})\, d\tau \\
&= \frac{e^{i\varphi}}{\pi}\int_{\mathbb{R}}\int_{\mathbb{C}} \frac{e^{-i\tau t}}{z-z_1}\,(i\tau)\, e_{\tau e^{i\varphi}}(z_1)\mu(z_1)\, d^2z_1\, d\tau \\
&= \frac{e^{i\varphi}}{\pi}\int_{\mathbb{R}}\int_{\mathbb{C}} (i\tau)\, \frac{e^{-i\tau(t-2\operatorname{Re}(e^{i\varphi}z_1))}}{z-z_1}\,\mu(z_1)\, d^2z_1\, d\tau \\
&= -2e^{i\varphi}\int_{\mathbb{C}} \frac{\delta'(t-2\operatorname{Re}(e^{i\varphi}z_1))}{z-z_1}\,\mu(z_1)\, d^2z_1,
\end{aligned} \tag{5-2}$$

with the integrals interpreted in the sense of distributions. Note that since t is dual to τ, which is the (signed) length of a frequency variable, for heuristic purposes t may be thought of as temporal.

5A. ***Microlocal analysis of $\hat{\omega}_1$.*** Fix $\Omega_0 \Subset \Omega_2 \Subset \Omega$ and assume once and for all that $\operatorname{supp}(\mu) \subset \Omega_0$, i.e., $\sigma \equiv 1$ on Ω_0^c. Let $\Omega_1 := (\bar{\Omega}_2)^c \supset \Omega^c \supset \partial\Omega$. Then the map $T_1 : \mathcal{E}'(\Omega_0) \to \mathcal{D}'(\Omega_1 \times \mathbb{R} \times \mathbb{S}^1)$, defined by

$$\mu(z_1) \to (T_1\mu)(z, t, e^{i\varphi}) := \hat{\omega}_1(z, t, e^{i\varphi}),$$

has Schwartz kernel

$$K_1(z, t, e^{i\varphi}, z_1) = -2e^{i\varphi} \frac{\delta'(t - 2\operatorname{Re}(e^{i\varphi} z_1))}{z - z_1}. \tag{5-3}$$

Note that $|z - z_1| \geq c > 0$ for $z \in \Omega_1$ and $z_1 \in \Omega_0$. For $z \in \partial\Omega$ and $z_1 \in \Omega_0$, the factor $(z - z_1)^{-1}$ in (5-3) is smooth, and T_1 acts on $\mu \in \mathcal{E}'(\Omega_0)$ as a standard Fourier integral operator (FIO). (See [Hörmander 1971] for the standard facts concerning FIOs which we use.) However, as we will see below, the amplitude $1/(z - z_1)$, although C^∞, both

(i) accounts for the fall-off rate in detectability of jumps, namely as the inverse of the distance from the boundary; and

(ii) causes artifacts, especially when some singularities of μ are close to the boundary, due to its large magnitude and the large gradient of its phase.

To see this, start by noting that the kernel K_1 is singular at the hypersurface,

$$L := \{(z, t, e^{i\varphi}, z_1) : t - 2\operatorname{Re}(e^{i\varphi} z_1) = 0\} \subset \mathbb{C} \times \mathbb{R} \times \mathbb{S}^1 \times \mathbb{C}.$$

Write $z = x + iy$, $z_1 = x' + iy'$", and use ζ, ζ' to denote their dual variables, (ξ, η), (ξ', η'). Using the defining function $t - 2\operatorname{Re}(e^{i\varphi} z_1) = t - 2(\cos(\varphi)x' - \sin(\varphi)y')$, identifying $\mathbb{C}$ with $\mathbb{R}^2$ as above and $\mathbb{S}^1$ with $[0, 2\pi)$, we see that the conormal bundle of L is

$$\Lambda := N^*L = \big\{\big(z, 2\operatorname{Re}(e^{i\varphi} z_1), e^{i\varphi}, x', y';\ 0, 0, \tau, 2\tau \operatorname{Im}(e^{i\varphi} z_1), -2\tau e^{-i\varphi}\big) : z \in \Omega_1,\ z_1 \in \Omega_0,\ e^{i\varphi} \in \mathbb{S}^1,\ \tau \in \mathbb{R} \setminus 0\big\}, \tag{5-4}$$

which is a Lagrangian submanifold of $T^*(\Omega_1 \times \mathbb{R} \times \mathbb{S}^1 \times \Omega_0) \setminus 0$. The kernel K_1 has the oscillatory representation

$$K_1(z, t, e^{i\varphi}, z_1) = \int_{\mathbb{R}} e^{i\tau(t - 2\operatorname{Re}(e^{i\varphi} z_1))} \frac{e^{i\varphi}(i\tau)}{\pi(z - z_1)}\, d\tau, \tag{5-5}$$

interpreted in the sense of distributions. The amplitude in (5-5) belongs to the standard space of symbols $S^1_{1,0}$ on $(\Omega_1 \times \mathbb{R} \times \mathbb{S}^1 \times \Omega_0) \times (\mathbb{R} \setminus 0)$ [Hörmander 1971]. Thus, using from that paper his notation and orders for Fourier integral (Lagrangian) distribution classes, K_1 is of order $1 + \frac{1}{2} - \frac{0}{4}$, i.e., $K_1 \in I^0(\Lambda)$. We conclude that T_1 is an FIO of order 0 associated with the canonical relation

$$C \subset (T^*(\Omega_1 \times \mathbb{R} \times \mathbb{S}^1) \setminus 0) \times (T^*\Omega_0 \setminus 0), \tag{5-6}$$

written $T_1 \in I^0(C)$, where

$$C = \Lambda' := \{(z, t, e^{i\varphi}, \zeta, \tau, \Phi; z_1, \zeta_1) : (z, t, e^{i\varphi}, z_1; \zeta, \tau, \Phi, -\zeta_1) \in \Lambda\}. \tag{5-7}$$

The wave front set of K_1 satisfies $\operatorname{WF}(K_1) \subset \Lambda$ (and actually, by the particular form of K_1, equality holds). Hence, by the Hörmander–Sato lemma [Hörmander 1971, Theorem 2.5.14], $\operatorname{WF}(T_1\mu) \subset C_0 \circ \operatorname{WF}(\mu)$, with C considered as a set-theoretic relation from $T^*\Omega_0 \setminus 0$ to $T^*(\Omega_1 \times \mathbb{R} \times \mathbb{S}^1) \setminus 0$.

We next consider the geometry of C, parametrized as

$$C = \left\{ \left(z, 2\,\mathrm{Re}(e^{i\varphi} z_1), e^{i\varphi}, 0, \tau, 2\tau\,\mathrm{Im}(e^{i\varphi} z_1); z_1, 2\tau e^{-i\varphi}\right) : z \in \Omega_1,\ z_1 \in \Omega_0,\ e^{i\varphi} \in \mathbb{S}^1,\ \tau \in \mathbb{R}\setminus 0 \right\}. \quad (5\text{-}8)$$

C is of dimension 6, while the natural projections to the left and right, $\pi_L : C \to T^*(\Omega_1 \times \mathbb{R} \times \mathbb{S}^1)\setminus 0$ and $\pi_R : C \to T^*\Omega_0 \setminus 0$, are into spaces of dimensions 8 and 4, respectively. C satisfies the Bolker condition [Guillemin 1985; Guillemin and Sternberg 1977]: π_L is an immersion (which is equivalent to π_R being a submersion) and is globally injective.

However, C in fact satisfies a much stronger condition than the Bolker condition: the geometry of C is *independent* of $z \in \Omega_1$, and it is a canonical graph in the remaining variables. If for any $z_0 \in \Omega_1$ we set $K_1^{z_0} = K_1|_{z=z_0}$, then one can factor $C = 0_{T^*\Omega_1} \times C_0$ (with the obvious reordering of the variables), where $0_{T^*\Omega_1}$ is the zero-section of $T^*\Omega_1$ and

$$\begin{aligned} C_0 := \mathrm{WF}(K_1^{z_0})' &= \left\{ \left(2\,\mathrm{Re}(e^{i\varphi} z_1), e^{i\varphi}, \tau, 2\tau\,\mathrm{Im}(e^{i\varphi} z_1); z_1, 2\tau e^{-i\varphi}\right) : z_1 \in \Omega_0,\ e^{i\varphi} \in \mathbb{S}^1,\ \tau \in \mathbb{R}\setminus 0 \right\} \\ &\subset (T^*(\mathbb{R}\times\mathbb{S}^1)\setminus 0)\times(T^*\Omega_0\setminus 0). \end{aligned} \quad (5\text{-}9)$$

(Note that $C_0 = N^* L_0'$, where

$$L_0 = \{(t, e^{i\varphi}, z_1) \in \mathbb{R} \times \mathbb{S}^1 \times \mathbb{C} : t - 2\,\mathrm{Re}(e^{i\varphi} z_1) = 0\}.)$$

From (5-8), (5-9) one can see that C satisfies the Bolker condition, but its product structure is in fact much more stringent.

Hence, it is reasonable to form determined (i.e., two-dimensional) data sets from two-dimensional slices of the full T_1 by fixing $z = z_0$; for these to correspond to boundary measurements, assume that $z_0 \in \partial\Omega \subset \Omega_1$. Thus, define $T_1^{z_0} : \mathcal{E}'(\Omega_0) \to \mathcal{D}'(\mathbb{R} \times \mathbb{S}^1)$ by $\mu(z_1) \to (T_1^{z_0}\mu)(t, \varphi) := \hat{\omega}_0(z_0, t, \varphi)$. $T_1^{z_0}$ has Schwartz kernel $K_1^{z_0}$ given by (5-5), but with z fixed at $z = z_0$, and thus $T_1^{z_0}$ is an FIO of order $1 + \frac{1}{2} - \frac{4}{4} = \frac{1}{2}$ with canonical relation C_0, i.e., $T_1^{z_0} \in I^{\frac{1}{2}}(C_0)$. Further, one easily checks from (5-9) that $\pi_R : C_0 \to T^*\Omega_0 \setminus 0$ and $\pi_L : C_0 \to T^*(\mathbb{R} \times \mathbb{S}^1) \setminus 0$ are local diffeomorphisms, injective if we either restrict $\tau > 0$ or $\phi \in [0, \pi)$, in which case C_0 becomes a global canonical graph.

Composing $T_1^{z_0}$ with the backprojection operator $(T_1^{z_0})^*$ then yields, by the transverse intersection calculus for FIOs [Hörmander 1971], a normal operator $(T_1^{z_0})^* T_1^{z_0}$ which is a ΨDO of order 1 on Ω_0, i.e., $(T_1^{z_0})^* T_1^{z_0} \in \Psi^1(\Omega_0)$. We will show that the normal operator is elliptic and thus admits a left parametrix, $Q(z, D) \in \Psi^{-1}(\mathbb{C})$, so that

$$Q(T_1^{z_0})^* T_0^{z_0} - I \text{ is a smoothing operator on } \mathcal{E}'(\Omega_0). \quad (5\text{-}10)$$

Therefore, $T_1^{z_0}\mu$ determines μ mod C^∞, making it possible to determine the singularities of the Beltrami multiplier μ, and hence those of the conductivity σ, from the singularities of $T_1^{z_0}\mu$. All of this follows from standard arguments once one shows that $T_1^{z_0}$ is an elliptic FIO.

To establish this ellipticity, we may, because $z_0 - z_1 \neq 0$ for $z_1 \in \Omega_0$, calculate the principal symbol $\sigma_{\mathrm{prin}}(T_1^{z_0})$ using (5-3). At a point of C_0, as given by the parametrization (5-9), we may calculate the

induced symplectic form $\varkappa_{C_0}$ on C_0,

$$\varkappa_{C_0} := \pi_R^*(\varkappa_{T^*\Omega_0}) = -2\tau\, d\varphi \wedge (s(\varphi)dx' + c(\varphi)dy') + 2d\tau \wedge (c(\varphi)dx' - s(\varphi)dy'), \tag{5-11}$$

so that $\varkappa_{C_0} \wedge \varkappa_{C_0} = 4\tau d\varphi \wedge d\tau \wedge dx' \wedge dy'$, and the half density satisfies

$$|\varkappa_{C_0} \wedge \varkappa_C|^{1/2} = 2|\tau|^{1/2}|d\varphi \wedge d\tau \wedge dx' \wedge dy|^{1/2}.$$

From this it follows that

$$\sigma_{\text{prin}}(T_1^{z_0}) = \frac{-2e^{i\varphi}(i\tau)}{2|\tau|^{1/2}(z_0 - z_1)} = \frac{(-ie^{i\varphi})\ \operatorname{sgn}(\tau)|\tau|^{1/2}}{z_0 - z_1},$$

which is elliptic of order $\frac{1}{2}$ on C_0.

Example. Although (5-10) allows imaging of general $\mu \in \mathcal{E}'(\Omega_0)$ from $\omega_1(z_0, \cdot, \cdot)$, consider the particular case where μ is a piecewise smooth function with jumps across an embedded smooth curve $\gamma = \{z : g(z) = 0\} \subset \Omega_0$ (not necessarily closed or connected), with unit normal n. In fact, consider the somewhat more general case of a μ which is *conormal of order* $m \in \mathbb{R}$, $m \le -1$, with respect to γ, i.e., is of the form

$$\mu(z) = \int_{\mathbb{R}} e^{ig(z)\theta}\, a_m(x, \theta)\, d\theta, \tag{5-12}$$

where a_m belongs to the standard symbol class $S^m_{1,0}(\Omega_0 \times (\mathbb{R} \setminus 0))$. (In general, we will denote the orders or biorders of symbols by subscripts.) A μ which is a piecewise smooth function with jumps across γ is of this form for $m = -1$; for $-2 < m < -1$, a μ given by (5-12) is piecewise smooth, as well as Hölder continuous of order $-m - 1$ across γ. (Recall that uniqueness in the Calderón problem for C^ω piecewise smooth conductivities was treated in [Kohn and Vogelius 1985] and some cases of conormal conductivities in [Greenleaf et al. 2003; Kim 2008].) As a Fourier integral distribution, $\mu \in I^m(\Gamma)$ for the Lagrangian manifold

$$\Gamma := N^*\gamma = \{(z_1, \theta\, n(z_1)) : z_1 \in \gamma,\ \theta \in \mathbb{R} \setminus 0\} \subset T^*\Omega_0 \setminus 0. \tag{5-13}$$

By the transverse intersection calculus, $T_1^{z_0}\mu \in I^{m+1/2}(\widetilde{\Gamma})$, where

$$\widetilde{\Gamma} := C \circ \Gamma = \left\{\left(2\operatorname{Re}(e^{i\varphi}z_1), e^{i\varphi}, \tau, 2\tau \operatorname{Im}(e^{i\varphi}z_1)\right) : z_1 \in \gamma,\ e^{i\varphi} = \overline{n(z_1)},\ \tau \in \mathbb{R}\setminus 0\right\} \subset T^*(\mathbb{R}\times\mathbb{S}^1)\setminus 0. \tag{5-14}$$

Thus, for φ fixed, $T_1^{z_0}\mu$ has singularities at those values of t of the form $t = 2\operatorname{Re}(e^{i\varphi}z_1)$ with z_1 ranging over the points of γ with $n(z_1) = e^{-i\varphi}$. (Under a finite order of tangency condition on γ, for each φ there are only a finite number of such points.) These values of t depend on φ but *are independent of* $z_0 \in \partial\Omega$; this reflects the complex principal-type geometry underlying the problem, which has propagated the singularities of μ out to *all* of the boundary points of Ω. Denoting these values of t by $t_j(e^{i\varphi})$, the distribution $T_1^{z_0}\mu$ has Lagrangian singularities conormal of order $m + \frac{1}{2}$ on $\mathbb{R}$ at $\{t_j\}$, and thus is of magnitude $\sim |t - t_j|^{-m-3/2}$ for $-\frac{3}{2} < m \le -1$. In particular, if μ is piecewise smooth with jumps, for which $m = -1$, the singularities have magnitude $\sim |t - t_j|^{-1/2}$.

Remark. More generally, since $T_1^{z_0}$ is an elliptic FIO of order $\frac{1}{2}$ associated to a canonical graph, if we denote the L^2-based Sobolev space of order $s \in \mathbb{R}$ by H^s, it follows that if $\mu \in H^s \setminus H^{s-1}$, then $T_1^{z_0}\mu \in H^{s-1/2} \setminus H^{s-3/2}$, allowing us to image general singularities of μ and hence σ.

5B. ***"Averages" of $\hat{\omega}_1$ and artifact removal.*** As described above, each $T_1^{z_0}$ is in $I^{1/2}(C_0)$; the symbol depends on z_0, the canonical relation (5-9) does not, and we now take advantage of this. For any $\mathbb{C}$-valued weight $a(\,\cdot\,)$ on $\partial\Omega$, define

$$\hat{\omega}_1^a(t, e^{i\varphi}) := \int_{\partial\Omega} \hat{\omega}_1(z_0, t, e^{i\varphi})\, a(z_0)\, d\boldsymbol{z_0}, \tag{5-15}$$

and denote by T_1^a the operator taking $\mu(z_1) \to \hat{\omega}_1^a(t, e^{i\varphi})$. (It will be clear from context when the superscript is a point $z_0 \in \partial\Omega$ and when it is a function $a(\,\cdot\,)$ on the boundary.) (We emphasize that (5-15) is a complex line integral.) Then T_1^a has kernel

$$\begin{aligned} K_1^a(t, e^{i\varphi}, z_1) &:= -2e^{i\varphi}\left[\int_{\partial\Omega} \frac{a(z_0)\, d\boldsymbol{z_0}}{z_0 - z_1}\right]\delta'(t - 2\,\mathrm{Re}(e^{i\varphi} z_1)) \\ &= -4\pi i e^{i\varphi}\alpha(z_1)\delta'(t - 2\,\mathrm{Re}(e^{i\varphi} z_1)), \end{aligned} \tag{5-16}$$

where

$$\alpha(z_1) = \frac{1}{2\pi i}\int_{\partial\Omega} \frac{a(z_0)\, d\boldsymbol{z_0}}{z_0 - z_1}, \quad z_1 \in \Omega,$$

is the Cauchy (line) integral of a. We thus have

$$\sigma_{\mathrm{prin}}(T_1^a) = 2\pi e^{i\varphi}\alpha(z_1)\,\mathrm{sgn}(\tau)|\tau|^{1/2} \quad \text{on } C_0,$$

and therefore $(T_1^a)^* T_1^a \in \Psi^1(\Omega_0)$, with

$$\sigma_{\mathrm{prin}}((T_1^a)^* T_1^a)(z, \zeta) = 2\pi^2 |\alpha(z)|^2 |\zeta|,$$

since, by (5-9), $|\tau| = \frac{1}{2}|\zeta'|$ on C_0. Thus,

$$(T_1^a)^* T_1^a = 2\pi^2 |\alpha|^2 \cdot |D_z| \bmod \Psi^0(\Omega_0).$$

By choosing $a \equiv (\pi\sqrt{2})^{-1}$ in (5-15), one has $\alpha \equiv (\pi\sqrt{2})^{-1}$ on Ω and $\sigma_{\mathrm{prin}}((T_1^a)^* T_1^a)(z, \zeta) = |\zeta|$, yielding

$$(T_1^a)^* T_1^a = |D_z| \bmod \Psi^0, \tag{5-17}$$

which faithfully reproduces the locations of the singularities of μ and accentuates their strength by one derivative. This is, in the context of our reconstruction method, an analogue of local (or Λ-) tomography [Faridani et al. 1992].

Alternatively (now with the choice of $a = 1/\pi$), one may obtain an exact *weighted, filtered backprojection* inversion formula,

$$(T_1^a)^*(|D_t|^{-1}) T_1^a = I \quad \text{on } L^2(\Omega_0). \tag{5-18}$$

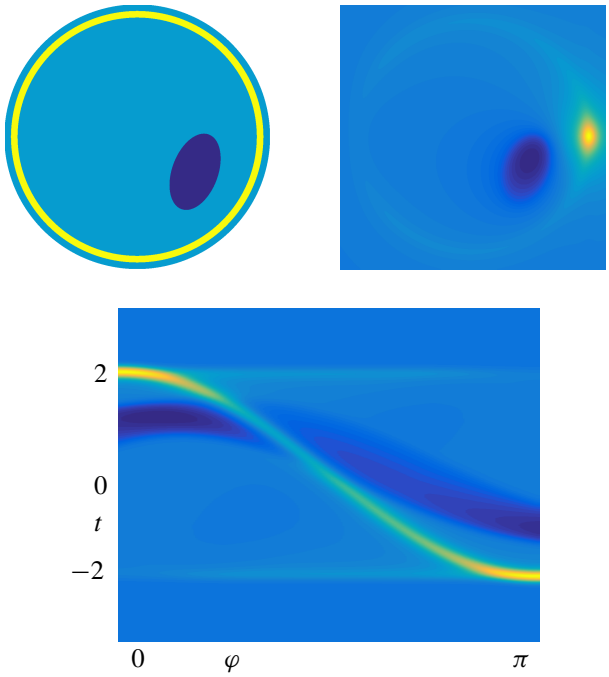

Figure 4. Artifacts from a single $T_1^{z_0}$. Top left: Phantom modeling hemorrhage (high conductivity inclusion) within skull (low conductivity shell). Bottom: $T_1^{z_0}\mu$ for $z_0 = 1$. Top right: backprojection applied to $T_1^{z_0}\mu$.

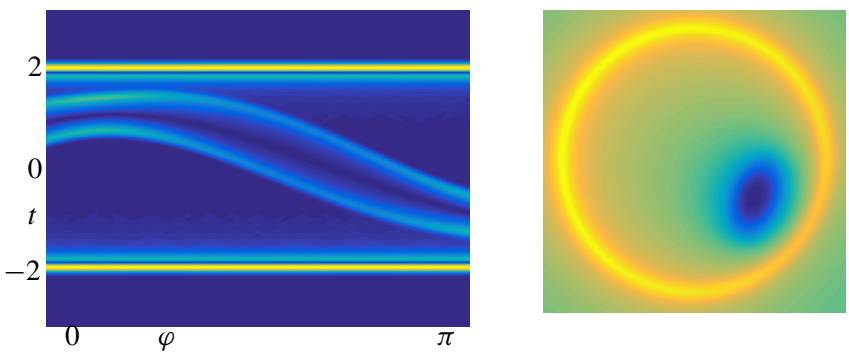

Figure 5. Artifact removal using weighted T_1^a. Left: $T_1^a\mu$ for phantom in Figure 4. Right: reconstruction from $T_1^a\mu$ using formula (5-18).

On the level of the principal symbol, this follows from the microlocal analysis above, again since $|\tau| = \frac{1}{2}|\zeta'|$ on C_0; for the exact result, note that

$$T_1^a = -\left(\frac{i\pi}{\sqrt{2}}\right)e^{i\varphi}\left(\frac{\partial}{\partial s}R\mu\right)\left(\frac{t}{2}, e^{i\varphi}\right), \tag{5-19}$$

where R is the standard Radon transform on $\mathbb{R}^2$,

$$(Rf)(s,\omega) = \int_{\boldsymbol{x}\cdot\omega=s} f(\boldsymbol{x})\, d^1\boldsymbol{x}, \quad (s,\omega)\in\mathbb{R}\times\mathbb{S}^1.$$

Remark. Note that if we take $\Omega=\mathbb{D}$, so that $\partial\Omega$ can be parametrized by $z_0=e^{i\theta}$, then (5-15) becomes

$$\hat\omega_1^a(t,e^{i\varphi}) = \int_0^{2\pi} \hat\omega_1(e^{i\theta},t,e^{i\varphi})\, ie^{i\theta}\, d\theta.$$

Thus, the weight is (slowly) oscillatory when expressed in terms of $d\theta$, but through destructive interference suppresses the artifacts present in each individual $\hat\omega_1^{z_0}$. Figure 5 illustrates, with a skull/hemorrhage phantom how using this simple weight removes the artifacts caused by the rapid change in the amplitude and phase of the Cauchy factor $(z_0-z_1)^{-1}$, shown in Figure 4.

6. Analysis of $\hat\omega_2$

Just as the introduction of polar coordinates and partial Fourier transform, applied to the zeroth-order term in the Neumann expansion (i.e., the Fréchet derivative of the scattering map at $\mu=0$), give rise to a term linear in μ, their application to the first-order term (4-10) gives rise to a term which is bilinear in μ. Wave front set analysis shows that this nonlinearity gives rise to two distinct types of singularities; we will see in Section 10 that both of these are visible in the numerics, and need to be taken into account to give good reconstruction based on $\hat\omega_1^a$.

We can rewrite (4-10) as

$$\omega_2(z,k) = P(\alpha(\bar P\bar\alpha)) + P(\nu(\bar S\bar\alpha)),$$

where the linear operators $\bar P$, $\bar S$ are defined by $\bar P(f)=\overline{P(\bar f)}$ and $\bar S(f)=\overline{S(\bar f)}$. The kernels of $\bar P$, $\bar S$ are just the complex conjugates of the kernels of P, S in (3-11), (3-12), respectively. We now denote the two interior variables in Ω_0 by z_1 and z_2; using (3-9), one sees that

$$\omega_2(z,k) = \frac{-k^2}{\pi^2}\int_{\mathbb{C}}\int_{\mathbb{C}} \frac{e^{-2i\,\mathrm{Re}(kz_1)}\mu(z_1)}{z_1-z}\,\frac{e^{2i\,\mathrm{Re}(kz_2)}\mu(z_2)}{\bar z_2-\bar z_1}\, d^2z_1\, d^2z_2 + \frac{ik}{\pi^2}\int_{\mathbb{C}}\int_{\mathbb{C}} \frac{e^{-2i\,\mathrm{Re}(kz_1)}\mu(z_1)}{z_1-z}\,\frac{e^{2i\,\mathrm{Re}(kz_2)}\mu(z_2)}{(\bar z_2-\bar z_1)^2}\, d^2z_1\, d^2z_2. \quad (6\text{-}1)$$

Thus, for $z_0\in\partial\Omega$,

$$\hat\omega_2(z_0,t,e^{i\varphi}) = \int_{\mathbb{R}} e^{-it\tau}\omega_1(z_0,\tau e^{i\varphi})\, d\tau = \int_{\mathbb{C}}\int_{\mathbb{C}} K_1(z_0,t,e^{i\varphi};z_1,z_2)\,\mu(z_1)\,\mu(z_2)\, d^2z_1\, d^2z_2 \quad (6\text{-}2)$$

is given by a bilinear operator acting on $\mu\otimes\mu$, with kernel

$$K_2^{z_0}(t,e^{i\varphi};z_1,z_2) = \frac{1}{\pi^2}\left(\frac{e^{2i\varphi}\delta''(t+2\,\mathrm{Re}(e^{i\varphi}(z_1-z_2)))}{(z_1-z_0)(\bar z_2-\bar z_1)} + \frac{e^{i\varphi}\delta'(t+2\,\mathrm{Re}(e^{i\varphi}(z_1-z_2)))}{(z_1-z_0)(\bar z_2-\bar z_1)^2}\right). \quad (6\text{-}3)$$

$K_2^{z_0}$ has multiple singularities, but, as in the case of K_1, the fact that $|z_1-z_0|\geq c>0$ for $z_0\in\partial\Omega$ and $z_1\in\mathrm{supp}(\mu)\subset\Omega_0$ eliminates the singularities at $\{z_1-z_0=0\}$. The remaining singularities put $K_2^{z_0}$ in

the general class of paired Lagrangian distributions introduced in [Melrose and Uhlmann 1979; Guillemin and Uhlmann 1981]. In fact, $K_2^{z_0}$ lies in a more restrictive class of *nested conormal distributions*, see [Greenleaf and Uhlmann 1990], associated with the pair (independent of z_0)

$$L_1 := \{t + 2\,\mathrm{Re}(e^{i\varphi}(z_1 - z_2)) = 0\} \supset L_3 := \{t + 2\,\mathrm{Re}(e^{i\varphi}(z_1 - z_2)) = 0,\ z_1 - z_2 = 0\} = \{t = 0,\ z_2 = z_1\}. \quad (6\text{-}4)$$

(The subscripts are chosen to indicate the respective codimensions in $\mathbb{R}_t \times \mathbb{S}^1_\varphi \times \Omega_{0,z_1} \times \Omega_{0,z_2}$.) These submanifolds have conormal bundles,

$$\Lambda_1 := N^*L_1, \quad \Lambda_3 := N^*L_3 \subset T^*(\mathbb{R}_t \times \mathbb{S}^1_\varphi \times \Omega_{0,z_1} \times \Omega_{0,z_2}) \setminus 0,$$

and $\mathrm{WF}(K_2^{z_0}) \subseteq \Lambda_1 \cup \Lambda_3$. (As with $K_1^{z_0}$, one can show from (6-3) that equality holds.)

6A. ***Bilinear wave front set analysis.*** Define $\hat\omega_2^{z_0} = \hat\omega_2|_{z=z_0}$. Since $\hat\omega_2^{z_0}(t, e^{i\varphi}) = \langle K_2^{z_0}(t, e^{i\varphi}, \cdot, \cdot), \mu\otimes\mu\rangle$, we have

$$\mathrm{WF}(\hat\omega_2^{z_0}) \subset \mathrm{WF}(K_2^{z_0})' \circ \mathrm{WF}(\mu\otimes\mu) \subset (\Lambda_1' \cup \Lambda_3') \circ \mathrm{WF}(\mu\otimes\mu).$$

Parametrizing Λ_1, Λ_3 in the usual way as conormal bundles, multiplying the variables dual to z_1, z_2 by -1 and then separating the variables on the left and right, we obtain canonical relations in $T^*(\mathbb{R}\times\mathbb{S}^1) \times T^*(\Omega_0\times\Omega_0)$,

$$C_1 := \Lambda_1' = \big\{\big(-2\,\mathrm{Re}(e^{i\varphi}(z_1 - z_2)), e^{i\varphi}, \tau, -2\tau\,\mathrm{Im}(e^{i\varphi}(z_1 - z_2)); z_1, z_2, -2\tau e^{i\varphi}, 2\tau e^{i\varphi}\big) : e^{i\varphi} \in \mathbb{S}^1,\ z_1, z_2 \in \Omega_0,\ \tau \in \mathbb{R}\setminus 0\big\}, \quad (6\text{-}5)$$

$$C_3 := \Lambda_3' = \{(0, e^{i\varphi}, \tau, 0; z_1, z_1, \zeta, -\zeta) : e^{i\varphi} \in \mathbb{S}^1,\ z_1 \in \Omega_0,\ (\tau, \zeta) \in \mathbb{R}^3 \setminus 0\}. \quad (6\text{-}6)$$

Representing $\mu\otimes\mu = \mu(z_1)\mu(z_2)$ as $(\mu\otimes 1)\cdot(1\otimes\mu)$, from a basic result concerning wave front sets of products [Hörmander 1971, Theorem 2.5.10], one sees that

$$\begin{aligned}\mathrm{WF}(\mu\otimes\mu) &\subseteq \mathrm{WF}(\mu\otimes 1) \cup \mathrm{WF}(1\otimes\mu) \cup (\mathrm{WF}(\mu\otimes 1) + \mathrm{WF}(1\otimes\mu))\\ &\subseteq (\mathrm{WF}(\mu) \times O_{T^*\Omega_0}) \cup (O_{T^*\Omega_0} \times \mathrm{WF}(\mu)) \cup (\mathrm{WF}(\mu) \times \mathrm{WF}(\mu)),\end{aligned} \quad (6\text{-}7)$$

where the sets are interpreted as subsets of $T^*\mathbb{C}^2 \setminus 0$, writing elements as either $(z_1, z_2; \zeta_1, \zeta_2)$ or $(z_1, \zeta_1; z_2, \zeta_2)$.

Since $\zeta_1 \neq 0,\ \zeta_2 \neq 0$ at all points of C_1, and $\zeta_1 = 0 \Longleftrightarrow \zeta_2 = 0$ on C_3, the relation $C_1 \cup C_3$, when applied to the first two terms on the right-hand side of (6-7), gives the empty set.

On the other hand, $C_1 \cup C_3$, when applied to $\mathrm{WF}(\mu) \times \mathrm{WF}(\mu)$, contributes nontrivially to $\mathrm{WF}(\hat\omega_2^{z_0})$. First, the application of C_3 gives

$$\big\{(0, e^{i\varphi}, \tau, 0) : \exists\, z_1 \text{ such that } (z_1, \tau e^{-i\varphi}) \in \mathrm{WF}(\mu)\big\} \subset N^*\{t = 0\}. \quad (6\text{-}8)$$

Secondly, C_1 yields a contribution to $\mathrm{WF}(\hat\omega_2^{z_0})$ contained in what we call the CGO *two-scattering* of μ, defined by

$$\mathrm{Sc}^{(2)}(\mu) := \big\{\big(-2\,\mathrm{Re}(e^{i\varphi}(z_1 - z_2)), e^{i\varphi}, \tau, -2\tau\,\mathrm{Im}(e^{i\varphi}(z_1 - z_2))\big) : \exists\, z_1, z_2 \in \Omega_0 \text{ such that } (z_1, \tau e^{-i\varphi}), (z_2, -\tau e^{-i\varphi}) \in \mathrm{WF}(\mu)\big\}. \quad (6\text{-}9)$$

Thus, pairs of points in $\mathrm{WF}(\mu)$ with spatial coordinates $z_1,\ z_2$ and antipodal covectors $\pm\tau e^{-i\varphi}$ give rise to elements of $\mathrm{WF}(\hat\omega_2^{z_0})$ at $t=-2\,\mathrm{Re}(e^{i\varphi}(z_1-z_2))$. Note that the expression in (6-8) is not necessarily contained in $\mathrm{Sc}^{(2)}(\mu)$, even if we allow $z_1=z_2$ in (6-9), since $\mathrm{WF}(\mu)$ is not necessarily symmetric under $(z,\zeta)\to(z,-\zeta)$ (although this does hold for μ which are smooth with jumps).

For later use, it is also convenient to define

$$\mathrm{Sc}^{(0)}(\mu):=N^*\{(t,e^{i\varphi}):t=0\}\quad\text{and}\quad\mathrm{Sc}^{(1)}(\mu):=C_0\circ\mathrm{WF}(\mu),\tag{6-10}$$

where C_0 is as in (5-9) above, so that the wave front set analysis so far can be summarized as

$$\mathrm{WF}(\hat\omega_1)\subset\mathrm{Sc}^{(1)}(\mu)\quad\text{and}\quad\mathrm{WF}(\hat\omega_2)\subset\mathrm{Sc}^{(0)}(\mu)\cup\mathrm{Sc}^{(2)}(\mu).\tag{6-11}$$

This is extended to general $\mathrm{WF}(\hat\omega_n)$ in (7-6) below.

Remarks. (1) Note that if the $\hat\omega_2^{z_0}$ are averaged out using a function $a(z_0)$ on $\partial\Omega$ as was done for $\hat\omega_1$, the wave front analysis above is still valid for the resulting $\hat\omega_2^a$, and we will refer to either as simply $\hat\omega_2$ in the following discussion.

(2) It follows from (6-8) that for any μ with $\mu\notin C^\infty$, and any $z_0\in\partial\Omega$, we always will see singularities of $\hat\omega_2$ at $t=0$. The only dependence on μ of these artifacts in $\mathrm{WF}(\hat\omega_2)$ is determined by the incident directions φ of the complex plane wave for which they occur, as dictated by (6-8).

(3) In addition, by (6-9), any spatially separated singularities of μ with antipodal covectors $\pm\zeta=\pm(\xi,\eta)$ give rise to singularities of $\hat\omega_2$ at $t=-2\,\mathrm{Re}(e^{i\varphi}(z_1-z_2))$, $\varphi=-\arg(\zeta)$. Under translations, neither the covectors nor the differences z_1-z_2 associated to such scatterings change, although the factor $(z_1-z)^{-1}$ in the kernel (6-3), which is evaluated at $z=z_0$, does. Hence, the locations and orders of these artifacts (but not their magnitude or phase) are essentially independent of translations within Ω_0 of inclusions present in μ.

Given the invertibility of T_1^a mod C^∞ (at least for constant weight $a(\,\cdot\,)$), from the point of view of our reconstruction method, the singularities of $\hat\omega_2^a$ at $t=0$ and at $\mathrm{Sc}^{(2)}(\mu)$, although part of $\hat\omega$, produce artifacts which interfere with reconstruction of the singularities of μ and should either be better characterized or filtered out. In the next subsection, we do the former for a class of μ which includes those which are piecewise smooth with jumps.

6B. ***Bilinear operator theory.*** Not only is $\mathrm{WF}(K_2^{z_0})\subset\Lambda_1\cup\Lambda_3$, but in fact $K_1^{z_0}$ belongs to the class of nested conormal distributions associated with the pair $L_1\supset L_3$, see [Greenleaf and Uhlmann 1990], and thus to the Lagrangian distributions associated with the cleanly intersecting pair $\Lambda_1,\ \Lambda_3$,

$$K_2^{z_0}\in I^{1,0}(\Lambda_1,\Lambda_3)+I^{1,-1}(\Lambda_1,\Lambda_3).$$

Any K_2^a is a linear superposition of these and thus belongs to the same class. The *linear* operators $T_2^{z_0}$, $T_2^a:\mathcal{E}'(\Omega_0\times\Omega_0)\to\mathcal{D}'(\mathbb{R}\times\mathbb{S}^1)$ with Schwartz kernels $K_2^{z_0}$, K_2^a respectively, which we will refer to simply as T_2, thus belong to a sum of spaces of singular Fourier integral operators, $I^{1,0}(C_1,C_3)+I^{1,-1}(C_1,C_3)$, and have some similarity to singular Radon transforms [Phong and Stein 1986], see also [Greenleaf and

Uhlmann 1990], but (i) this underlying geometry has to our knowledge not been studied before; and (ii) we are interested in *bilinear* operators with these kernels. Rather than pursuing optimal bounds for T_2 on function spaces, we shall focus on the goal of characterizing the singularities of $\hat\omega_2$ when μ is piecewise smooth with jumps. We will show that, away from $t=0$, $\hat\omega_2$ is half a derivative smoother than $\hat\omega_1$. On the other hand, at $t=0$ it is possible for $\hat\omega_2$ to be as singular as the strongest singularities of $\hat\omega_1$; this is present in the full $\hat\omega$ (computed from the DN data) and produces strong artifacts, which can be seen in numerics when attempting to reconstruct μ. For this reason, data should be either preprocessed by filtering out a neighborhood of $t=0$ before applying backprojection, or alternatively one should resort to the subtraction techniques discussed in Section 8.

It will be helpful to work (as with the example (5-12) above) in the slightly greater generality of distributions (still denoted by μ) that are conormal for a curve $\gamma\subset\Omega_0$, having an oscillatory integral representation such as (5-12) with an amplitude of some order $m\in\mathbb{R}$. For such a μ (even for one not coming from a conductivity), we may still define both $\hat\omega_1^{z_0}$ and $\hat\omega_1^{a}$ (denoted generically by $\hat\omega_1$), and they belong to $I^{m+1/2}(\widetilde\Gamma)$, where $\widetilde\Gamma = C_0\circ N^*\gamma\subset T^*(\mathbb{R}\times\mathbb{S}^1)\setminus 0$ is as in (5-14). We also define $\hat\omega_2 := T_2^{z_0}(\mu\otimes\mu)$ or $T_2^a(\mu\otimes\mu)$.

To make the microlocal analysis of $\hat\omega_2$ tractable, we now impose a curvature condition on γ: since $\nabla g(z)\perp T_z\gamma$ at a point $z\in\gamma$, we have $i\nabla g(z)\in T_z g$; thus, γ has nonzero Gaussian curvature at z if and only if

$$(i\nabla g(z))^t\nabla^2 g(z)(i\nabla g(z))\neq 0, \tag{6-12}$$

which *we henceforth assume holds at all points of* γ (or at least at all $z\in\operatorname{sing\,supp}\mu\subset\gamma$, which is all that matters).

Note that (6-12) implies the finite-order tangency condition referred to in the Example of Section 5A, so that for each $e^{i\varphi}\in\mathbb{S}^1$, we have $\hat\omega_0(\,\cdot\,,e^{i\varphi})$ is singular at a finite number of values $t=t_j(e^{i\varphi})$.

Theorem 6.1. *Under the curvature assumption (6-12),*

(i) $\mathrm{Sc}^{(2)}(\mu)$, *defined as in (6-9), is a smooth Lagrangian manifold in* $T^*(\mathbb{R}\times\mathbb{S}^1)\setminus 0$; *and*

(ii) *if* μ *is as in (5-12) for some* $m\in\mathbb{R}$, *then*

$$\hat\omega_2 = T_2^{z_0}(\mu\otimes\mu)\in I^{2m+3/2,-1/2}(\mathrm{Sc}^{(2)}(\mu),\mathrm{Sc}^{(0)}(\mu)). \tag{6-13}$$

Microlocally away from $\Lambda_0\cap\Lambda_1$, a distribution $u\in I^{p,l}(\Lambda_0,\Lambda_1)$ belongs to $I^p(\Lambda_1\setminus\Lambda_0)$ and to $I^{p+l}(\Lambda_0\setminus\Lambda_1)$ [Melrose and Uhlmann 1979; Guillemin and Uhlmann 1981]. Thus, $\hat\omega_2\in I^{2m+1}(\mathrm{Sc}^{(2)}(\mu))$ on $\mathrm{Res}^{(2)}(\mu)\setminus N^*\{t=0\}$ and hence is smoother than $\hat\omega_1\in I^{m+1/2}(\widetilde\Gamma)$ if $m<-\frac12$. In contrast, on $N^*\{t=0\}\setminus\mathrm{Sc}^{(2)}(\mu)$, one has $\hat\omega_2\in I^{2m+3/2}(N^*\{t=0\})$, which is guaranteed to be smoother than $\hat\omega_1$ only if $m<-1$.

In particular, for $m=-1$, corresponding to σ (and hence μ) being piecewise smooth with jumps, one has $\hat\omega_2\in I^{-1}(\mathrm{Sc}^{(2)}(\mu))$, while $\hat\omega_1\in I^{-1/2}(\widetilde\Gamma)$, so that these artifacts are half a derivative smoother than the faithful image of μ encoded by $\hat\omega_1$. On the other hand, the singularity of $\hat\omega_2$ at $N^*\{t=0\})$ can be just as strong as the singularity of $\hat\omega_1$ at $\widetilde\Gamma$.

To summarize: for conductivities with jumps, applying standard Radon transform backprojection methods to the full data $\hat{\omega}$, or even its approximation $\hat{\omega}_1+\hat{\omega}_2$, rather than just $\hat{\omega}_1$ (which is not measurable directly) can result in artifacts which are smoother than the leading singularities only if one filters out a neighborhood of $t=0$.

To see (i) and (ii), start by noting from (6-3) that $T_2(\mu\otimes\mu)(t,e^{i\varphi})$ is a sum of two terms of the form

$$\int e^{i\Phi}\, a_{p,l}(*;\tau;\sigma)\, b_m(z_1;\theta_1)\, b_m(z_2;\theta_2)\, d\theta_1\, d\theta_2\, dz_1\, dz_2\, d\sigma\, d\tau, \tag{6-14}$$

where (recalling that g is a defining function for γ),

$$\begin{aligned}\Phi &= \Phi(t,e^{i\varphi},z_1,z_2,\tau,\sigma,\theta_1,\theta_2)\\ &:= \tau(t+2\operatorname{Re}(e^{i\varphi}(z_1-z_2)))+\sigma\cdot(z_1-z_2)+\theta_1 g(z_1)+\theta_2 g(z_2),\end{aligned} \tag{6-15}$$

$b_m\in S^m_{1,0}(\Omega_0\times(\mathbb{R}\setminus 0))$, and the $a_{p,l}$ are product-type symbols satisfying

$$|\partial^\gamma_{t,\varphi,z_1,z_2}\partial^\beta_\sigma\partial^\alpha_\tau a_{p,l}(*;\tau;\sigma)|\lesssim\langle\tau\rangle^{p-\alpha}\langle\sigma\rangle^{l-|\beta|}$$

on $(\mathbb{R}\times\mathbb{S}^1\times\Omega_0\times\Omega_0)\times\mathbb{R}_\tau\times\mathbb{R}^2_\sigma$, (the $*$ denoting all of the spatial variables) of biorders $(p,l)=(2,-1)$ and $(1,0)$, respectively. As can be seen from (6-5), (6-6),

$$C_1,\ C_3\subset\{\zeta_2=-\zeta_1,\ |\zeta_1|=2|\tau|\}\subset\{|\zeta_1|=|\zeta_2|=2|\tau|\},$$

so one can microlocalize the amplitudes in (6-14) to $\{|\theta_1|\sim|\theta_2|\sim|\tau|\}$ and thus replace the $a_{p,l}\cdot b_m\cdot b_m$ by amplitudes

$$a_{p+2m,l}(*;(\tau,\theta_1,\theta_2);\sigma)\in S^{p+2m,l}(\mathbb{R}\times\mathbb{S}^1\times\Omega_0\times\Omega_0\times(\mathbb{R}^3_{\tau,\theta_1,\theta_2}\setminus 0)\times\mathbb{R}^2_\sigma)$$

with biorders $(2m+2,-1)$ and $(2m+1,0)$, respectively.

Now homogenize the variables $z_1,\ z_2$, by defining phase variables $\eta_j:=\tau z_j\ \ j=1,2$. In terms of the estimates for derivatives, the new phase variables are grouped with the elliptic variables (τ,θ_1,θ_2); furthermore, the change of variables involves a Jacobian factor of τ^{-4}, so that, mod C^∞, (6-14) becomes

$$\int e^{i\widetilde{\Phi}}\, a_{\tilde{p},\tilde{l}}(*;(\tau,\theta_1,\theta_2,\eta_1,\eta_2);\sigma)\, d\tau\, d\theta_1\, d\theta_2\, d\eta_1\, d\eta_2\, d\sigma, \tag{6-16}$$

with

$$\begin{aligned}\widetilde{\Phi} &= \widetilde{\Phi}(t,e^{i\varphi};\ \tau,\theta_1,\theta_2,\eta_1,\eta_2;\ \sigma)\\ &:= \tau t+2\operatorname{Re}(e^{i\varphi}(\eta_1-\eta_2))+\theta_1 g\left(\frac{\eta_1}{\tau}\right)+\theta_2 g\left(\frac{\eta_2}{\tau}\right)+\sigma\cdot\left(\frac{\eta_1-\eta_2}{\tau}\right)\end{aligned} \tag{6-17}$$

on $(\mathbb{R}\times\mathbb{S}^1)\times(\mathbb{R}^7_{\tau,\theta_1,\theta_2,\eta_1,\eta_2}\setminus 0)\times\mathbb{R}^2_\sigma$ and with amplitude biorders $(\tilde{p},\tilde{l})=(2m-2,-1)$ and $(2m-3,0)$, respectively. We interpret $\widetilde{\Phi}$ as (a slight variation of) a multiphase function in the sense of [Mendoza 1982]: one can check that $\widetilde{\Phi}_0:=\widetilde{\Phi}|_{\sigma=0}$ is a nondegenerate phase function (i.e., clean with excess $e_0=0$) which parametrizes $\mathrm{Sc}^{(2)}(\mu)$ (which is thus a smooth Lagrangian). One does this by verifying, using (6-12), that $d^2_{(t,\phi,\tau,\theta_1,\theta_2,\eta_1,\eta_2),(\tau,\theta_1,\theta_2,\eta_1,\eta_2)}\widetilde{\Phi}_0$ has maximal rank at $\{d_{(\tau,\theta_1,\theta_2,\eta_1,\eta_2)}\widetilde{\Phi}_0=0\}$, namely $=7$. On the other hand, the full phase function $\widetilde{\Phi}$ parametrizes $N^*\{t=0\}$, but rather than being nondegenerate,

it is clean with excess $e_1 = 1$, i.e., $d^2_{(t,\phi,\tau,\theta_1,\theta_2,\eta_1,\eta_2,\sigma),(\tau,\theta_1,\theta_2,\eta_1,\eta_2,\sigma)}\widetilde{\Phi}$ has constant rank $9-1=8$ at $\{d_{(\tau,\theta_1,\theta_2,\eta_1,\eta_2,\sigma)}\widetilde{\Phi}=0\}$. (See [Hörmander 1985] for a discussion of clean phase functions.) A slight modification of the results in [Mendoza 1982] yields the following.

Proposition 6.2. *Suppose two smooth conic Lagrangians $\Lambda_0,\ \Lambda_1 \subset T^*\mathbb{R}^n \setminus 0$ intersect cleanly in codimension k. Let $\phi(x,\theta,\sigma)$ be a phase function on $\mathbb{R}^n \times (\mathbb{R}^{N+M}\setminus 0)$ such that parametrizes Λ_1 cleanly with excess $e_1 \geq 0$ and $\phi_0(x,\theta) := \phi|_{\sigma=0}$ parametrizes Λ_0 cleanly with excess $e_0 \geq 0$. Suppose further that $a \in S^{\tilde{p},\tilde{l}}(\mathbb{R}^n \times (\mathbb{R}^N\setminus 0)\times \mathbb{R}^M)$. Then,*

$$u(x) := \int_{\mathbb{R}^{N+M}} e^{i\phi_1(x,\theta,\sigma)} a(x,\theta,\sigma)\, d\theta\, d\sigma \in I^{p',l'}(\Lambda_0,\Lambda_1),$$

with

$$p' = \tilde{p}+\tilde{l}+\frac{N+M+e_0+e_1}{2}-\frac{n}{4}, \quad l' = -\tilde{l}-\frac{M+e_1}{2}.$$

Applying the proposition to each of the two biorders $(\tilde{p},\tilde{l}) = (2m-2,-1)$ and $(2m-3,0)$ from above, we see that $T_2^{z_0}(\mu\otimes\mu)$, as given by the expression (6-16), is a sum of two terms,

$$\hat{\omega}_2^{z_0} = T_2^{z_0}(\mu\otimes\mu) \in (I^{2m+3/2,-1/2}+I^{2m+3/2,-3/2})(\mathrm{Sc}^{(2)}(\mu), N^*\{t=0\}).$$

Recalling that $N^*\{t=0\} =: \mathrm{Sc}^{(0)}(\mu)$ and also that $I^{p',l''}\subset I^{p',l'}$ for $l''\leq l'$, this yields (6-13), finishing the proof of Theorem 6.1. □

7. Higher-order terms

7A. ***Multilinear wave front set analysis.*** For $n\geq 3$, and for any conductivity σ, one can analyze $\mathrm{WF}(\hat{\omega}_n^{z_0})$ and $\mathrm{WF}(\hat{\omega}_n^a)$ by n-linear versions of the case $n=2$ treated in Section 6A, starting with the kernels. For $\hat{\omega}_n^{z_0}$, we denote these by $K_n(t,e^{i\varphi},z_1,\dots,z_n)$; i.e., $\hat{\omega}_n^{z_0}$ is given by

$$\begin{aligned}\hat{\omega}_n^{z_0}(t,e^{i\varphi}) &= T_n^{z_0}(\mu\otimes\cdots\otimes\mu)\\ &:= \int_{\mathbb{C}^n} K_n^{z_0}(t,e^{i\varphi};z_1,\dots,z_n)\,\mu(z_1)\cdots\mu(z_{n+1})\,d^2z_1\cdots d^2z_n.\end{aligned} \tag{7-1}$$

The kernel for $\hat{\omega}_n^a$ has the same geometry and orders, but amplitudes $a(\,\cdot\,)$-averaged in z_0, which does not affect the following analysis.

$K_n^{z_0}$ is a sum of 2^{n-1} terms of the form, for $\vec{\epsilon}\in\{0,1\}^{n-1}$,

$$c_{\vec{\epsilon}}\cdot\frac{\delta^{(n+1-|\vec{\epsilon}|)}\big(t+(-1)^{n+1}2\,\mathrm{Re}\big(e^{i\varphi}\sum_{j=1}^n(-1)^j z_j\big)\big)}{(z_0-z_1)(\bar{z}_1-\bar{z}_2)^{1+\epsilon_1}(\bar{z}_2-\bar{z}_3)^{1+\epsilon_2}\cdots(\bar{z}_{n-1}-\bar{z}_n)^{1+\epsilon_{n-1}}}, \tag{7-2}$$

each with total homogeneity $-(2n+1)$ in $(t,z_0,\dots,z_n)$. These have singularities all in the same locations, namely on a lattice of submanifolds of $\mathbb{R}\times\mathbb{S}^1\times\mathbb{C}^n$. For each $J\in\mathcal{J}=\{J: J\subset\{1,\dots,n-1\}\}$, as in (4-14), let

$$L_n^J := \left\{t+(-1)^{n+1}2\,\mathrm{Re}\left(e^{i\varphi}\sum_{j=1}^n(-1)^j z_j\right)=0 : z_j - z_{j+1}=0 \text{ for all } j\in J\right\}. \tag{7-3}$$

One has $\operatorname{codim}(L_n^J) = 1 + 2|J|$ and $L_n^J \supset L_n^{J'}$ if and only if $J \subset J'$. Rather than using set notation, we sometimes simply list the elements of J. The unique maximal element of the lattice is the hypersurface

$$L_n^{\varnothing} := \left\{ t + (-1)^{n+1} 2 \operatorname{Re}\left(e^{i\varphi} \sum_{j=1}^{n} (-1)^j z_j \right) = 0 \right\},$$

while the unique minimal one is

$$L_n^{12\cdots(n-1)} = \{ t = 0,\ z_1 = z_2 = \cdots = z_n \}.$$

(This notation replaces that used earlier for $n = 1, 2$: what was previously denoted by L_0 is now L_1^{ϕ}, and $L_1 = L_2^{\phi},\ L_3 = L_2^1$.)

As stated above,

$$\operatorname{sing\,supp}(K_n^{z_0}) = \bigcup_{J \in \mathcal{J}} L_n^J$$

and, in fact,

$$\operatorname{WF}(K_n^{z_0}) = \bigcup_{J \in \mathcal{J}} N^* L_n^J, \tag{7-4}$$

with the fact that equality holds (rather than just the $\subset$ containment) following from the nonvanishing in all directions at infinity of the Fourier transforms of $\delta^{(m)}$, $\bar{z}^{-1}$ and $\bar{z}^{-2}$. (However, we only need the containment, not equality, in what follows.)

Define canonical relations

$$C_n^J := N^*(L_n^J)' \subset (T^*(\mathbb{R} \times \mathbb{S}^1) \times T^*\mathbb{C}^n) \setminus 0,$$

sometimes also denoting $C_n^{\varnothing}$ simply by C_n. The *linear* operators $T_n^{z_0} : \mathcal{E}'(\mathbb{C}^n) \to \mathcal{D}'(\mathbb{R} \times \mathbb{S}^1)$ with kernels $K_n^{z_0}$ are (as n varies) interesting prototypes of generalized Fourier integral operators associated with the lattices $\{C_n^J : J \in \mathcal{J}\}$ of canonical relations intersecting cleanly pairwise. There is to our knowledge no general theory of such operators, but in any case, we can describe the wave front relation as follows. Let $\widetilde{\Sigma}^m$ denote the alternating sum

$$\widetilde{\Sigma}^m := z_1 - z_2 + \cdots + (-1)^{m+1} z_m.$$

Definition 7.1. In $T^*(\mathbb{R} \times \mathbb{S}^1) \setminus 0$, define

$$\operatorname{Sc}^{(0)}(\mu) = \left\{ (0, e^{i\varphi}, \tau, 0) : \exists z \in \Omega \text{ such that } (z, \tau e^{-i\varphi}) \in \operatorname{WF}(\mu) \right\} \subset N^*\{t = 0\},$$

and, for $m \geq 1$, let

$$\operatorname{Sc}^{(m)}(\mu) = \Big\{ \big((-1)^{m+1} 2 \operatorname{Re}(e^{i\varphi} \widetilde{\Sigma}^m),\ e^{i\varphi},\ \tau,\ (-1)^m 2\tau \operatorname{Im}(e^{i\varphi} \widetilde{\Sigma}^m) \big) : \exists\, z_1, \ldots, z_m \text{ such that } (z_j, (-1)^{j+1} \tau e^{-i\varphi}) \in \operatorname{WF}(\mu),\ 1 \leq j \leq m \Big\}. \tag{7-5}$$

Definition 7.1 extends the definitions (6-10) for $m = 0, 1$ and (6-9) for $m = 2$. The next theorem extends the WF containments (6-11) for $\hat{\omega}_1,\ \hat{\omega}_2$, to higher n, locating microlocally the singularities of $\hat{\omega}_n$.

Theorem 7.2. *For any conductivity $\sigma \in L^\infty(\Omega)$ and all $n \geq 1$,*

$$\mathrm{WF}(\hat{\omega}_n) \subset \bigcup \{\mathrm{Sc}^{(m)}(\mu) : 0 \leq m \leq n,\ m \equiv n \bmod 2\}. \tag{7-6}$$

Proof. This will follow from (7-1) and the Hörmander–Sato lemma [Hörmander 1971, Theorem 2.5.14]. First, to formulate the n-fold version of (6-7), we introduce the following notation. For sets $A, B \subset T^*\mathbb{C}$ and

$$I \in \mathcal{I} := \{I : I \subset \{1, \dots, n\}\},$$

let

$$\prod_{i\in I} A_i \times \prod_{i'\in I^c} B_{i'} := \{(z,\zeta) \in T^*\mathbb{C}^{n+1} :\ (z_i, \zeta_i) \in A \text{ for all } i \in I,\ (z_{i'}, \zeta_{i'}) \in B, \text{ for all } i' \in I^c\}.$$

For $I \in \mathcal{I}$, if we set

$$\mathrm{WF}^I(\mu) := \prod_{i\in I} \mathrm{WF}(\mu)_i \times \prod_{i'\in I^c} 0_{T^*\mathbb{C},i'}, \tag{7-7}$$

then the analogue of (6-7), which follows from it by induction, is

$$\mathrm{WF}\Big(\bigotimes^n \mu\Big) \subset \bigcup_{I\in\mathcal{I},\, I\neq\phi} \mathrm{WF}^I(\mu). \tag{7-8}$$

Next, for $J \in \mathcal{J}$, define

$$\bar{J} := \{i \in \{1, \dots, n\} :\ i \in J \text{ or } i-1 \in J\} \in \mathcal{I}.$$

Then, $|\bar{J}|$ is even, and thus

$$|\bar{J}^c| = |\{1, \dots, n\} \setminus \bar{J}| \equiv n \bmod 2.$$

We can partition $\bar{J} = \bar{J}_+ \cup \bar{J}_- \cup \bar{J}_\pm$, where

$$\begin{aligned} \bar{J}_+ &:= \{i \in \bar{J} : i \in J,\ i-1 \notin J\},\\ \bar{J}_- &:= \{i \in \bar{J} : i-1 \in J,\ i \notin J\},\\ \bar{J}_\pm &:= \{i \in \bar{J} : i-1 \in J,\ i \in J\}. \end{aligned} \tag{7-9}$$

The submanifold $L_n^J \subset \mathbb{R} \times \mathbb{S}^1 \times \mathbb{C}^n$ is given by defining functions $f_0,\ \{f_j\}_{j\in J}$, where

$$\begin{aligned} f_0(t,\varphi,z) &= t + (-1)^{n+1} 2\,\mathrm{Re}\Big(e^{i\varphi} \sum_{i=1}^n (-1)^i z_i\Big),\\ f_j(t,\varphi,z) &= z_j - z_{j+1}, \quad j \in J. \end{aligned}$$

The twisted conormal bundles are parametrized by

$$C_n^J = \left\{\left(t, \varphi, \tau d_{t,\varphi} f_0; z, -\left(\tau d_z f_0 + \sum_{j\in J} \sigma_j \cdot d_z f_j\right)\right) : (t, e^{i\varphi}, z) \in L_n^J,\ (\tau,\sigma) \in (\mathbb{R} \times \mathbb{C}^{|J|}) \setminus 0\right\}.$$

The twisted gradients $df' := (d_{t,\varphi} f, -d_z f)$ of the defining functions are

$$df_0' = \left(1, (-1)^{n+1} 2\,\mathrm{Im}\Big(e^{i\varphi} \sum_{i=1}^n (-1)^i z_i\Big), (-1)^n 2E(\varphi)\right),$$

with $E(\varphi) = (e^{-i\varphi}, -e^{-i\varphi}, e^{-i\varphi}, \dots, (-1)^n e^{-i\varphi})$, where we identify $\pm e^{-i\varphi} \in \mathbb{C}$ with a real covector $(\xi_i, \eta_i) \in T^*\mathbb{C}$, and

$$df_j' = -\sigma_j \cdot dz_j + \sigma_j \cdot dz_{j+1}, \quad j \in J,$$

similarly identifying $\sigma_j \in \mathbb{C}$ with $(\operatorname{Re}\sigma_j, \operatorname{Im}\sigma_j) \in T^*\mathbb{C}$. Thus,

$$\begin{aligned} C_n^J = \Bigg\{ \Bigg((-1)^n 2 \operatorname{Re}\Bigg(e^{i\varphi} \sum_{i=1}^n (-1)^i z_i \Bigg), e^{i\varphi}, \tau, (-1)^{n+1} 2\tau \operatorname{Im}\Bigg(e^{i\varphi} \sum_{i=1}^n (-1)^i z_i \Bigg); \\ z, (-1)^n 2\tau E(\varphi) + \sum_{i \in \bar{J}_+} \sigma_i \cdot dz_i - \sum_{i \in \bar{J}_-} \sigma_i \cdot dz_i + \sum_{i \in \bar{J}_\pm} (-\sigma_{i-1} + \sigma_i) \cdot dz_i \Bigg) \\ : e^{i\varphi} \in \mathbb{S}^1,\ z_j - z_{j+1} = 0,\ j \in J,\ (\tau, \sigma) \in (\mathbb{R} \times \mathbb{C}^{|J|}) \setminus 0 \Bigg\}. \end{aligned} \tag{7-10}$$

Since $\operatorname{WF}(K_n^{z_0})' = \bigcup_{J \in \mathcal{J}} C_n^J$, to prove (7-6), it suffices to show that each of the $2^{n-1}(2^n - 1)$ compositions $C_n^J \circ \operatorname{WF}^I$, $J \in \mathcal{J}$, $I \in \mathcal{I} \setminus \{\varnothing\}$, is contained in one of the $\operatorname{Sc}^{(m)}(\mu)$ for some $0 \le m \le n$ with $m \equiv n \mod 2$. In fact, from (7-7) and the representation of C_n^J above, one sees that each $C_n^J \circ \operatorname{WF}^I$ is either empty (e.g., if $\bar{J}^c \cap I^c \neq \phi$), or a (potentially) nonempty subset of $\operatorname{Sc}^{(m)}(\mu)$, when $m = |\bar{J}^c| \equiv n \mod 2$, yielding (7-6) and finishing the proof of Theorem 7.2. □

8. Parity symmetry

We now come to an important symmetry property which significantly improves the imaging obtained via our reconstruction method. Recall that what we have been denoting by $\hat{\omega}$ is in fact $\hat{\omega}^+$, the partial Fourier transform of the correction term ω^+ in the CGO solution (3-6) of the Beltrami equation (3-4) with multiplier μ. Similarly, the solution ω^- in (3-6) corresponding to $-\mu$ has partial Fourier transform $\hat{\omega}^-$. Astala and Päivärinta [2006a] showed that both ω^+ and ω^- can be reconstructed from the Dirichlet-to-Neumann map Λ_σ. We show that by taking their difference we can suppress the $\hat{\omega}_n$ for *even* n, and thus suppress some of the singularities described in the preceding sections, most importantly the strong singularity at $\operatorname{Sc}^{(0)}(\mu) \subset N^*\{t = 0\}$ coming from $\hat{\omega}_2$.

Start by writing the two Neumann series

$$\hat{\omega}^+ \sim \sum_{n=1}^{+\infty} \hat{\omega}_n^+ = \hat{\omega}_{\text{odd}}^+ + \hat{\omega}_{\text{even}}^+, \qquad \hat{\omega}^- \sim \sum_{n=1}^{+\infty} \hat{\omega}_n^- = \hat{\omega}_{\text{odd}}^- + \hat{\omega}_{\text{even}}^-,$$

where $\hat{\omega}_{\text{odd}}^\pm$ (resp. $\hat{\omega}_{\text{even}}^\pm$) consists of the n odd (resp. even) terms in the expansion corresponding to $\hat{\omega}^\pm$. Recall that, as a function of μ, $\hat{\omega}_n^\pm$ is a multilinear form of degree n.

Proposition 8.1. *Each of $\hat{\omega}_{\text{odd}}^+$ and $\hat{\omega}_{\text{even}}^+$ has the same parity in t as the multilinear degrees of its terms*; *i.e.*,

$$\hat{\omega}_{\text{odd}}^+ = -\hat{\omega}_{\text{odd}}^- \quad \textit{and} \quad \hat{\omega}_{\text{even}}^+ = \hat{\omega}_{\text{even}}^-. \tag{8-1}$$

Equivalently,

$$\hat{\omega}_{\text{odd}}^+ = \frac{\hat{\omega}^+ - \hat{\omega}^-}{2} \quad \textit{and} \quad \hat{\omega}_{\text{even}}^+ = \frac{\hat{\omega}^+ + \hat{\omega}^-}{2}. \tag{8-2}$$

Proof. Let $\bar{u}^{\pm} = -\bar{\partial}\omega^{\pm}$. As in Section 3, $u^{\pm}$ is the solution of the integral equation (3-14),

$$(I + A^{\pm}\rho)u^{\pm} = \mp\bar{\alpha}, \tag{8-3}$$

where $A^{\pm} = \mp(\bar{\alpha}P + \bar{\nu}S)$, and α and ν were defined in (3-8). Since $A^{+} = -A^{-}$ we have $u_1^{+} = -\bar{\alpha} = -u_1^{-}$, $u_2^{+} = -A^{+}\bar{u}_1^{+} = -(-A^{-}(-\bar{u}_1^{-})) = u_2^{-}$ and by induction, for $n \geq 1$,

$$u_{n+2}^{+} = A^{+}\overline{A^{+}\bar{u}_n^{+}} = (-1)^n A^{-}\overline{A^{-}\bar{u}_n^{-}} = (-1)^n u_{n+2}^{-}.$$

(Another way of seeing this is that $\mu \to \hat{\omega}_n$ is a form of degree n, with the same multilinear kernel applied to both $\pm\mu$.) □

Proposition 8.1 provides a method to isolate the even and the odd terms in the expansion of $\hat{\omega}$. In particular, by imaging using $\hat{\omega}_{\text{odd}}^{+}$, we can eliminate the strong singularities of $\hat{\omega}_2$ at $\text{Sc}^{(0)}(\mu) = N^*\{t = 0\}$, described in (6-13), and in fact the singularities there of all the even terms since, by (7-6), these only arise from $\hat{\omega}_n$ for even n.

9. Multilinear operator theory

Following the analysis of $\hat{\omega}_2$, one can also describe the singularities of $\hat{\omega}_3$, but now having to restrict away from $t = 0$. The singularities of $\hat{\omega}_3$ are of interest, since, after the symmetrization considerations from the previous section are applied, $\hat{\omega}_3$ is the first higher-order term encountered after $\hat{\omega}_1$. Recall from above that, if μ is a piecewise smooth function with jumps ($m = -1$), $\hat{\omega}_2$ has a singularity at $\text{Sc}^{(0)}(\mu) = N^*\{t = 0\}$ as strong as that of $\hat{\omega}_1$ at $\text{Sc}^{(1)}(\mu)$, and that its presence is due to the singularity of $K_2^{z_0}$ at the submanifold $L_2^1 = \{t = 0\} \subset L_2^{\varnothing} \subset \mathbb{R} \times \mathbb{S}^1 \times \mathbb{C}^2$. Similarly, in order to analyze $\hat{\omega}_3$, we will need to localize $K_3^{z_0}$ away from $L_3^{12} = \{t = 0\} \subset \mathbb{R} \times \mathbb{S}^1 \times \mathbb{C}^3$, which results in a kernel that can then be decomposed into a sum of two kernels, each having singularities on one of two nested pairs, $L_3^1 \subset L_3^{\varnothing}$ or $L_3^2 \subset L_3^{\varnothing}$, but not at $L_3^1 \cap L_3^2 = L_3^{12} = \{t = 0\}$. We will show that applying these to $\mu \otimes \mu \otimes \mu$, as in (7-1), does not just result in terms with WF contained in $\text{Sc}^{(3)}(\mu) \cup \text{Sc}^{(1)}(\mu)$, as was shown in Theorem 7.2, but a more precise statement can be made:

Theorem 9.1. *If $\mu \in I^m(\gamma)$ with γ satisfying the curvature condition (6-12), then $\text{Sc}^{(3)}(\mu)$, defined as in (6-9), is a smooth Lagrangian manifold in $T^*(\mathbb{R} \times \mathbb{S}^1) \setminus 0$, and*

$$\hat{\omega}_3|_{t\neq 0} \in I^{3m+2,-1/2}(\text{Sc}^{(3)}(\mu), \text{Sc}^{(1)}(\mu)). \tag{9-1}$$

Remark. For $m = -1$, this is in $I^{-1}(\text{Sc}^{(1)}(\mu) \setminus \text{Sc}^{(3)}(\mu))$, and thus is half a derivative smoother than $\hat{\omega}_1$ on $\text{Sc}^{(1)}(\mu)$. On the other hand, it is also in $I^{-3/2}(\text{Sc}^{(3)}(\mu) \setminus \text{Sc}^{(1)}(\mu))$, which is a full derivative smoother than $\hat{\omega}_1$.

To put this in perspective we first discuss what should be the leading terms contributing to $\hat{\omega}_n$ for general $n \geq 3$. The analysis for $\hat{\omega}_3|_{t\neq 0}$ given below applies more generally to $\hat{\omega}_n$ if we localize $K_n^{z_0}$ even more strongly: not just away from $t = 0$, but away from *all* of the submanifolds $L_n^J \subset \mathbb{R} \times \mathbb{S}^1 \times \mathbb{C}^n$ with $|J| \geq 2$. Now, for $j \neq j'$, we have $L_n^j \cap L_n^{j'} = L_n^{jj'}$; by localizing away from all of the L_n^J with $|J| = 2$, by a partition of unity the kernel $K_n^{z_0}$ can be decomposed into a sum of $n - 1$ terms, each a nested conormal

distribution associated with the pair $L_n^\phi \supset L_n^j$, $j = 1, \dots, n-1$ respectively. When these pieces of $K_n^{z_0}$ are applied to $\bigotimes^n \mu$, as in (7-1), the results have WF in $\mathrm{Sc}^{(n)}(\mu) \cup \mathrm{Sc}^{(n-2)}(\mu)$, and again can be shown to belong to $I^{p,l}(\mathrm{Sc}^{(n)}(\mu), \mathrm{Sc}^{(n-2)}(\mu))$. However, as this requires localizing away from $\bigcup_{|J|\geq 2} L_n^J$, which is strictly larger than $L_n^{12\cdots(n-1)}$ if $n \geq 4$, the analysis here is inconclusive concerning the singularities of $\hat{\omega}_n|_{t\neq 0}$, and thus we only present the details for $\hat{\omega}_3$.

We now start the proof of Theorem 9.1 by noting that, for $n = 3$, the lattice of submanifolds (7-3) to which the trilinear operator $T_3^{z_0}$ is associated is a simple diamond, $L_3^\varnothing \supset L_3^1$, $L_3^2 \supset L_3^{12}$. In the region $\{t \neq 0\}$, the two submanifolds L_3^1 and L_3^2 are disjoint. Hence, by a partition of unity in the spatial variables, we can write

$$\hat{\omega}_3|_{t\neq 0} = \langle K_3^1 + K_3^2, \mu \otimes \mu \otimes \mu \rangle, \tag{9-2}$$

where each K_3^j is associated with the nested pair $L_3^\varnothing \supset L_3^j$, $j = 1, 2$. Since these two terms are so similar, we just treat the K_3^2 term.

The submanifolds $L_2^2 \subset L_3^\phi \subset \mathbb{R} \times \mathbb{S}^1 \times \mathbb{C}^3$ are given by

$$\begin{aligned} L_3^\varnothing &= \{t - 2\,\mathrm{Re}(e^{i\varphi}(z_1 - z_2 + z_3)) = 0\}, \\ L_3^2 &= \{t - 2\,\mathrm{Re}(e^{i\varphi}(z_1 - z_2 + z_3)) = 0,\ z_2 - z_3 = 0\}. \end{aligned} \tag{9-3}$$

For K_3^2 we are localizing away from L_3^1, so that $z_1 - z_2 \neq 0$ on the support of the kernels below. Thus, the factors $(\bar{z}_1 - \bar{z}_2)^{-1+\epsilon_1}$ in (7-2) are smooth, and their dependence on ϵ_1 is irrelevant for this analysis. Thus, K_3^2 is a sum of two terms, each of which we will still denote K_3^2, given by

$$K_3^2 = \int_{\mathbb{R}^3} e^{i[\tau(t - 2\,\mathrm{Re}(e^{i\varphi}(z_1 - z_2 + z_3))) + (z_2 - z_3)\cdot\sigma]} a_{p,l}(*; \tau; \sigma)\, d\tau\, d\sigma, \tag{9-4}$$

where $*$ denotes the spatial variables and $a_{p,l}$ is a symbol-valued symbol of biorder $(3, -1)$ and $(2, 0)$, respectively.

If, for any $c > 0$, we introduce a smooth cutoff into the amplitude which is a function of $|\sigma|/|\tau|$ and supported in the region $\{|\sigma| \geq c|\tau|\}$, the amplitude becomes a standard symbol of order $p + l = 2$ in the phase variables $(\tau, \sigma) \in \mathbb{R}^3 \setminus 0$. The phase function is nondegenerate and parametrizes the canonical relation, with C_0 as in (5-9),

$$\begin{aligned} C_{0\times N} &:= C_0 \times N^*\{z_2 = z_3\} \\ &= \big\{\big(2\,\mathrm{Re}(e^{i\varphi} z_1), e^{i\varphi}, \tau, 2\tau\,\mathrm{Im}(e^{i\varphi} z_1); z_1, z_2, z_2, 2\tau e^{-i\varphi}, \zeta_2, -\zeta_2\big) \\ &\qquad : e^{i\varphi} \in \mathbb{S}^1, (z_1, z_2) \in \mathbb{C}^2, (\tau, \zeta_2) \in \mathbb{R}^3 \setminus 0\big\}. \end{aligned}$$

This is a nondegenerate canonical relation: the projection $\pi_R : C_{0\times N} \to T^*\mathbb{C}^3 \setminus 0$ is an immersion and the projection $\pi_L : C_{0\times N} \to T^*(\mathbb{R} \times \mathbb{S}^1) \setminus 0$ is a submersion. Thus, this contribution to K_3^2 belongs to $I^{2+3/2-8/4}(C_{0\times N}) = I^{3/2}(C_{0\times N})$. Due to the support of the amplitude of this term, $\pi_R(C_{0\times N}) \subset \{|\zeta_1| \sim |\zeta_2| = |\zeta_3|\}$, and by reasoning similar to that used in the analysis of $\hat{\omega}_1$, one concludes that

$\mu\otimes\mu\otimes\mu\in I^{3m}(N^*(\gamma\times\gamma\times\gamma))$ microlocally on this region. Hence, the composition

$$C_{0\times N}\circ N^*(\gamma\times\gamma\times\gamma)\subset C_0\circ N^*\gamma =: \mathrm{Sc}^{(1)}(\mu)$$

is covered by the transverse intersection calculus, and this contribution to $\hat\omega_3$ belongs to

$$I^{3m+3/2}(\mathrm{Sc}^{(1)}(\mu)). \tag{9-5}$$

Now consider the contribution to (9-4) from the region $\{|\sigma|\le\frac12|\tau|\}$. Writing out the representations of each of the three μ factors in (9-2) as conormal distributions, we first note that, using the parametrization in (7-10) for C_3^2 and the constraint $|\sigma|\le\frac12|\tau|$, we can read off that, on π_R of the wave front relation,

$$|\zeta_1|=2|\tau| \qquad\text{and}\qquad |\zeta_j|=|\pm(\sigma-2\tau e^{-i\varphi})|\ge\tfrac32|\tau|,\quad j=2,3.$$

Hence, again we are acting on a part of $\mu\otimes\mu\otimes\mu$ which is microlocalized where $|\zeta_1|\sim|\zeta_2|\sim|\zeta_3|$. As a result, in (9-6) below, the θ_j are grouped with τ as "elliptic" variables for the symbol-valued symbol estimates. Mimicking the analysis in and following (6-16), homogenize z_1,z_2,z_3 by setting $\eta_j=\tau z_j$, $j=1,2,3$. This leads to the expression

$$\int e^{i\widetilde\Psi}\,a_{\tilde p,\tilde l}(*;(\tau,\theta_1,\theta_2,\theta_3,\eta_1,\eta_2,\eta_3);\sigma)\,d\tau\,d\theta_1\,d\theta_2\,d\theta_3\,d\eta_1\,d\eta_2\,d\eta_3\,d\sigma, \tag{9-6}$$

with phase

$$\begin{aligned}\widetilde\Psi&=\widetilde\Psi(t,e^{i\varphi};\ \tau,\theta_1,\theta_2,\theta_3,\eta_1,\eta_2,\eta_3;\ \sigma)\\ &:=\tau t-2\,\mathrm{Re}(e^{i\varphi}(\eta_1-\eta_2+\eta_3))+\theta_1 g\left(\frac{\eta_1}{\tau}\right)+\theta_2 g\left(\frac{\eta_2}{\tau}\right)+\theta_3 g\left(\frac{\eta_3}{\tau}\right)+\sigma\cdot\left(\frac{\eta_2-\eta_3}{\tau}\right)\end{aligned}$$

on $(\mathbb{R}\times\mathbb{S}^1)\times(\mathbb{R}^{10}_{\tau,\theta_1,\theta_2,\theta_3,\eta_1,\eta_2,\eta_3}\setminus 0)\times\mathbb{R}^2_\sigma$ and symbol-valued symbols with biorders $(\tilde p,\tilde l)=(3m-3,-1)$ and $(3m-4,0)$, respectively. As with the phase $\widetilde\Phi$ that arose in the analysis of $\hat\omega_1$, $\widetilde\Psi$ is a multiphase function: $\widetilde\Psi_0=\widetilde\Psi|_{\sigma=0}$ is nondegenerate (excess $e_0=0$) and parametrizes $\mathrm{Sc}^{(3)}(\mu)$, while the full $\widetilde\Psi$ is clean (excess $e_1=1$) and parametrizes $\mathrm{Sc}^{(1)}(\mu)$. Applying Proposition 6.2, with $N=10$, $M=2$, the terms in (9-6) with amplitudes of biorders $(3m-3,-1)$ and $(3m-4,0)$, respectively, yield elements of $I^{3m+2,-1/2}(\mathrm{Sc}^{(3)}(\mu),\mathrm{Sc}^{(1)}(\mu))$ and $I^{3m+2,-3/2}(\mathrm{Sc}^{(3)}(\mu),\mathrm{Sc}^{(1)}(\mu))$, respectively; since the former space contains the latter, and furthermore contains the space in (9-5), we conclude that $\hat\omega_3|_{t\ne0}\in I^{3m+2,-1/2}(\mathrm{Sc}^{(3)}(\mu),\mathrm{Sc}^{(1)}(\mu))$. This finishes the proof of Theorem 9.1. □

10. Computational studies

In the idealized infinite-bandwidth model discussed above, knowledge of $\omega_1(z_0,k)$ for all complex frequencies k, and thus $T_1^{z_0}\mu=\hat\omega(z_0,t,e^{i\varphi})$ for all $(t,\ e^{i\varphi})$, determines μ mod C^∞. A more physically realistic model, band-limiting to $|k|\le k_{\max}$, requires a windowed Fourier transform; see [Isaacson et al. 2004; 2006; Knudsen 2003; Knudsen et al. 2004; 2007; Vainikko 2000]. This corresponds to convolving in the t-variable with a smooth cutoff at length-scale $\sim k_{\max}^{-1}$, rendering the reconstruction less accurate. This section examines numerical simulations and how they are affected by this bandwidth issue.

We first introduce a new reconstruction algorithm from the Dirichlet-to-Neumann map Λ_σ, as well as the algorithm used in the simulations. Then we will present our numerical results. In this section we take Ω to be the unit disk, $\Omega = D(0, 1)$.

10A. ***Reconstruction algorithm.*** The results presented in the preceding sections give rise to a linear reconstruction scheme to approximately recover a conductivity σ from its Dirichlet-to-Neumann map Λ_σ. This can be summarized in the following steps:

(i) Find $f_{\pm\mu}(z, k)$, and so $\omega^{\pm}(z, k)$, for $z \in \partial\Omega$ and $k \in \mathbb{C}$, by solving the boundary integral equation

$$f_{\pm\mu}(z, k) + e^{ikz} = (\mathcal{P}_{\pm\mu} + \mathcal{P}_0^k) f_{\pm\mu}(z, k), \quad z \in \partial\Omega, \tag{10-1}$$

where $\mathcal{P}_{\pm\mu}$ and $\mathcal{P}_0^k$ are projection operators constructed from Λ_σ. See [Astala and Päivärinta 2006a] and [Mueller and Siltanen 2012, Section 16.3.3] for full details.

(ii) Write $k = \tau e^{i\varphi}$. Apply the one-dimensional Fourier transform $\mathcal{F}_{\tau\mapsto t}$ and the complex average (5-15) in order to obtain $\hat{\omega}^{a,\pm}(t, e^{i\varphi})$, with $a \equiv 1/\sqrt{2}$.

(iii) Taking into account the parity result Proposition 8.1, define $\hat{\omega}^a_{\text{diff}} := \frac{1}{2}(\hat{\omega}^{a,+} - \hat{\omega}^{a,-})$. Apply either the exact inversion formula (5-18) or the Λ-tomography analogue (5-17) with $\hat{\omega}^a_{\text{diff}}$ instead of $\hat{\omega}^a_1$, in order to obtain an approximation μ_{appr} to μ.

(iv) The approximate conductivity is found with the identity $\sigma_{\text{appr}} = (1 - \mu_{\text{appr}})/(1 + \mu_{\text{appr}})$.

10B. ***High-precision data assumption.*** In the numerical reconstructions presented below, the spectral parameter k ranges in the disk $\{|k| < R\}$ with cutoff frequency $R = 60$. Such a large radius R is needed for demonstrating the crucial properties of the new method; with a smaller radius the windowing of the Fourier transform would smooth out important features in the CGO solutions.

Using such a large R in practice would require very high precision EIT measurements, which cannot be achieved by current technology. However, it is possible to evaluate the needed CGO solutions computationally when σ is known. (Remark: it is possible to compute useful reconstructions from real EIT measurements using the new method combined with sparsity-promoting inversion algorithms, but we do not discuss such approaches further in this paper.) This is done as in [Astala et al. 2014] by solving the Beltrami equation

$$\bar{\partial}_z f_\mu(z, k) = \mu(z)\, \overline{\partial_z f_\mu(z, k)}, \tag{10-2}$$

which yields very accurate solutions even for large $|k|$. From the point of view of the classical $\bar{\partial}$ reconstruction method [Knudsen et al. 2009; Mueller and Siltanen 2003; 2012; Siltanen et al. 2000] for C^2 conductivities, this is the analogue of solving the Lippmann–Schwinger equation to construct the CGO solutions.

In this section the CGO remainders $\omega^{\pm}(z, k)$, with $z \in \partial\Omega$ and $|k| < 60$, are constructed by solving the Beltrami equation following the approach of Huhtanen and Perämäki [2012] (see also [Astala et al. 2014] and Section 3 for more details). We then follow steps (ii)–(iv) of the algorithm in Section 10A to obtain two-dimensional reconstructions.

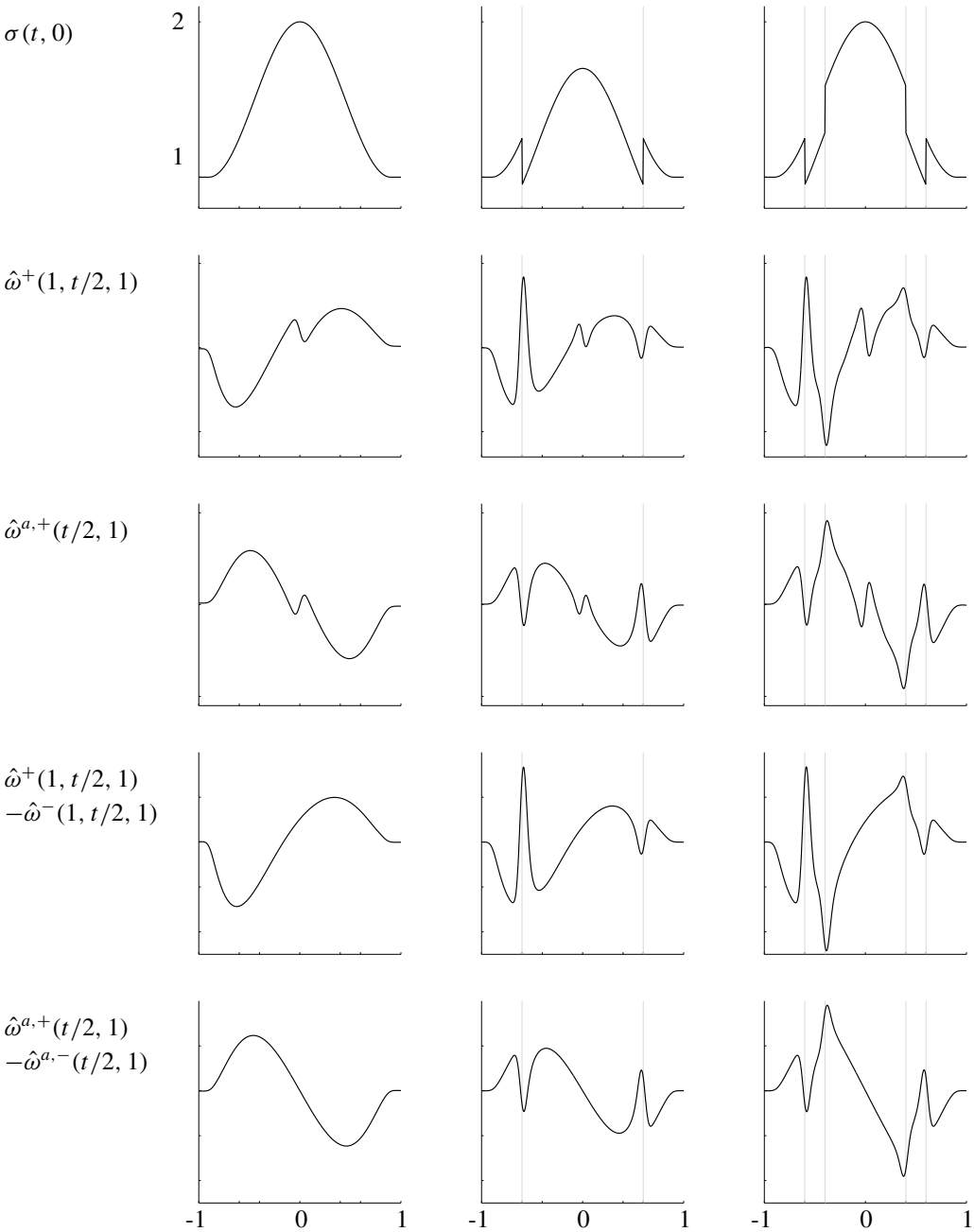

Figure 6. Top: profiles of three radial conductivities along the real axis. The middle conductivity has a jump along the circle $|z| = 0.6$; the one on the right has jumps on both $|z| = 0.4$ and $|z| = 0.6$. Rows 2 and 3: the functions $\hat{\omega}^+(1, t/2, 1)$ and $\hat{\omega}^{a,+}(t/2, 1)$, respectively; note the artifacts at $t = 0$. Rows 3 and 4: as described in Section 8, the artifacts are eliminated by subtracting $\hat{\omega}^-$, $\hat{\omega}^{a,-}$, respectively.

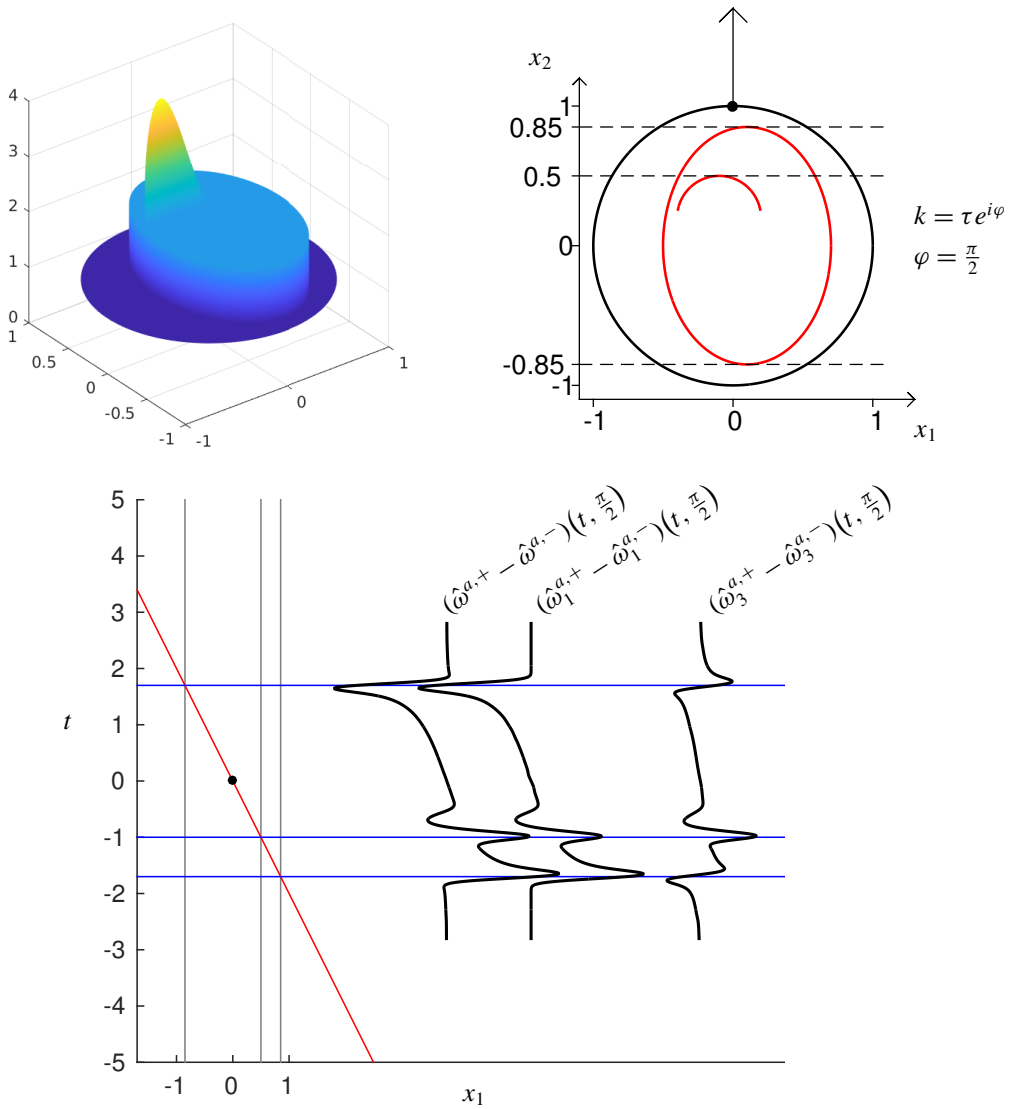

Figure 7. Diagram showing the propagation of singularities for the HME phantom with zero background. The virtual direction is $k = i$.

10C. ***Rotationally symmetric cases.*** We study three rotationally symmetric conductivities defined in the unit disc. The first conductivity σ_1 is smooth. The second conductivity is defined as

$$\sigma_2 = \sigma_1 - 0.3\chi_{D(0,0.6)}$$

and therefore has a jump of magnitude 0.3 along the circle centered at the origin and radius 0.6. The third rotationally symmetric conductivity is defined as

$$\sigma_3 = \sigma_2 + 0.3\chi_{D(0,0.4)}$$

and has jumps of magnitude 0.3 along the circles centered at the origin and radii 0.4 and 0.6.

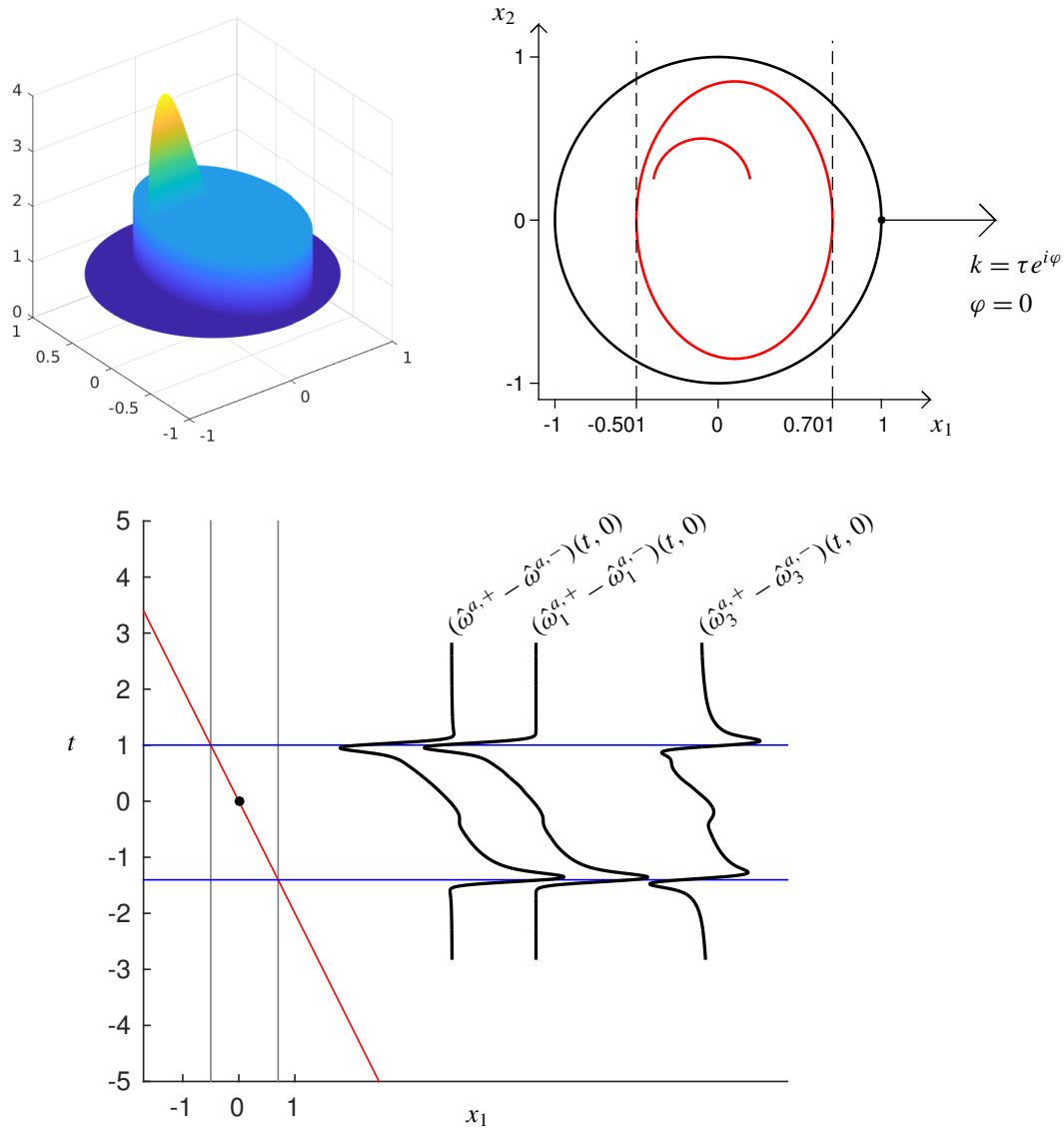

Figure 8. Diagram showing the propagation of singularities for the HME phantom with zero background. The virtual direction is $k = 1$.

In Figure 6 we show the profiles of $\hat{\omega}(1, t, 1)$ for three rotationally symmetric conductivity phantoms. The first phantom is smooth, while the second and the third have jumps. The position and the sign of each jump is clearly visible from the CGO-Fourier data. Note that the artifact singularity appearing around 0 in the second and third rows vanishes when considering the difference of the two CGO functions, in the fourth and fifth rows. This confirms the parity symmetry analyzed in Section 8.

10D. ***Half-moon and ellipse*** **(*HME*).** This conductivity phantom has a large elliptical inclusion and another smaller inclusion inside the ellipse. The smaller inclusion has a jump along an almost complete

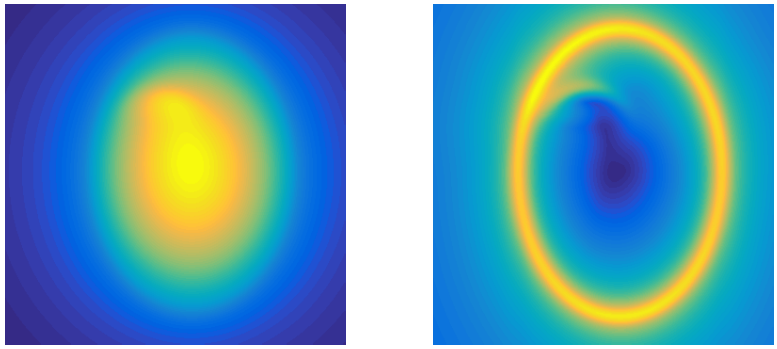

Figure 9. Reconstructions from the averaged full series $\hat{\omega}^a_{\text{diff}}$. Left: exact inversion formula. Right: Λ-tomography like reconstruction.

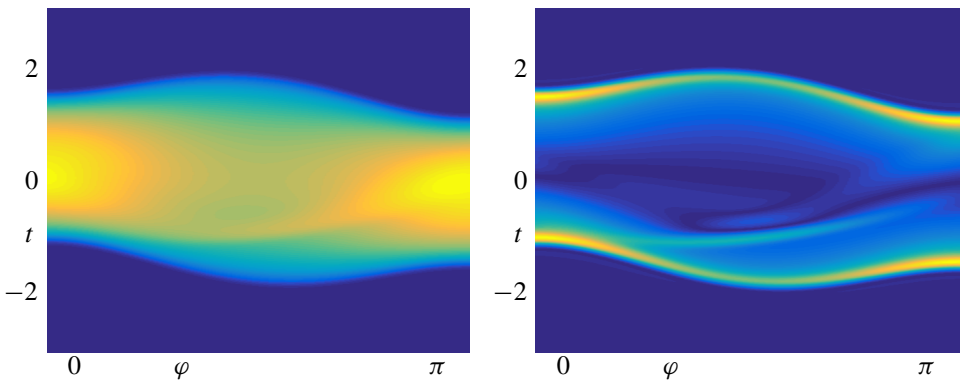

Figure 10. Sinograms of the averaged full series $\hat{\omega}^a_{\text{diff}}$. Left: exact reconstruction sinogram. Right: Λ-tomography like sinogram.

half-circle. This example was chosen because it has two nontrivial features in the wave front set for the horizontal direction and three for the vertical. Figures 8 and 7 show, in particular, *ladder* diagrams of the propagation of singularities in the directions $k = i$ and $k = 1$, respectively: the zeroth- and second-order terms of the Neumann series for $\hat{\omega}^a$ are displayed, as well as the full series of the difference of the CGOs: $\hat{\omega}_{\text{diff}} = (\hat{\omega}^+ - \hat{\omega}^-)/2$.

Figure 9 shows two-dimensional reconstructions obtained using the new algorithm, with the two different inversion formulas. In Figure 10 we show the values of $\hat{\omega}^a_{\text{diff}}(t, e^{i\varphi})$ for $t \in [-3, 3]$ and $\varphi \in [0, \pi]$. We borrow the term *sinogram* to describe these plots, because of the clear similarity with the sinograms of X-ray tomography.

11. Conclusion

We introduce a novel and robust method for recovering singularities of conductivities from electric boundary measurements. It is unique in its capability of recovering inclusions within inclusions in an unknown inhomogeneous background conductivity. This method provides a new connection between

diffuse tomography (EIT) and classical parallel-beam X-ray tomography and filtered back-projection algorithms.

Full analysis of the higher-order terms $\hat{\omega}_n$ remains an open problem. We point out that there is a strong formal similarity between the multilinear forms $\mu \to \hat{\omega}_n$ and multilinear operators considered by Brown [2001], Nie and Brown [2011] and Perry and Christ [Perry 2016]. Indeed, any Born-type expansion naturally leads to expressions of this general form, with the places of the Cauchy and Beurling kernels for ω_n or $\hat{\omega}_n$ here being taken by the appropriate Green's functions. However, an important feature here is that the singular coefficient in a Beltrami equation occurs in the top-order term, rather than as a potential as in the works cited above. For the application needed in this setting, useful function space estimates do not seem to follow from existing results, which would require higher regularity of μ, and this is an interesting topic for future investigation.

References

[Alessandrini 1988] G. Alessandrini, "Stable determination of conductivity by boundary measurements", *Appl. Anal.* **27**:1-3 (1988), 153–172. MR Zbl

[Alessandrini and Di Cristo 2005] G. Alessandrini and M. Di Cristo, "Stable determination of an inclusion by boundary measurements", *SIAM J. Math. Anal.* **37**:1 (2005), 200–217. MR Zbl

[Alessandrini and Scapin 2017] G. Alessandrini and A. Scapin, "Depth dependent resolution in electrical impedance tomography", *J. Inverse Ill-Posed Probl.* **25**:3 (2017), 391–402. MR Zbl

[Assenheimer et al. 2001] M. Assenheimer, O. Laver-Moskovitz, D. Malonek, D. Manor, U. Nahaliel, R. Nitzan, and A. Saad, "The T-SCAN technology: electrical impedance as a diagnostic tool for breast cancer detection", *Physiolog. Meas.* **22**:1 (2001), 1–8.

[Astala and Päivärinta 2006a] K. Astala and L. Päivärinta, "A boundary integral equation for Calderón's inverse conductivity problem", *Collect. Math.* Spec. Iss. (2006), 127–139. MR Zbl

[Astala and Päivärinta 2006b] K. Astala and L. Päivärinta, "Calderón's inverse conductivity problem in the plane", *Ann. of Math.* (2) **163**:1 (2006), 265–299. MR Zbl

[Astala et al. 2009] K. Astala, T. Iwaniec, and G. Martin, *Elliptic partial differential equations and quasiconformal mappings in the plane*, Princeton Mathematical Series **48**, Princeton Univ. Press, 2009. MR Zbl

[Astala et al. 2010] K. Astala, J. L. Mueller, L. Päivärinta, and S. Siltanen, "Numerical computation of complex geometrical optics solutions to the conductivity equation", *Appl. Comput. Harmon. Anal.* **29**:1 (2010), 2–17. MR Zbl

[Astala et al. 2011] K. Astala, J. L. Mueller, L. Päivärinta, A. Perämäki, and S. Siltanen, "Direct electrical impedance tomography for nonsmooth conductivities", *Inverse Probl. Imaging* **5**:3 (2011), 531–549. MR Zbl

[Astala et al. 2014] K. Astala, L. Päivärinta, J. M. Reyes, and S. Siltanen, "Nonlinear Fourier analysis for discontinuous conductivities: computational results", *J. Comput. Phys.* **276** (2014), 74–91. MR Zbl

[Astala et al. 2016] K. Astala, M. Lassas, and L. Päivärinta, "The borderlines of invisibility and visibility in Calderón's inverse problem", *Anal. PDE* **9**:1 (2016), 43–98. MR Zbl

[Brown 2001] R. M. Brown, "Estimates for the scattering map associated with a two-dimensional first-order system", *J. Nonlinear Sci.* **11**:6 (2001), 459–471. MR Zbl

[Brown and Uhlmann 1997] R. M. Brown and G. A. Uhlmann, "Uniqueness in the inverse conductivity problem for nonsmooth conductivities in two dimensions", *Comm. Partial Differential Equations* **22**:5-6 (1997), 1009–1027. MR Zbl

[Brühl and Hanke 2000] M. Brühl and M. Hanke, "Numerical implementation of two noniterative methods for locating inclusions by impedance tomography", *Inverse Problems* **16**:4 (2000), 1029–1042. MR Zbl

[Calderón 1980] A.-P. Calderón, "On an inverse boundary value problem", pp. 65–73 in *Seminar on Numerical Analysis and its Applications to Continuum Physics* (Rio de Janeiro, 1980), Soc. Brasil. Mat., Rio de Janeiro, 1980. MR Zbl

[Candès and Fernandez-Granda 2013] E. J. Candès and C. Fernandez-Granda, "Super-resolution from noisy data", *J. Fourier Anal. Appl.* **19**:6 (2013), 1229–1254. MR Zbl

[Candès and Fernandez-Granda 2014] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution", *Comm. Pure Appl. Math.* **67**:6 (2014), 906–956. MR Zbl

[Caro and Rogers 2016] P. Caro and K. M. Rogers, "Global uniqueness for the Calderón problem with Lipschitz conductivities", *Forum Math. Pi* **4** (2016), art. id. e2. MR Zbl

[Chan and Tai 2004] T. F. Chan and X.-C. Tai, "Level set and total variation regularization for elliptic inverse problems with discontinuous coefficients", *J. Comput. Phys.* **193**:1 (2004), 40–66. MR Zbl

[Chen et al. 2001] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit", *SIAM Rev.* **43**:1 (2001), 129–159. MR Zbl

[Cheney and Isaacson 1992] M. Cheney and D. Isaacson, "Distinguishability in impedance imaging", *IEEE Trans. Biomed. Eng.* **39**:8 (1992), 852–860.

[Cheney et al. 1999] M. Cheney, D. Isaacson, and J. C. Newell, "Electrical impedance tomography", *SIAM Rev.* **41**:1 (1999), 85–101. MR Zbl

[Chung et al. 2005] E. T. Chung, T. F. Chan, and X.-C. Tai, "Electrical impedance tomography using level set representation and total variational regularization", *J. Comput. Phys.* **205**:1 (2005), 357–372. MR Zbl

[Dobson and Santosa 1994] D. C. Dobson and F. Santosa, "An image-enhancement technique for electrical impedance tomography", *Inverse Problems* **10**:2 (1994), 317–334. MR Zbl

[van den Doel and Ascher 2006] K. van den Doel and U. M. Ascher, "On level set regularization for highly ill-posed distributed parameter estimation problems", *J. Comput. Phys.* **216**:2 (2006), 707–723. MR Zbl

[Duistermaat and Hörmander 1972] J. J. Duistermaat and L. Hörmander, "Fourier integral operators, II", *Acta Math.* **128**:3-4 (1972), 183–269. MR Zbl

[Faridani et al. 1992] A. Faridani, E. L. Ritman, and K. T. Smith, "Examples of local tomography", *SIAM J. Appl. Math.* **52**:4 (1992), 1193–1198. MR Zbl

[Faridani et al. 1997] A. Faridani, D. V. Finch, E. L. Ritman, and K. T. Smith, "Local tomography, II", *SIAM J. Appl. Math.* **57**:4 (1997), 1095–1127. MR Zbl

[Garde and Knudsen 2016] H. Garde and K. Knudsen, "Sparsity prior for electrical impedance tomography with partial data", *Inverse Probl. Sci. Eng.* **24**:3 (2016), 524–541. MR Zbl

[Garde and Knudsen 2017] H. Garde and K. Knudsen, "Distinguishability revisited: depth dependent bounds on reconstruction quality in electrical impedance tomography", *SIAM J. Appl. Math.* **77**:2 (2017), 697–720. MR Zbl

[Greenleaf and Uhlmann 1990] A. Greenleaf and G. Uhlmann, "Estimates for singular Radon transforms and pseudodifferential operators with singular symbols", *J. Funct. Anal.* **89**:1 (1990), 202–232. MR Zbl

[Greenleaf and Uhlmann 2001] A. Greenleaf and G. Uhlmann, "Local uniqueness for the Dirichlet-to-Neumann map via the two-plane transform", *Duke Math. J.* **108**:3 (2001), 599–617. MR Zbl

[Greenleaf et al. 2003] A. Greenleaf, M. Lassas, and G. Uhlmann, "The Calderón problem for conormal potentials, I: Global uniqueness and reconstruction", *Comm. Pure Appl. Math.* **56**:3 (2003), 328–352. MR Zbl

[Guillemin 1985] V. Guillemin, "On some results of Gelfand in integral geometry", pp. 149–155 in *Pseudodifferential operators and applications* (Notre Dame, IN, 1984), edited by F. Trèves, Proc. Sympos. Pure Math. **43**, Amer. Math. Soc., Providence, RI, 1985. MR Zbl

[Guillemin and Sternberg 1977] V. Guillemin and S. Sternberg, *Geometric asymptotics*, Mathematical Surveys **14**, Amer. Math. Soc., Providence, RI, 1977. MR Zbl

[Guillemin and Uhlmann 1981] V. Guillemin and G. Uhlmann, "Oscillatory integrals with singular symbols", *Duke Math. J.* **48**:1 (1981), 251–267. MR Zbl

[Haberman 2015] B. Haberman, "Uniqueness in Calderón's problem for conductivities with unbounded gradient", *Comm. Math. Phys.* **340**:2 (2015), 639–659. MR Zbl

[Haberman and Tataru 2013] B. Haberman and D. Tataru, "Uniqueness in Calderón's problem with Lipschitz conductivities", *Duke Math. J.* **162**:3 (2013), 496–516. MR Zbl

[Hamilton et al. 2012] S. J. Hamilton, C. N. L. Herrera, J. L. Mueller, and A. Von Herrmann, "A direct D-bar reconstruction algorithm for recovering a complex conductivity in 2D", *Inverse Problems* **28**:9 (2012), art. id. 095005. MR Zbl

[Hamilton et al. 2014] S. J. Hamilton, A. Hauptmann, and S. Siltanen, "A data-driven edge-preserving D-bar method for electrical impedance tomography", *Inverse Probl. Imaging* **8**:4 (2014), 1053–1072. MR Zbl

[Hamilton et al. 2016] S. J. Hamilton, J. M. Reyes, S. Siltanen, and X. Zhang, "A hybrid segmentation and D-bar method for electrical impedance tomography", *SIAM J. Imaging Sci.* **9**:2 (2016), 770–793. MR Zbl

[Harrach and Ullrich 2013] B. Harrach and M. Ullrich, "Monotonicity-based shape reconstruction in electrical impedance tomography", *SIAM J. Math. Anal.* **45**:6 (2013), 3382–3403. MR Zbl

[Harrach and Ullrich 2015] B. Harrach and M. Ullrich, "Resolution guarantees in electrical impedance tomography", *IEEE. Trans. Med. Imaging* **34**:7 (2015), 1513–1521.

[Holder 1992a] D. S. Holder, "Detection of cerebral ischaemia in the anaesthetised rat by impedance measurement with scalp electrodes: implications for non-invasive imaging of stroke by electrical impedance tomography", *Clin. Phys. Physiol. Meas.* **13**:1 (1992), 63–75.

[Holder 1992b] D. S. Holder, "Electrical impedance tomography with cortical or scalp electrodes during global cerebral ischaemia in the anaesthetised rat", *Clin. Phys. Physiol. Meas.* **13**:1 (1992), 87–98.

[Hörmander 1971] L. Hörmander, "Fourier integral operators, I", *Acta Math.* **127**:1-2 (1971), 79–183. MR Zbl

[Hörmander 1985] L. Hörmander, *The analysis of linear partial differential operators, IV: Fourier integral operators*, Grundlehren der Mathematischen Wissenschaften **275**, Springer, 1985. MR Zbl

[Huhtanen and Perämäki 2012] M. Huhtanen and A. Perämäki, "Numerical solution of the $\mathbb{R}$-linear Beltrami equation", *Math. Comp.* **81**:277 (2012), 387–397. MR Zbl

[Ide et al. 2007] T. Ide, H. Isozaki, S. Nakata, S. Siltanen, and G. Uhlmann, "Probing for electrical inclusions with complex spherical waves", *Comm. Pure Appl. Math.* **60**:10 (2007), 1415–1442. MR Zbl

[Ikehata 2000] M. Ikehata, "Reconstruction of the support function for inclusion from boundary measurements", *J. Inverse Ill-Posed Probl.* **8**:4 (2000), 367–378. MR Zbl

[Ikehata and Siltanen 2000] M. Ikehata and S. Siltanen, "Numerical method for finding the convex hull of an inclusion in conductivity from boundary measurements", *Inverse Problems* **16**:4 (2000), 1043–1052. MR Zbl

[Ikehata and Siltanen 2004] M. Ikehata and S. Siltanen, "Electrical impedance tomography and Mittag–Leffler's function", *Inverse Problems* **20**:4 (2004), 1325–1348. MR Zbl

[Isaacson 1986] D. Isaacson, "Distinguishability of conductivities by electric current computed tomography", *IEEE Trans. Med. Imaging* **5**:2 (1986), 91–95.

[Isaacson et al. 2004] D. Isaacson, J. L. Mueller, J. C. Newell, and S. Siltanen, "Reconstructions of chest phantoms by the D-bar method for electrical impedance tomography", *IEEE Trans. on Med. Imag.* **23** (2004), 821–828.

[Isaacson et al. 2006] D. Isaacson, J. L. Mueller, J. C. Newell, and S. Siltanen, "Imaging cardiac activity by the D-bar method for electrical impedance tomography", *Physiolog. Meas.* **27**:5 (2006), S43–S50.

[Jin and Maass 2012] B. Jin and P. Maass, "Sparsity regularization for parameter identification problems", *Inverse Problems* **28**:12 (2012), art. id. 123001. MR Zbl

[Kaipio et al. 2000] J. P. Kaipio, V. Kolehmainen, E. Somersalo, and M. Vauhkonen, "Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography", *Inverse Problems* **16**:5 (2000), 1487–1522. MR Zbl

[Kim 2008] S. E. Kim, "Calderón's problem for Lipschitz piecewise smooth conductivities", *Inverse Problems* **24**:5 (2008), art. id. 055016. MR Zbl

[Kirsch 1998] A. Kirsch, "Characterization of the shape of a scattering obstacle using the spectral data of the far field operator", *Inverse Problems* **14**:6 (1998), 1489–1512. MR Zbl

[Knudsen 2003] K. Knudsen, "A new direct method for reconstructing isotropic conductivities in the plane", *Physiolog. Meas.* **24**:2 (2003), 391–403.

[Knudsen et al. 2004] K. Knudsen, J. Mueller, and S. Siltanen, "Numerical solution method for the dbar-equation in the plane", *J. Comput. Phys.* **198**:2 (2004), 500–517. MR Zbl

[Knudsen et al. 2007] K. Knudsen, M. Lassas, J. L. Mueller, and S. Siltanen, "D-bar method for electrical impedance tomography with discontinuous conductivities", *SIAM J. Appl. Math.* **67**:3 (2007), 893–913. MR Zbl

[Knudsen et al. 2009] K. Knudsen, M. Lassas, J. L. Mueller, and S. Siltanen, "Regularized D-bar method for the inverse conductivity problem", *Inverse Probl. Imaging* **3**:4 (2009), 599–624. MR Zbl

[Kohn and Vogelius 1985] R. V. Kohn and M. Vogelius, "Determining conductivity by boundary measurements, II: Interior results", *Comm. Pure Appl. Math.* **38**:5 (1985), 643–667. MR Zbl

[Kuchment 2014] P. Kuchment, *The Radon transform and medical imaging*, CBMS-NSF Regional Conference Series in Applied Mathematics **85**, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2014. MR Zbl

[Lechleiter 2006] A. Lechleiter, "A regularization technique for the factorization method", *Inverse Problems* **22**:5 (2006), 1605–1625. MR Zbl

[Lechleiter et al. 2008] A. Lechleiter, N. Hyvönen, and H. Hakula, "The factorization method applied to the complete electrode model of impedance tomography", *SIAM J. Appl. Math.* **68**:4 (2008), 1097–1121. MR Zbl

[Malone et al. 2014] E. Malone, M. Jehl, S. Arridge, T. Betcke, and D. Holder, "Stroke type differentiation using spectrally constrained multifrequency EIT: evaluation of feasibility in a realistic head model", *Physiolog. Meas.* **35**:6 (2014), 1051–1066.

[Mandache 2001] N. Mandache, "Exponential instability in an inverse problem for the Schrödinger equation", *Inverse Problems* **17**:5 (2001), 1435–1444. MR Zbl

[Melrose and Uhlmann 1979] R. B. Melrose and G. A. Uhlmann, "Lagrangian intersection and the Cauchy problem", *Comm. Pure Appl. Math.* **32**:4 (1979), 483–519. MR Zbl

[Mendoza 1982] G. Mendoza, "Symbol calculus associated with intersecting Lagrangians", *Comm. Partial Differential Equations* **7**:9 (1982), 1035–1116. MR Zbl

[Mueller and Siltanen 2003] J. L. Mueller and S. Siltanen, "Direct reconstructions of conductivities from boundary measurements", *SIAM J. Sci. Comput.* **24**:4 (2003), 1232–1266. MR Zbl

[Mueller and Siltanen 2012] J. L. Mueller and S. Siltanen, *Linear and nonlinear inverse problems with practical applications*, Computational Science & Engineering **10**, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2012. MR Zbl

[Nachman 1996] A. I. Nachman, "Global uniqueness for a two-dimensional inverse boundary value problem", *Ann. of Math.* (2) **143**:1 (1996), 71–96. MR Zbl

[Nagayasu et al. 2009] S. Nagayasu, G. Uhlmann, and J.-N. Wang, "A depth-dependent stability estimate in electrical impedance tomography", *Inverse Problems* **25**:7 (2009), art. id. 075001. MR Zbl

[Nie and Brown 2011] Z. Nie and R. M. Brown, "Estimates for a family of multi-linear forms", *J. Math. Anal. Appl.* **377**:1 (2011), 79–87. MR Zbl

[Perry 2016] P. A. Perry, "Global well-posedness and long-time asymptotics for the defocussing Davey–Stewartson II equation in $H^{1,1}(\mathbb{C})$", *J. Spectr. Theory* **6**:3 (2016), 429–481. MR Zbl

[Phong and Stein 1986] D. H. Phong and E. M. Stein, "Hilbert integrals, singular integrals, and Radon transforms, I", *Acta Math.* **157**:1-2 (1986), 99–157. MR Zbl

[Rondi and Santosa 2001] L. Rondi and F. Santosa, "Enhanced electrical impedance tomography via the Mumford–Shah functional", *ESAIM Control Optim. Calc. Var.* **6** (2001), 517–538. MR Zbl

[Siltanen et al. 2000] S. Siltanen, J. Mueller, and D. Isaacson, "An implementation of the reconstruction algorithm of A. Nachman for the 2D inverse conductivity problem", *Inverse Problems* **16**:3 (2000), 681–699. MR Zbl

[Sylvester and Uhlmann 1987] J. Sylvester and G. Uhlmann, "A global uniqueness theorem for an inverse boundary value problem", *Ann. of Math.* (2) **125**:1 (1987), 153–169. MR Zbl

[Tanushev and Vese 2007] N. M. Tanushev and L. A. Vese, "A piecewise-constant binary model for electrical impedance tomography", *Inverse Probl. Imaging* **1**:2 (2007), 423–435. MR Zbl

[Uhlmann and Wang 2008] G. Uhlmann and J.-N. Wang, "Reconstructing discontinuities using complex geometrical optics solutions", *SIAM J. Appl. Math.* **68**:4 (2008), 1026–1044. MR Zbl

[Vainikko 2000] G. Vainikko, "Fast solvers of the Lippmann–Schwinger equation", pp. 423–440 in *Direct and inverse problems of mathematical physics* (Newark, DE, 1997), edited by R. P. Gilbert et al., Int. Soc. Anal. Appl. Comput. **5**, Kluwer, Dordrecht, 2000. MR Zbl

[Zhou et al. 2015] Z. Zhou, G. S. dos Santos, T. Dowrick, J. Avery, Z. Sun, H. Xu, and D. S. Holder, "Comparison of total variation algorithms for electrical impedance tomography", *Physiolog. Meas.* **36**:6 (2015), 1193–1209.

Received 12 Dec 2016. Revised 21 Sep 2017. Accepted 14 Nov 2017.

ALLAN GREENLEAF: allan.greenleaf@rochester.edu
Department of Mathematics, University of Rochester, Rochester, NY, United States

MATTI LASSAS: matti.lassas@helsinki.fi
Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland

MATTEO SANTACESARIA: matteo.santacesaria@helsinki.fi
Dipartimento di Matematica, Politecnico di Milano, Milano, Italy

SAMULI SILTANEN: samuli.siltanen@helsinki.fi
Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland

GUNTHER UHLMANN: guntheruhlmann@gmail.com
Department of Mathematics, University of Washington, Seattle, WA, United States

ANALYSIS AND PDE
Vol. 11, No. 8, 2018
dx.doi.org/10.2140/apde.2018.11.1945

QUANTITATIVE STOCHASTIC HOMOGENIZATION AND REGULARITY THEORY OF PARABOLIC EQUATIONS

SCOTT ARMSTRONG, ALEXANDRE BORDAS AND JEAN-CHRISTOPHE MOURRAT

We develop a quantitative theory of stochastic homogenization for linear, uniformly parabolic equations with coefficients depending on space and time. Inspired by recent works in the elliptic setting, our analysis is focused on certain subadditive quantities derived from a variational interpretation of parabolic equations. These subadditive quantities are intimately connected to spatial averages of the fluxes and gradients of solutions. We implement a renormalization-type scheme to obtain an algebraic rate for their convergence, which is essentially a quantification of the weak convergence of the gradients and fluxes of solutions to their homogenized limits. As a consequence, we obtain estimates of the homogenization error for the Cauchy–Dirichlet problem which are optimal in stochastic integrability. We also develop a higher regularity theory for solutions of the heterogeneous equation, including a uniform $C^{0,1}$-type estimate and a Liouville theorem of every finite order.

1. Introduction

1A. ***Motivation and informal summary of results.*** In this paper, we develop a quantitative theory of stochastic homogenization for linear, uniformly parabolic equations with coefficients depending on both the space and time variables. We consider equations of the form

$$\partial_t u^\varepsilon - \nabla \cdot \left(\boldsymbol{a}\left(\frac{t}{\varepsilon^2}, \frac{x}{\varepsilon} \right) \nabla u^\varepsilon \right) = 0 \quad \text{in } I \times U, \tag{1-1}$$

MSC2010: primary 35B27, 35B45; secondary 60K37, 60F05.
Keywords: stochastic homogenization, parabolic equation, large-scale regularity, variational methods.

where $I \subseteq \mathbb{R}$ is an open interval, U is a bounded Lipschitz domain of $\mathbb{R}^d$, and $(t, x) \mapsto \boldsymbol{a}(t, x)$ is a stationary random field taking values in the set of real d-by-d matrices satisfying, for a fixed constant $\Lambda \in [1, \infty)$,

$$\text{for all } \xi \in \mathbb{R}^d, \quad \xi \cdot \boldsymbol{a}(t, x)\xi \geq \Lambda^{-1}|\xi|^2 \quad \text{and} \quad |\boldsymbol{a}(t, x)\xi| \leq \Lambda|\xi|. \tag{1-2}$$

Here the symbol ∇ denotes the gradient in the space variables only; that is, $\nabla w = (\partial_{x_1} w, \ldots, \partial_{x_d} w)$. We let $\mathbb{P}$ be the law of the random field $\boldsymbol{a}(t, x)$, which we assume to be invariant under translations by elements of $\mathbb{Z} \times \mathbb{Z}^d$ and to have a finite range of dependence. (See the following subsection for the precise assumptions.)

We are interested in the behavior of the solutions $u^\varepsilon(t, x)$ for $0 < \varepsilon \ll 1$. It is well known that, under very general qualitative assumptions on the coefficients (stationarity and ergodicity), (1-1) homogenizes to an effective limiting equation of the form

$$\partial_t u - \nabla \cdot (\bar{\boldsymbol{a}} \nabla u) = 0 \quad \text{in } I \times U, \tag{1-3}$$

where $\bar{\boldsymbol{a}}$ is a deterministic d-by-d matrix. This principle can be formulated in various ways, but it means for example that the solutions u^ε of (1-1), subject to appropriate initial-boundary conditions, converge as $\varepsilon \to 0$, $\mathbb{P}$-almost surely and in some appropriate function space, to solutions of the homogenized equation (1-3).[1] Such a result is usually proved by soft arguments, using an abstract version of the ergodic theorem, which unfortunately does not give *quantitative* information concerning the convergence.

There has been a lot of recent interest in quantitative stochastic homogenization for elliptic equations, particularly in the case of linear, uniformly elliptic equations. This essentially began with the work of Gloria and Otto [2011; 2012], who proved the first quantitative results which are optimal in the scaling of the parameter ε; see also [Gloria et al. 2015]. Their work motivated a great number of subsequent works, and we refer to the recently completed [Armstrong et al. 2017b] for more background, references and historical information.

In this paper, motivated by the desire to obtain quantitative homogenization results — in particular, explicit estimates of the homogenization error — we develop an analytic approach for parabolic equations with random coefficients based on the ideas recently introduced in [Armstrong and Smart 2016; Armstrong and Mourrat 2016; Armstrong et al. 2016; 2017a], which are perhaps best presented in [Armstrong et al. 2017b]. Those papers developed a rather complete quantitative theory of elliptic homogenization starting from the observation that certain energy quantities — which are very natural from a variational perspective — are also rather convenient for studying the homogenization process. This is because: (i) they efficiently encode information about the weak convergence of the fluxes, gradients, and energy densities of solutions; and (ii) they are amenable to renormalization arguments in the sense that we can obtain rates of convergence for the quantities by iterating the length scale. This variational approach allows one to circumvent the need for nonlinear concentration inequalities, because it reveals a "linear" structure of the randomness: while the solutions are very nonlinear functions of the coefficients, the energy quantities

[1] We remark that we are unaware of a reference which proves this specific result in the parabolic setting. Nevertheless, we maintain that it is essentially well known, since the classical qualitative proof given in the elliptic case, see for instance [Papanicolaou and Varadhan 1981; Bensoussan et al. 1978; Jikov et al. 1994], can be straightforwardly generalized to the parabolic setting.

turn out to be essentially linear. This observation greatly simplifies the theory and allows one to derive estimates which are optimal both in the scaling of ε as well as in stochastic integrability. A related approach inspired by [Armstrong and Smart 2016; Armstrong and Mourrat 2016; Armstrong et al. 2016; 2017a] has also recently been developed in [Gloria and Otto 2015; 2016].

The two main results of this paper are (i) a quantitative estimate on the homogenization error for Cauchy–Dirichlet problems (Theorem 1.1) and (ii) a complete large-scale regularity theory (Theorem 1.2). It has already been observed in the elliptic case, see [Armstrong et al. 2017b], that results of this type are the first step towards optimal quantitative estimates and scaling limits for first-order correctors as well as optimal error estimates for boundary-value problems. At the same time, the results in this paper are the first quantitative stochastic homogenization results, to our knowledge, for parabolic equations with coefficients with space-time dependence.

The starting point for adapting the techniques of [Armstrong et al. 2017b] to the parabolic case is a variational characterization of divergence-form parabolic equations that was first discovered by Brezis and Ekeland [1976a; 1976b]. We give a self-contained presentation of this characterization in Appendix A, where we also give a convex analytic proof of the well-posedness of general Cauchy–Dirichlet problems inspired by [Ghoussoub and Tzou 2004]. Based on this variational principle, we introduce subadditive quantities for the homogenization problem in Section 2 and adapt the methods of [Armstrong et al. 2017b], using an iteration of scales and a renormalization-type argument, to obtain an algebraic rate of convergence in Section 4. Compared to the elliptic case, the main sources of additional difficulty in the iteration argument have to do with the need to control certain weak Sobolev norms of the time derivatives of the solutions. We accomplish this with the help of some functional inequalities we prove in Section 3. In Section 5, we show that the convergence of the subadditive quantities gives us approximate first-order correctors with good quantitative bounds, which allows us to prove Theorem 1.1. In the last section, we obtain the regularity result of Theorem 1.2. In the rest of this introduction, we state the assumptions, notation and main results.

1B. ***Assumptions.*** We fix a spatial dimension $d \geq 2$ and a parameter $\Lambda \in [1, \infty)$. We let Ω denote the set of all possible coefficient fields $\boldsymbol{a}(t, x)$, which are assumed to be measurable maps from $\mathbb{R} \times \mathbb{R}^d$ into the set Ω_0 of matrices $\boldsymbol{a}$ satisfying

$$\text{for all } \xi \in \mathbb{R}^d, \quad \xi \cdot \boldsymbol{a}\xi \geq \Lambda^{-1}|\xi|^2 \quad \text{and} \quad |\boldsymbol{a}\xi| \leq \Lambda|\xi|. \tag{1-4}$$

That is, we define

$$\Omega_0 := \{\boldsymbol{a} \in \mathbb{R}^{d\times d} : \boldsymbol{a} \text{ satisfies (1-4)}\}$$

and then set

$$\Omega := \{\boldsymbol{a} : \mathbb{R} \times \mathbb{R}^d \to \Omega_0 \text{ is Lebesgue-measurable}\}. \tag{1-5}$$

For every Borel subset $V \subseteq \mathbb{R} \times \mathbb{R}^d$, we define $\mathcal{F}_V$ to be the σ-algebra representing the information obtaining by observing the coefficient field in V. Formally,

$$\begin{aligned} \mathcal{F}_V := {} & \text{the } \sigma\text{-algebra generated by the random elements} \\ & \boldsymbol{a} \mapsto \int_V \varphi(t, x)\, \boldsymbol{a}(t, x)\, dt\, dx, \quad \varphi \in C_c^\infty(\mathbb{R} \times \mathbb{R}^d). \end{aligned} \tag{1-6}$$

The largest of the σ-algebras in this family is $\mathcal{F} := \mathcal{F}_{\mathbb{R}\times\mathbb{R}^d}$. We assume that $\mathbb{P}$ is a given probability measure on the measurable space $(\Omega, \mathcal{F})$ which satisfies the following two assumptions:

(P1) $\mathbb{P}$ is *stationary* with respect to $\mathbb{Z}\times\mathbb{Z}^d$-translations. For every $z \in \mathbb{Z}\times\mathbb{Z}^d$ and event $A \in \mathcal{F}$,

$$\mathbb{P}[A] = \mathbb{P}[T_z A].$$

(P2) $\mathbb{P}$ has a *unit range of dependence*. For every pair of Borel subsets $U, V \subseteq \mathbb{R}\times\mathbb{R}^d$,

$$\operatorname{dist}(U, V) \geq 1 \quad \Longrightarrow \quad \mathcal{F}_U \text{ and } \mathcal{F}_V \text{ are } \mathbb{P}\text{-independent.}$$

Here "dist" is defined with respect to the usual Euclidean distance on $\mathbb{R}\times\mathbb{R}^d$. We denote by $\mathbb{E}[X]$ the expectation of an $\mathcal{F}$-measurable random variable X with respect to $\mathbb{P}$. While we assume that the coefficient field has a finite range of dependence for simplicity, we point out that this hypothesis can be weakened using arguments similar to those exposed in [Armstrong and Mourrat 2016].

1C. ***Notation.*** We unfortunately must introduce quite a bit of notation, particularly since we are working with parabolic equations which require us to define various function spaces. We collect the notation needed in this subsection, which the reader is encouraged to skim and consult as a reference.

General notation. We denote the set of natural numbers by $\mathbb{N} := \{0, 1, 2, \ldots\}$. We use the symbols $\wedge$ and $\vee$ to denote minimum and maximum, respectively; for example $r \wedge s = \min\{r, s\}$ for $r, s \in \mathbb{R}$. For every $r \in \mathbb{R}$, we also define $r_+ := r \vee 0$ and $r_- := r \wedge 0$. For any $m \in \mathbb{N}$ and measurable subset $E \subseteq \mathbb{R}^m$, the Lebesgue measure of E is denoted by $|E|$, unless E is a finite set, in which case $|E|$ is the cardinality of E. This is often used for $m \in \{1, d, 1+d\}$. A slash through the integral denotes normalization by the Lebesgue measure: $\fint_E := (1/|E|)\int_E$. The mean of a function $f \in L^1(E)$ is also denoted by $(f)_E := \fint_E f$.

A *parabolic cylinder* is any set of the form $I \times U$ where $I = (I_-, I_+) \subseteq \mathbb{R}$ is a bounded open interval and $U \subseteq \mathbb{R}^d$ is a bounded Lipschitz domain. We denote the parabolic boundary of $I \times U$ by

$$\partial_\sqcup(I \times U) := (I \times \partial U) \cup (\{I_-\} \times U).$$

We denote the Euclidean ball of $\mathbb{R}^d$ of radius $r \in (0, \infty]$ centered at $x \in \mathbb{R}^d$ by $B_r(x)$, and put $B_r := B_r(0)$. Throughout, we work with the triadic cubes defined for every $n \in (0, \infty)$ by

$$I_n := \left(-\frac{3^{2n}}{2}, \frac{3^{2n}}{2}\right), \quad \square_n := \left(-\frac{3^n}{2}, \frac{3^n}{2}\right)^d, \quad \boxdot_n := I_n \times \square_n.$$

Note that the parabolic cylinder $\boxdot_n$ is evidently not a cube per se since its sides have a scaling which matches the parabolic scaling. However, we note that for each $m, n \in \mathbb{N}$ with $m < n$, we can write $\boxdot_n$ as the disjoint union (up to a set of Lebesgue measure zero) of exactly $3^{(2+d)(m-n)}$ cubes of the form $z + \boxdot_m$ with $z \in 3^{2m}\mathbb{Z}\times 3^m\mathbb{Z}$.

We also use the following notation for parabolic cylinders: for each $r \in (0, \infty]$ and $(t, x) \in \mathbb{R}\times\mathbb{R}^d$, we define

$$\tilde{I}_r := (-r^2, 0], \quad Q_r(t, x) := (t, x) + \tilde{I}_r \times B_r, \quad \text{and} \quad Q_r := Q_r(0, 0). \tag{1-7}$$

Function spaces. For every bounded Lipschitz domain $U \subseteq \mathbb{R}^d$ with $|U| < \infty$ and $p \in [1, \infty)$, we denote the normalized $L^p(U)$ norm of a function $f \in L^p(U)$ by

$$\|f\|_{\underline{L}^p(U)} := \left(\fint_U |f|^p \right)^{\frac{1}{p}} = |U|^{-\frac{1}{p}} \|f\|_{L^p(U)}. \tag{1-8}$$

For $p = \infty$, we define $\|f\|_{\underline{L}^\infty(U)} := \|f\|_{L^\infty(U)}$. We use similar notation to denote normalized (scale-invariant) Sobolev norms: for every $p \in [1, \infty)$ and $f \in W^{1,p}(U)$,

$$\|f\|_{\underline{W}^{1,p}(U)} := |U|^{-\frac{1}{d}} \|f\|_{\underline{L}^p(U)} + \|\nabla f\|_{\underline{L}^p(U)}.$$

In the case $p = 2$ we use the notation $\|f\|_{\underline{H}^1(U)} := \|f\|_{\underline{W}^{1,2}(U)}$. As usual, $H_0^1(U)$ and $W_0^{1,p}(U)$ respectively denote the closure in $H^1(U)$ and $W^{1,p}(U)$, respectively, of the compactly supported smooth functions in U. The dual spaces to $W^{1,p}(U)$ and $W_0^{1,p}(U)$ are denoted by $\widehat{W}^{-1,p'}(U)$ and $W^{-1,p'}(U)$, respectively, where $p' := p/(p-1)$ is the Hölder conjugate exponent of p. The normalized, scale-invariant dual norms are respectively defined by

$$\|v\|_{\underline{\widehat{W}}^{-1,p'}(U)} := \sup\left\{ \fint_U uv \,:\, u \in W^{1,p}(U),\ \|u\|_{\underline{W}^{1,p}(U)} \le 1 \right\},$$

$$\|v\|_{\underline{W}^{-1,p'}(U)} := \sup\left\{ \fint_U uv \,:\, u \in W_0^{1,p}(U),\ \|u\|_{\underline{W}^{1,p}(U)} \le 1 \right\}.$$

Here we are abusing notation by denoting the natural pairing $\langle u, w \rangle$ between the two dual spaces (up to a constant) by the normalized integral. This is done to emphasize the normalization that we wish to enforce, which extends the action of an element $w \in C_c^\infty(U)$ on $W^{1,p}(U)$ by $u \mapsto \fint_U uw$. For $p = 2$, we also write $\|\cdot\|_{\underline{\widehat{H}}^{-1}(U)} := \|v\|_{\underline{\widehat{W}}^{-1,2}(U)}$ and $\|\cdot\|_{\underline{H}^{-1}(U)} := \|v\|_{\underline{W}^{-1,2}(U)}$.

We next introduce function spaces designed for parabolic equations. For each $n \in \mathbb{N}$, bounded Lipschitz domain $U \subseteq \mathbb{R}^n$, Banach space X and $p \in [1, \infty)$, we denote by $L^p(U; X)$ the space of Lebesgue-measurable mappings $u : U \to X$ such that

$$\|u\|_{\underline{L}^p(U;X)} := \left(\fint_U \|u(x)\|_X^p \, dx \right)^{\frac{1}{p}} < \infty.$$

For every interval $I = (I_-, I_+) \subseteq \mathbb{R}$ and bounded Lipschitz domain $U \subseteq \mathbb{R}^d$, we define the function space

$$H^1_{\mathrm{par}}(I \times U) := \{u \in L^2(I; H^1(U)) : \partial_t u \in L^2(I; H^{-1}(U))\}, \tag{1-9}$$

which is the closure of bounded smooth functions on $I \times U$ with respect to the norm

$$\|u\|_{\underline{H}^1_{\mathrm{par}}(I\times U)} := \|u\|_{\underline{L}^2(I;\underline{H}^1(U))} + \|\partial_t u\|_{\underline{L}^2(I;\underline{H}^{-1}(U))}. \tag{1-10}$$

We denote by $H^1_{\mathrm{par},\sqcup}(I \times U)$ the closure in $H^1_{\mathrm{par}}(I \times U)$ of the set of smooth functions with compact support in $(I_-, I_+] \times U$. In other words, a function in $H^1_{\mathrm{par},\sqcup}(I \times U)$ has zero trace on the lateral boundary $I \times \partial U$ and the initial time $\{I_-\} \times U$ but does not necessarily vanish at the final time.

We let $H^1_{\mathrm{par},0}(I\times U)$ denote the completion of the set of smooth functions with compact support in $I\times U$ with respect to the norm

$$\|v\|_{\underline{H}^1_{\mathrm{par},0}(I\times U)} := \|v\|_{\underline{L}^2(I;\underline{H}^1(U))} + \|\partial_t v\|_{\underline{L}^2(I;\underline{\widehat{H}}^{-1}(U))}. \tag{1-11}$$

Note that compared with (1-10), here we require the time derivative $\partial_t v(t,\cdot)$ to be an element of $\widehat{H}^{-1}(U)$ instead of $H^{-1}(U)$. In particular, for $v\in H^1_{\mathrm{par},0}(I\times U)$, the spatial average of $\partial_t v$ over $I\times U$ is well-defined, since constant functions belong to $L^2(I;H^1(U))$, while they do not belong to $L^2(I;H^1_0(U))$. Moreover, the boundary condition imposes that for every $v\in H^1_{\mathrm{par},0}(I\times U)$,

$$\int_{I\times U}\partial_t v = 0. \tag{1-12}$$

This identity is indeed clear if v is smooth and compactly supported in $I\times U$, and we can then obtain the general case by density.

In certain situations, it is useful to work with variations of $H^1_{\mathrm{par}}(I\times U)$ in which the exponent of integrability is $p\in(1,\infty)$ rather than 2. So we also define the function space

$$W^{1,p}_{\mathrm{par}}(I\times U) := \{u\in L^p(I;W^{1,p}(U)) : \partial_t u\in L^p(I;W^{-1,p}(U))\}, \tag{1-13}$$

which is the closure of bounded smooth functions on $I\times U$ with respect to the norm

$$\|u\|_{\underline{W}^{1,p}_{\mathrm{par}}(I\times U)} := \|u\|_{\underline{L}^p(I;\underline{W}^{1,p}(U))} + \|\partial_t u\|_{\underline{L}^p(I;\underline{W}^{-1,p}(U))}. \tag{1-14}$$

Similarly to $H^1_{\mathrm{par},\sqcup}(I\times U)$, we denote by $W^{1,p}_{\mathrm{par},\sqcup}(I\times U)$ the closure in $W^{1,p}_{\mathrm{par}}(I\times U)$ of the set of smooth functions with compact support in $(I_-,I_+]\times U$. Finally, for every parabolic cylinder V, we denote by $W^{1,p}_{\mathrm{par,\,loc}}(V)$, $H^1_{\mathrm{par,\,loc}}(V)$, and so forth, the functions on V which are, respectively, elements of $W^{1,p}_{\mathrm{par}}(W)$ and $H^1_{\mathrm{par}}(W)$, etc., for every subcylinder $W\subseteq V$ with $\overline{W}\subseteq V$.

We next turn to the definitions of the negative parabolic Sobolev spaces. We denote by $\widehat{H}^{-1}_{\mathrm{par}}(V)$ and $H^{-1}_{\mathrm{par}}(V)$ the dual spaces to $H^1_{\mathrm{par}}(V)$ and $H^1_{\mathrm{par},\sqcup}(V)$, respectively, with (normalized, scale-invariant) dual norms given by

$$\begin{aligned}
\|f\|_{\underline{\widehat{H}}^{-1}_{\mathrm{par}}(V)} &:= \sup\left\{ \fint_V fw \,:\, w\in H^1_{\mathrm{par}}(V),\ \|w\|_{\underline{H}^1_{\mathrm{par}}(V)}\le 1\right\},\\
\|f\|_{\underline{H}^{-1}_{\mathrm{par}}(V)} &:= \sup\left\{ \fint_V fw \,:\, w\in H^1_{\mathrm{par},\sqcup}(V),\ \|w\|_{\underline{H}^1_{\mathrm{par}}(V)}\le 1\right\}.
\end{aligned} \tag{1-15}$$

As explained above, the notation $\fint_V fw$ should be interpreted as the canonical pairing between $f\in\widehat{H}^{-1}_{\mathrm{par}}(V)$ or $f\in H^{-1}_{\mathrm{par}}(V)$, respectively, and $w\in H^1_{\mathrm{par}}(V)$ or $w\in H^1_{\mathrm{par},\sqcup}(V)$, which extends the action of bounded smooth functions on $H^1_{\mathrm{par}}(V)$ or $H^1_{\mathrm{par},\sqcup}(V)$. We similarly define the space $W^{-1,p}_{\mathrm{par}}(V)$ to be the dual space of the Banach space $W^{1,p'}_{\mathrm{par},\sqcup}(V)$, where $p' := p/(p-1)$, and endow it with the (normalized, scale-invariant) norm

$$\|f\|_{\underline{W}^{-1,p}_{\mathrm{par}}(V)} := \sup\left\{ \fint_V fw \,:\, w\in W^{1,p'}_{\mathrm{par},\sqcup}(V),\ \|w\|_{\underline{W}^{1,p'}_{\mathrm{par}}(V)}\le 1\right\}. \tag{1-16}$$

Recall that negative Sobolev norms arise naturally when one wishes to quantify *weak* convergence in L^p or positive Sobolev spaces; see [Armstrong et al. 2017b, Section 1.4]. This is indeed their purpose in this paper.

The $\mathcal{O}_s$ notation. Since the random variables we encounter in this paper are very often the sum of a deterministic quantity and a "small" random part, it is useful to work with the notation introduced in [Armstrong et al. 2017a] for expressing the sizes of random variables (essentially, an alternative notation for certain Orlicz norms). It is intended to remind us of "big-O" notation and is convenient because it compresses some of our computations and makes our inequalities easier to understand at a glance.

If X is a random variable and $s, k \in (0, \infty)$, then we write

$$X \le \mathcal{O}_s(k)$$

as a shorthand for the statement

$$\mathbb{E}\left[\exp\left(\left(\frac{X_+}{k}\right)^s\right)\right] \le 2. \tag{1-17}$$

Roughly, this means that "X is of order k with stretched exponential tails with exponent s." More precisely, we can use Chebyshev's inequality to see that

$$X \le \mathcal{O}_s(k) \quad \Longrightarrow \quad \text{for all } \lambda > 0, \ \ \mathbb{P}[X > \lambda k] \le 2\exp(-\lambda^s). \tag{1-18}$$

The converse of this statement is almost true: for every $k \ge 0$,

$$\text{for all } \lambda \ge 0, \ \ \mathbb{P}[X \ge \lambda k] \le \exp(-\lambda^s) \quad \Longrightarrow \quad X \le \mathcal{O}_s(2^{\frac{1}{s}}\theta). \tag{1-19}$$

This can be obtained by integration. We also use the notation

$$X = \mathcal{O}_s(k) \quad \Longleftrightarrow \quad X \le \mathcal{O}_s(k) \ \text{ and } \ -X \le \mathcal{O}_s(k).$$

Similarly, we write $X \le Y + \mathcal{O}_s(k)$ to mean that $X - Y \le \mathcal{O}_s(k)$ and $X = Y + \mathcal{O}_s(k)$ to mean that $X - Y = \mathcal{O}_s(k)$. If $s \in [1, \infty)$, then Jensen's inequality gives us a triangle inequality for $\mathcal{O}_s(\,\cdot\,)$ in the following sense: for any measure space $(E, \mathcal{S}, \mu)$, measurable function $K : E \to (0, \infty)$ and jointly measurable family $\{X(z)\}_{z\in E}$ of nonnegative random variables, we have

$$\text{for all } z \in E, \ \ X(z) \le \mathcal{O}_s(K(z)) \quad \Longrightarrow \quad \int_E X \, d\mu \le \mathcal{O}_s\left(\int_E K \, d\mu\right). \tag{1-20}$$

If $s \in (0, 1]$, then the statement is true after adding a prefactor constant $C_s > 1$ to the right side. For a proof of (1-18)–(1-20), see [Armstrong et al. 2017b, Appendix A].

1D. ***Statement of the main results.*** We present two main results. The first provides an algebraic convergence rate for the homogenization limit of the Cauchy–Dirichlet initial-value problem in a parabolic cylinder $I \times U$, where $U \subseteq \mathbb{R}^d$ is a bounded Lipschitz domain. This is a parabolic counterpart of a theorem proved in the elliptic setting in [Armstrong and Smart 2016]; see also [Armstrong et al. 2017b, Theorem 2.16].

Theorem 1.1. *Fix $s \in (0, 2+d)$, a bounded Lipschitz domain $U \subseteq B_1$, an interval $I := (I_-, 0) \subseteq \left(-\frac{1}{4}, 0\right)$ and an exponent $\delta > 0$. Put $V := I \times U$. There exist an exponent $\beta(\delta, V, d, \Lambda) > 0$, a constant $C(s, V, \delta, d, \Lambda) < \infty$ and a random variable $\mathcal{X}$ satisfying*

$$\mathcal{X} = \mathcal{O}_1(C)$$

such that the following convergence result holds: for each $\varepsilon \in \left(0, \frac{1}{2}\right]$ and initial-boundary condition $f \in W^{1,2+\delta}_{\mathrm{par}}(V)$, defining

$$\boldsymbol{a}^\varepsilon(t,x) := \boldsymbol{a}\left(\frac{t}{\varepsilon^2}, \frac{x}{\varepsilon}\right)$$

and taking u^ε, $u \in f + H^1_{\mathrm{par},\sqcup}(V)$ to be the solutions of the Cauchy–Dirichlet problems

$$\begin{cases} \partial_t u^\varepsilon - \nabla \cdot (\boldsymbol{a}^\varepsilon \nabla u^\varepsilon) = 0 & \text{in } V, \\ u^\varepsilon = f & \text{on } \partial_\sqcup V, \end{cases} \quad \text{and} \quad \begin{cases} \partial_t u - \nabla \cdot (\bar{\boldsymbol{a}} \nabla u) = 0 & \text{in } V, \\ u = f & \text{on } \partial_\sqcup V, \end{cases} \tag{1-21}$$

we have the estimate

$$\|\nabla u^\varepsilon - \nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(V)} + \|\boldsymbol{a}^\varepsilon \nabla u^\varepsilon - \bar{\boldsymbol{a}} \nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(V)} + \|u^\varepsilon - u\|_{L^2(V)} \le C \|\nabla f\|_{W^{1,2+\delta}_{\mathrm{par}}(V)} (\varepsilon^{\beta(2+d-s)} + \mathcal{X}\varepsilon^s). \tag{1-22}$$

As well as estimating the homogenization error, notice that the estimate (1-22) quantifies the weak convergence in $L^2(V)$ of the gradients and fluxes of u^ε to those of u. The random part of the error, namely $\mathcal{X}\varepsilon^s$ for an s arbitrarily close to $2+d$, is very small compared to the deterministic part, $\varepsilon^{\beta(2+d-s)}$. It is also important for applications to observe that $\mathcal{X}$ is independent of the initial-boundary condition f.

On the right side of (1-22), we have split the error into a possibly rather large deterministic part (large, since we do not control the smallness of $\beta > 0$) plus a random error. While the typical size of the error is estimated suboptimally, since $\beta > 0$ is small, the *tail behavior* of this random part is sharply estimated. In particular, we see that the probability for the term $(\varepsilon^{\beta(2+d-s)} + \mathcal{X}\varepsilon^s)$ to be $O(1)$ is smaller than $\exp(-\varepsilon^{-s}/C)$, for arbitrary $s < 2+d$. This estimate is sharp, in the sense that it would be false for any $s > 2+d$. We refer to [Armstrong et al. 2017b, Remark 2.5 and Section 3.5] for similar considerations in the elliptic setting.

The second theorem we present here is a large-scale regularity result, a parabolic counterpart to [Armstrong et al. 2017b, Theorem 3.6]. In particular, we seek to classify all ancient solutions of the parabolic equation which exhibit at most polynomial growth at infinity and backwards in time. This requires us to introduce some additional notation.

We denote polynomials in the variables $t, x_1, \ldots, x_d$ by $\mathcal{P}(\mathbb{R} \times \mathbb{R}^d)$. The *parabolic degree* $\deg_p(w)$ of an element $w \in \mathcal{P}(\mathbb{R} \times \mathbb{R}^d)$ is the degree of the polynomial $(t, x) \mapsto w(t^2, x)$. For each $k \in \mathbb{N}$ we let $\mathcal{P}_k(\mathbb{R} \times \mathbb{R}^d)$ be the subset of $\mathcal{P}(\mathbb{R} \times \mathbb{R}^d)$ of polynomials with parabolic degree at most k. For $\alpha > 0$, we say that a function $\phi : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ or $\phi : (-\infty, 0] \times \mathbb{R}^d \to \mathbb{R}$ is *parabolically α-homogeneous* if

$$\text{for all } \lambda \in \mathbb{R}, \quad \phi(\lambda^2 t, \lambda x) = \lambda^\alpha \phi(t, x).$$

Any element of $\mathcal{P}_k(\mathbb{R} \times \mathbb{R}^d)$ can be written as a sum of at most $C(d, k) < \infty$ many parabolically homogeneous polynomials.

We denote by $\bar{\mathcal{A}}_k(Q_\infty)$ the set of $\bar{\boldsymbol{a}}$-caloric functions on Q_∞ with growth which is strictly less than a polynomial of parabolic degree $k+1$:

$$\bar{\mathcal{A}}_k(Q_\infty) := \Big\{ w \in H^1_{\mathrm{par}}(Q_\infty) \,:\, \limsup_{r\to\infty} r^{-(k+1)} \|w\|_{\underline{L}^2(Q_r)} = 0,\ \partial_t w - \nabla\cdot(\bar{\boldsymbol{a}}\nabla w) = 0 \text{ in } Q_\infty \Big\}.$$

It turns out that $\bar{\mathcal{A}}_k(Q_\infty)$ coincides with the set of *$\bar{\boldsymbol{a}}$-caloric polynomials*[2] of parabolic degree at most k. That is,

$$\bar{\mathcal{A}}_k(Q_\infty) = \Big\{ w|_{Q_\infty} \,:\, w \in \mathcal{P}_k(\mathbb{R}\times\mathbb{R}^d),\ \partial_t w - \nabla\cdot(\bar{\boldsymbol{a}}\nabla w) = 0 \text{ in } \mathbb{R}\times\mathbb{R}^d \Big\}. \tag{1-23}$$

The vector space of n-homogeneous $\bar{\boldsymbol{a}}$-caloric polynomials is isomorphic to that of n-homogeneous polynomials of $\mathbb{R}^d$. This can be shown by backwards uniqueness and the fact that this vector space is spanned by products of homogeneous $\bar{\boldsymbol{a}}$-caloric polynomials depending only on t and one of the space variables; see for instance [Widder 1961] or [Nualart 2006, Proposition 1.1.1]. In any case, we have that $\dim(\bar{\mathcal{A}}_k(Q_\infty)) = \binom{d+k}{d} < \infty$.

In the next result, we generalize the parabolic Liouville theorem implicit in (1-23) to $\boldsymbol{a}(x)$-caloric functions. At the same time we provide a *quantitative* version of this Liouville principle, in other words, a $C^{k,1}$-type regularity estimate. Define, for every parabolic cylinder $I\times U \subseteq \mathbb{R}\times\mathbb{R}^d$,

$$\mathcal{A}(I\times U) := \{ w \in H^1_{\mathrm{par,\,loc}}(I\times U) : \partial_t w - \nabla\cdot(\boldsymbol{a}\nabla w) = 0 \text{ in } I\times U \}$$

and, for every $k\in\mathbb{N}$,

$$\mathcal{A}_k(Q_\infty) := \Big\{ w \in \mathcal{A}(Q_\infty) : \limsup_{r\to\infty} r^{-(k+1)} \|w\|_{\underline{L}^2(Q_r)} = 0 \Big\}.$$

Note that these vector spaces are *random* since they depend on $\boldsymbol{a}\in\Omega$. The following theorem is a parabolic analogue of [Armstrong et al. 2017b, Theorem 3.6].

Theorem 1.2 (parabolic higher regularity theory). *Fix $s\in(0,2+d)$. There exist an exponent $\delta(s,d,\Lambda)\in\big(0,\frac12\big]$ and a random variable $\mathcal{X}_s$ satisfying the estimate*

$$\mathcal{X}_s \le \mathcal{O}_s(C(s,d,\Lambda)) \tag{1-24}$$

such that the following statements hold, for every $k\in\mathbb{N}$:

(i)$_k$ *There exists $C(k,d,\Lambda)<\infty$ such that, for every $u\in\mathcal{A}_k(Q_\infty)$, there exists $p\in\bar{\mathcal{A}}_k(Q_\infty)$ such that for every $R\ge\mathcal{X}_s$,*

$$\|u-p\|_{\underline{L}^2(Q_R)} \le CR^{-\delta}\|p\|_{\underline{L}^2(Q_R)}. \tag{1-25}$$

(ii)$_k$ *For every $p\in\bar{\mathcal{A}}_k(Q_\infty)$, there exists $u\in\mathcal{A}_k(Q_\infty)$ satisfying (1-25) for every $R\ge\mathcal{X}_s$.*

(iii)$_k$ *There exists $C(k,d,\Lambda)<\infty$ such that, for every $R\ge\mathcal{X}_s$ and $u\in\mathcal{A}(Q_R)$, there exists $\phi\in\mathcal{A}_k(Q_\infty)$ such that, for every $r\in[\mathcal{X}_s,R]$, we have the estimate*

$$\|u-\phi\|_{\underline{L}^2(Q_r)} \le C\left(\frac{r}{R}\right)^{k+1}\|u\|_{\underline{L}^2(Q_R)}. \tag{1-26}$$

[2] $\bar{\boldsymbol{a}}$-caloric polynomials are often called *heat polynomials* in the literature, in the case $\bar{\boldsymbol{a}} = I_d$.

In particular, we have, $\mathbb{P}$-almost surely, for every $k \in \mathbb{N}$,

$$\dim(\mathcal{A}_k(Q_\infty)) = \dim(\bar{\mathcal{A}}_k(Q_\infty)) = \binom{d+k}{d}. \tag{1-27}$$

Observe that, as in the elliptic case, even for $k = 0$ the third statement of Theorem 1.2 gives us an important gradient estimate on solutions. Indeed, the combination of statement $(\mathrm{iii})_0$ and the Caccioppoli inequality yields that, for every $R \geq \mathcal{X}_s$, $u \in \mathcal{A}(Q_R)$ and $r \in [\mathcal{X}_s, R]$, we have

$$\|\nabla u\|_{\underline{L}^2(Q_r)} \leq C\|\nabla u\|_{\underline{L}^2(Q_R)}.$$

This should be seen as a $C^{0,1}$-type estimate and compared to pointwise gradient bounds for the solutions of the heat equation.

The proof of Theorem 1.2 is obtained as a consequence of Theorem 1.1 and a routine adaptation of the proof of [Armstrong et al. 2017b, Theorem 3.6], which is the statement of the analogous result in the elliptic case. In Section 6, we explain the modifications required in the parabolic setting.

Soon after the first version of this paper was submitted and posted to the arXiv, a new preprint of Bella, Chiarini and Fehrman [Bella et al. 2018] appeared which contains a large-scale regularity result which has some overlap with Theorem 1.2. In particular, under qualitative assumptions, they obtain the statement of Theorem 1.2 in the case $k = 1$ with the estimate (1-24) on $\mathcal{X}_s$ replaced by the qualitative bound $\mathbb{P}[\mathcal{X}_s < \infty] = 1$.

1E. ***Outline of the paper.*** In the next section, we introduce the subadditive quantities inherited from the variational structure of the equation and record some of their basic properties. For convenience, the variational formulation of uniformly parabolic equations is recalled in a self-contained presentation in Appendix A. In Section 3, we present several functional inequalities which are needed later in the paper. Of particular interest are inequalities giving us control of certain weak norms of functions in terms of the spatial averages of the functions in cubes as well as Caccioppoli-type inequalities giving us control of strong norms of solutions in terms of weak norms. Section 4 is the heart of the paper, where we prove the convergence of the subadditive quantities by an iteration over the length scales. In Section 5, we demonstrate how to pass from control of the convergence of the subadditive quantities to general homogenization results. Finally, in Section 6 we summarize the passage from the quantitative homogenization results to the higher regularity theory (which is entirely analogous to the elliptic setting). In Appendix B, we give local and global versions of the Meyers higher integrability estimate for gradients of solutions. We remark that the statement of the global Meyers estimate we prove appears to be new and somewhat sharper compared to what has previously appeared in the literature.

2. Variational structure and subadditive quantities

2A. ***Variational formulation of parabolic equations.*** As we now explain, the solution of a parabolic equation can be obtained as the minimizer of a uniformly convex functional. This is an entirely deterministic statement, valid for an arbitrary fixed coefficient field $\boldsymbol{a} \in \Omega$.

The following proposition states the solvability of parabolic equations. It relies on convex analysis and calculus of variations, and is close to the main result of [Ghoussoub and Tzou 2004]; see also [Ghoussoub

2009]. We provide a self-contained proof in Appendix A in the more general setting of maximal monotone operators, and for a larger set of pairs (w, w^*); see Proposition A.1.

Proposition 2.1 (parabolic variational principle). *Let $\mathcal{J}$ be defined below in (2-4). For each $w \in H^1_{\mathrm{par}}(I \times U)$ and $w^* \in L^2(I; H^{-1}(U))$, the mapping*

$$w + H^1_{\mathrm{par},\sqcup}(I \times U) \to \mathbb{R}, \qquad u \mapsto \mathcal{J}[u, w^*],$$

is uniformly convex. Moreover, its minimum is zero, and the associated minimizer is the unique $u \in w + H^1_{\mathrm{par},\sqcup}(I \times U)$ that is a solution of

$$(\partial_t - \nabla \cdot \boldsymbol{a} \nabla) u = w^* \quad \text{in } I \times U. \tag{2-1}$$

Equation (2-1) is interpreted as

$$\text{for all } \phi \in L^2(I; H^1_0(U)), \quad \int_{I\times U} \nabla\phi \cdot \boldsymbol{a} \nabla u = \int_{I\times U} \phi(w^* - \partial_t u). \tag{2-2}$$

The left side of (2-2) can be more explicitly written as

$$\int_{I\times U} \nabla\phi(t,x) \cdot \boldsymbol{a}(t,x)\, \nabla u(t,x)\, dt\, dx,$$

while the right side of (2-2) could be more properly written as

$$\int_I \langle \phi(t,\cdot), (w^* - \partial_t u)(t,\cdot)\rangle\, dt,$$

with $\langle \cdot, \cdot \rangle$ the duality pairing between $H^1_0(U)$ and $H^{-1}(U)$.

We proceed to define the functional $\mathcal{J}$ appearing in Proposition 2.1. To start, we decompose the matrix $\boldsymbol{a}$ into its symmetric and skew-symmetric parts:

$$\boldsymbol{s}(t,x) := \frac{\boldsymbol{a}(t,x) + \boldsymbol{a}^{\mathsf{t}}(t,x)}{2}, \quad \boldsymbol{m}(x) := \frac{\boldsymbol{a}(t,x) - \boldsymbol{a}^{\mathsf{t}}(t,x)}{2},$$

and set

$$A(p,q,t,x) := \tfrac{1}{2} p \cdot \boldsymbol{s}(t,x) p + \tfrac{1}{2}(q - \boldsymbol{m}(t,x)p) \cdot \boldsymbol{s}^{-1}(t,x)(q - \boldsymbol{m}(t,x)p), \tag{2-3}$$

so that the following lemma holds.

Lemma 2.2. *There exists a constant $C(\Lambda) < \infty$ such that, for every $(t,x) \in \mathbb{R} \times \mathbb{R}^d$,*

$$(p,q) \mapsto A(p,q,t,x) - C^{-1}(|p|^2 + |q|^2) \quad \text{is convex,}$$

and

$$(p,q) \mapsto A(p,q,t,x) - C(|p|^2 + |q|^2) \quad \text{is concave.}$$

Moreover, for every $(t,x) \in \mathbb{R} \times \mathbb{R}^d$ and $p, q \in \mathbb{R}^d$,

$$A(p,q,t,x) \geq p \cdot q,$$

with equality if and only if $q = \boldsymbol{a}(t,x)p$.

Proof. We briefly recall the proof; see also [Armstrong and Mourrat 2016, (2.6)]. The fact that $(p, q) \mapsto A(p, q, t, x)$ is uniformly convex and $C^{1,1}$ follows from the definition of Ω in (1-5). The second part of the lemma is a consequence of the identity

$$A(p, q, t, x) - p \cdot q = \tfrac{1}{2}(\boldsymbol{a}(t, x)p - q)\, \boldsymbol{s}^{-1}(t, x)(\boldsymbol{a}(t, x)p - q). \qquad \square$$

The functional $\mathcal{J}$ appearing in Proposition 2.1 is defined, for every $u \in H^1_{\mathrm{par}}(I \times U)$ and $u^* \in L^2(I; H^{-1}(U))$, by

$$\mathcal{J}[u, u^*] := \inf\left\{ \int_{I\times U} (A(\nabla u, \boldsymbol{g}, \cdot) - \nabla u \cdot \boldsymbol{g}) : -\nabla \cdot \boldsymbol{g} = u^* - \partial_t u \right\}. \tag{2-4}$$

In the infimum above, we understand that $\boldsymbol{g} \in L^2(I \times U; \mathbb{R}^d)$, and the last condition is interpreted as

$$\text{for all } \phi \in L^2(I; H^1_0(U)), \quad \int_{I\times U} \nabla\phi \cdot \boldsymbol{g} = \int_{I\times U} \phi(u^* - \partial_t u).$$

In the integral on the right side of (2-4), the dot in the expression $A(\nabla u, \boldsymbol{g}, \cdot)$ stands for the time-space variable; that is,

$$\int_{I\times U} (A(\nabla u, \boldsymbol{g}, \cdot) - \nabla u \cdot \boldsymbol{g}) = \int_I \int_U \big(A(\nabla u(t, x), \boldsymbol{g}(t, x), t, x) - \nabla u(t, x) \cdot \boldsymbol{g}(t, x)\big)\, dx\, dt.$$

2B. ***Subadditive quantities and basic properties.*** In this subsection, we define the subadditive quantities and collect their basic properties. Although their definitions are actually very natural and intuitive, many readers will not find them to be on first reading. In order to understand the motivation for studying them, it is best to first have some familiarity with the elliptic case with symmetric coefficients, which is described in [Armstrong et al. 2017b]. Indeed, much of what appears below can be compared to Chapter 2 of that book, and in fact this paper can be seen as a generalization of Chapters 1–3 of that book to the parabolic setting. Now, since the subadditive quantities are endowed from the variational structure of the equation, it is natural that the parabolic versions should be somewhat more complicated than the elliptic ones. A similar issue was encountered in [Armstrong and Mourrat 2016], where subadditive quantities were defined and analyzed for "nonvariational" elliptic equations.

In any case, the most convincing demonstration that these are the "right" quantities will have to wait until Section 5, where we prove that quantitative information about the convergence of the subadditive quantities can be translated directly into control of the first-order correctors and therefore into estimates on the rate of homogenization.

Without further ado, we give the definitions of the subadditive quantities. For every Lipschitz domain $U \subseteq \mathbb{R}^d$, bounded interval $I \subseteq \mathbb{R}$ and $p, q \in \mathbb{R}^d$, we define

$$\mu(I \times U, p, q) := \inf_{(\nabla v, \boldsymbol{h}) \in \mathcal{C}_0(I\times U)} \fint_{I\times U} A(p + \nabla v, q + \boldsymbol{h}, \cdot), \tag{2-5}$$

where the infimum is taken over $(\nabla v, \boldsymbol{h})$ ranging in the space

$$\mathcal{C}_0(I\times U) := \Big\{ (\nabla v, \boldsymbol{h}) \in L^2(I\times U, \mathbb{R}^d)^2 : v \in H^1_{\mathrm{par},0}(I\times U) \text{ and } \\ \text{for all } \phi \in L^2(I; H^1(U)),\ \int_{I\times U} \nabla\phi\cdot\boldsymbol{h} = -\int_{I\times U} \phi\,\partial_t v \Big\}. \tag{2-6}$$

Since $\phi \in L^2(I; H^1(U))$ and $\partial_t v \in L^2(I; \widehat{H}^{-1}(U))$, the last integral is well-defined, in the usual interpretation as

$$\int_I \langle \phi(t,\cdot), \partial_t v(t,\cdot)\rangle\, dt,$$

where here $\langle\cdot,\cdot\rangle$ denotes the duality pairing between $H^1(U)$ and $\widehat{H}^{-1}(U)$. Testing the condition in (2-6) against the function $\phi(t,x) := p\cdot x$ and integrating by parts in time, we see that any candidate $\boldsymbol{h}$ must satisfy

$$\int_{I\times U} \boldsymbol{h} = 0. \tag{2-7}$$

The dual subadditive quantity μ^* is defined, for every $p^*, q^* \in \mathbb{R}^d$, by

$$\mu^*(I\times U, q^*, p^*) := \sup_{(\nabla u, \boldsymbol{g})\in\mathcal{C}(I\times U)} \fint_{I\times U} \big(-A(\nabla u, \boldsymbol{g}, \cdot) + q^*\cdot\nabla u + p^*\cdot\boldsymbol{g}\big), \tag{2-8}$$

where the supremum is taken over $(\nabla u, \boldsymbol{g})$ ranging in the space

$$\mathcal{C}(I\times U) := \Big\{ (\nabla u, \boldsymbol{g}) \in L^2(I\times U; \mathbb{R}^d)^2 : u \in H^1_{\mathrm{par}}(I\times U), \\ \text{for all } \phi \in L^2(I; H^1_0(U)),\ \int_{I\times U} \nabla\phi\cdot\boldsymbol{g} = -\int_{I\times U} \phi\,\partial_t u \Big\}. \tag{2-9}$$

Note that for each $p\in\mathbb{R}^d$ and $(\nabla v, \boldsymbol{h}) \in \mathcal{C}_0(I\times U)$, we have $(p+\nabla v, q+\boldsymbol{h}) \in \mathcal{C}(I\times U)$. Using also (1-12) and (2-7), we thus deduce that for every $p, q, p^*, q^* \in \mathbb{R}^d$,

$$\mu^*(I\times U, q^*, p^*) \ge q^*\cdot p + p^*\cdot q - \mu(I\times U, p, q). \tag{2-10}$$

That is, the function $(q^*, p^*) \mapsto \mu^*(I\times U, q^*, p^*)$ is bounded below by the convex dual of the function $(p,q)\mapsto \mu(I\times U, p, q)$. As in the elliptic case, see [Armstrong et al. 2016, Lemma 3.1; 2017a], we will combine μ and μ^* into a master quantity denoted by J which monitors the defect in this convex duality pairing. For concision, we set

$$V := I\times U, \tag{2-11}$$

and define a $2d$-by-$2d$ matrix field $\boldsymbol{A} : \mathbb{R}\times\mathbb{R}^d \to \mathbb{R}^{2d\times 2d}$ by

$$\boldsymbol{A}(t,x) := \begin{pmatrix} \boldsymbol{s}(t,x) - \boldsymbol{m}(t,x)\,\boldsymbol{s}^{-1}(t,x)\,\boldsymbol{m}(t,x) & \boldsymbol{m}(t,x)\,\boldsymbol{s}^{-1}(t,x) \\ -\boldsymbol{s}^{-1}(t,x)\,\boldsymbol{m}(t,x) & \boldsymbol{s}^{-1}(t,x) \end{pmatrix},$$

so that

$$A(p,q,t,x) = \frac{1}{2}\begin{pmatrix} p \\ q \end{pmatrix} \cdot \boldsymbol{A} \begin{pmatrix} p \\ q \end{pmatrix}.$$

This notation allows us to rewrite the definitions of μ and μ^* in (2-5) and (2-8) in more compact notation: for every $X, X^* \in \mathbb{R}^{2d}$, we have

$$\mu(V, X) = \inf_{S \in X + \mathcal{C}_0(V)} \fint_V \tfrac{1}{2} S \cdot \boldsymbol{A} S, \tag{2-12}$$

$$\mu^*(V, X^*) = \sup_{S \in \mathcal{C}(V)} \fint_V \left(-\tfrac{1}{2} S \cdot \boldsymbol{A} S + X^* \cdot S\right), \tag{2-13}$$

and the inequality (2-10) can be rewritten as

$$\mu^*(V, X^*) \geq X \cdot X^* - \mu(V, X). \tag{2-14}$$

We now set

$$\mathcal{S}(V) := \left\{ (\nabla v, \boldsymbol{h}) \in \mathcal{C}(V) \,:\, \text{for all } (\nabla \phi, \boldsymbol{f}) \in \mathcal{C}_0(V),\ \int_V \begin{pmatrix} \nabla \phi \\ \boldsymbol{f} \end{pmatrix} \cdot \boldsymbol{A} \begin{pmatrix} \nabla v \\ \boldsymbol{h} \end{pmatrix} = 0 \right\}, \tag{2-15}$$

and for every $X, X^* \in \mathbb{R}^{2d}$,

$$J(V, X, X^*) := \sup_{S \in \mathcal{S}(V)} \fint_V \left(-\tfrac{1}{2} S \cdot \boldsymbol{A} S - X \cdot \boldsymbol{A} S + X^* \cdot S\right). \tag{2-16}$$

The master quantity J can be rewritten in the following more explicit notation:

$$J\left(V, \begin{pmatrix} p \\ q \end{pmatrix}, \begin{pmatrix} q^* \\ p^* \end{pmatrix}\right) := \sup_{(\nabla v, \boldsymbol{g}) \in \mathcal{S}(V)} \fint_V \left(-\frac{1}{2} \begin{pmatrix} \nabla v \\ \boldsymbol{g} \end{pmatrix} \cdot \boldsymbol{A} \begin{pmatrix} \nabla v \\ \boldsymbol{g} \end{pmatrix} - \begin{pmatrix} p \\ q \end{pmatrix} \cdot \boldsymbol{A} \begin{pmatrix} \nabla v \\ \boldsymbol{g} \end{pmatrix} + \begin{pmatrix} q^* \\ p^* \end{pmatrix} \cdot \begin{pmatrix} \nabla v \\ \boldsymbol{g} \end{pmatrix} \right). \tag{2-17}$$

The next lemma shows that J indeed monitors the defect in convex duality between μ and μ^*.

Lemma 2.3. *For every $X, X^* \in \mathbb{R}^{2d}$,*

$$J(V, X, X^*) = \mu(V, X) + \mu^*(V, X^*) - X \cdot X^*. \tag{2-18}$$

Moreover, the maximizer $S(\cdot, V, X, X^*)$ in (2-16) is the difference between the maximizer of $\mu^*(V, X^*)$ in (2-13) and the minimizer of $\mu(V, X)$ in (2-12).

Proof. We first argue that, for every $X^* \in \mathbb{R}^{2d}$,

$$\mu^*(V, X^*) = \sup_{S \in \mathcal{S}(V)} \fint_V \left(-\tfrac{1}{2} S \cdot \boldsymbol{A} S + X^* \cdot S\right). \tag{2-19}$$

Let $S^* \in \mathcal{C}(V)$ denote the maximizer in the definition of $\mu^*(V, X^*)$. Note that, for every $X^* \in \mathbb{R}^{2d}$ and $S \in \mathcal{C}_0(V)$,

$$\fint_U X^* \cdot S = 0. \tag{2-20}$$

By the first variation for μ^*, we deduce that

$$\text{for all } S' \in \mathcal{C}_0(V), \quad \int_V S' \cdot \boldsymbol{A} S^* = 0.$$

That is, $S^* \in \mathcal{S}(U)$, and thus (2-19) holds.

Let $X = (p, q) \in \mathbb{R}^{2d}$ and

$$S_0 = (\nabla v, \boldsymbol{h}) \in X + \mathcal{C}_0(V) \tag{2-21}$$

denote the minimizer in the definition of $\mu(V, p, q) = \mu(V, X)$. For every $S \in \mathcal{S}(V)$,

$$\mu(V, X) + \fint_V \big(X^* \cdot S - \tfrac{1}{2} S \cdot \boldsymbol{A} S\big) - X \cdot X^* = \fint_V \big(\tfrac{1}{2} S_0 \cdot \boldsymbol{A} S_0 - \tfrac{1}{2} S \cdot \boldsymbol{A} S + X^* \cdot S\big) - X \cdot X^*. \tag{2-22}$$

By (2-21), we have $X = \fint_V S_0$. For each $S \in \mathcal{S}(V)$,

$$\fint_V S \cdot \boldsymbol{A} S_0 = \fint_V S \cdot \boldsymbol{A} X,$$

and this last identity holds true in particular for $S = S_0$. We obtain that the left side of (2-22) is equal to

$$\fint_V \big(-\tfrac{1}{2}(S - S_0) \cdot \boldsymbol{A}(S - S_0) - X \cdot \boldsymbol{A}(S - S_0) + X^* \cdot (S - S_0)\big).$$

We compare this result to the identities (2-19) and (2-16) to obtain the lemma. □

The next lemma collects elementary properties of J and its minimizer. It can be compared with [Armstrong et al. 2017b, Lemma 2.2].

Lemma 2.4 (basic properties of J). *The quantity $J(V, X, X^*)$ and its maximizer $S(\cdot, V, X, X^*)$ satisfy the following properties*:

- *The mapping $(X, X^*) \mapsto J(V, X, X^*)$ is quadratic.*
- ***Uniformly convex and $C^{1,1}$ in X and X^* separately.*** *There exists a constant $C(d, \Lambda) < \infty$ such that, for every $X_1, X_2, X^* \in \mathbb{R}^{2d}$,*

$$\frac{1}{C}|X_1 - X_2|^2 \le \tfrac{1}{2} J(V, X_1, X^*) + \tfrac{1}{2} J(V, X_2, X^*) - J\big(V, \tfrac{1}{2}X_1 + \tfrac{1}{2}X_2, X^*\big) \le C|X_1 - X_2|^2 \tag{2-23}$$

and, for every $X_1^*, X_2^*, X \in \mathbb{R}^{2d}$,

$$\frac{1}{C}|X_1^* - X_2^*|^2 \le \tfrac{1}{2} J(V, X, X_1^*) + \tfrac{1}{2} J(V, X, X_2^*) - J\big(V, X, \tfrac{1}{2}X_1^* + \tfrac{1}{2}X_2^*\big) \le C|X_1^* - X_2^*|^2. \tag{2-24}$$

- ***Subadditivity.*** *Let $V_1, \ldots, V_N \subseteq V$ be parabolic cylinders that partition V, in the sense that $V_i \cap V_j = \varnothing$ if $i \ne j$ and*

$$\Big|V \setminus \bigcup_{i=1}^N V_i\Big| = 0.$$

For every $X, X^* \in \mathbb{R}^{2d}$, we have

$$J(V, X, X^*) \le \sum_{i=1}^N \frac{|V_i|}{|V|} J(V_i, X, X^*). \tag{2-25}$$

- ***First variation for J.*** *For $X, X^* \in \mathbb{R}^{2d}$, the function $S(\cdot, V, X, X^*)$ is the unique element of $\mathcal{S}(V)$ such that*

$$\text{for all } T \in \mathcal{S}(V), \quad \fint_V T \cdot \boldsymbol{A} S(\cdot, V, X, X^*) = \fint_V (-X \cdot \boldsymbol{A} T + X^* \cdot T). \tag{2-26}$$

- ***Quadratic response.*** *For every $X, X^* \in \mathbb{R}^{2d}$ and $T \in \mathcal{S}(V)$,*

$$\frac{1}{C} \fint_V |T - S(\cdot, V, X, X^*)|^2 \le J(V, X, X^*) - \fint_V \left(-\tfrac{1}{2} T \cdot \boldsymbol{A} T - X \cdot \boldsymbol{A} T + X^* \cdot T\right) \le C \fint_V |T - S(\cdot, V, X, X^*)|^2. \tag{2-27}$$

- ***Formulas for derivatives of J.*** *For every $X, X^* \in \mathbb{R}^{2d}$,*

$$\nabla_X J(V, X, X^*) = -\fint_V \boldsymbol{A} S(\cdot, V, X, X^*), \tag{2-28}$$

$$\nabla_{X^*} J(V, X, X^*) = \fint_V S(\cdot, V, X, X^*). \tag{2-29}$$

Proof. Since these properties are easy to check and their proofs are almost the same of those of [Armstrong et al. 2017b, Lemma 2.2], we omit the details. □

Remark 2.5. Since $X^* \mapsto \mu^*(V, X^*)$ is a quadratic form, we obtain from (2-28) and Lemma 2.3 that

$$\fint_V S(\cdot, V, X, 0) = \nabla_{X^*} J(V, X, 0) = -X, \tag{2-30}$$

a property which also follows directly from the definition of μ in (2-12) and the identification of $-S(\cdot, V, X, 0)$ as the minimizer in this definition. From (2-29) and Lemma 2.3, we also obtain the dual identity

$$\fint_V \boldsymbol{A} S(\cdot, V, 0, X^*) = X^*.$$

In the next lemma, we relate the space $\mathcal{S}(I \times U)$ with the space of solutions of the parabolic equation and of its dual. Define the vector space $\mathcal{A}(I \times U)$ to be the set of weak solutions $u \in H^1_{\mathrm{par}}(I \times U)$ of the equation

$$\partial_t u - \nabla \cdot (\boldsymbol{a} \nabla u) = 0 \quad \text{in } I \times U,$$

and the vector space $\mathcal{A}^*(I \times U)$ to be the set of weak solutions $u^* \in H^1_{\mathrm{par}}(I \times U)$ of the dual equation

$$\partial_t u^* + \nabla \cdot (\boldsymbol{a}^{\mathrm{t}} \nabla u^*) = 0 \quad \text{in } I \times U.$$

Note that the direction of time is reversed in the dual equation. Precisely,

$$\mathcal{A}(I \times U) := \left\{ u \in H^1_{\mathrm{par}}(I \times U) \,:\, \text{for all } w \in L^2(I; H^1_0(U)),\ \int_{I \times U} w\, \partial_t u = -\int_{I \times U} \nabla w \cdot \boldsymbol{a} \nabla u \right\},$$

$$\mathcal{A}^*(I \times U) := \left\{ u^* \in H^1_{\mathrm{par}}(I \times U) \,:\, \text{for all } w \in L^2(I; H^1_0(U)),\ \int_{I \times U} w\, \partial_t u^* = \int_{I \times U} \nabla w \cdot \boldsymbol{a}^{\mathrm{t}} \nabla u^* \right\}.$$

Lemma 2.6. *We have*

$$\mathcal{S}(V) := \{(\nabla u + \nabla u^*, \boldsymbol{a} \nabla u - \boldsymbol{a}^{\mathrm{t}} \nabla u^*) \,:\, u \in \mathcal{A}(V),\ u^* \in \mathcal{A}^*(V)\}. \tag{2-31}$$

Proof. Recall that $V = I \times U$, and denote by $\mathcal{S}'(I \times U)$ the set on the right side of (2-31). The condition

$$\int_{I\times U} (\nabla\phi, \boldsymbol{f}) \cdot \boldsymbol{A}(\nabla v, \boldsymbol{h}) = 0$$

appearing in (2-15) can be rewritten more explicitly as

$$\int_{I\times U} \big(\nabla\phi \cdot \boldsymbol{s}\nabla v + (\boldsymbol{f} - \boldsymbol{m}\nabla\phi) \cdot \boldsymbol{s}^{-1}(\boldsymbol{h} - \boldsymbol{m}\nabla v)\big) = 0. \tag{2-32}$$

We first verify that $\mathcal{S}(I \times U) \subseteq \mathcal{S}'(I \times U)$. The space $\{0\} \times L^2(I, L^2_{\mathrm{sol},0}(U))$ is a subspace of $\mathcal{C}_0(I \times U)$. Hence, if (2-32) holds for every $(\nabla\phi, \boldsymbol{h}) \in \mathcal{C}_0(I \times U)$, then in particular

$$\text{for all } \boldsymbol{f} \in L^2(I; L^2_{\mathrm{sol},0}(U)), \quad \int_{I\times U} \boldsymbol{f} \cdot \boldsymbol{s}^{-1}(\boldsymbol{h} - \boldsymbol{m}\nabla v) = 0.$$

In other words, $\boldsymbol{s}^{-1}(\boldsymbol{h} - \boldsymbol{m}\nabla v)$ belongs to the space orthogonal to $L^2(I; L^2_{\mathrm{sol},0}(U))$ in $L^2(I \times U)$. That is, there exists $w \in L^2(I; H^1(U))$ such that

$$\boldsymbol{s}^{-1}(\boldsymbol{h} - \boldsymbol{m}\nabla v) = \nabla w,$$

and we deduce that for every $(\nabla\phi, \boldsymbol{f}) \in \mathcal{C}_0(I \times U)$,

$$\int_{I\times U} (\nabla\phi \cdot (\boldsymbol{s}\nabla v + \boldsymbol{m}\nabla w) - w\, \partial_t\phi) = 0.$$

Denoting by Δ_N^{-1} the solution operator for the Laplace equation on U with null Neumann boundary condition, we observe that for each $\phi \in H^1_{\mathrm{par},0}(I \times U)$, the pair $(\nabla\phi, \nabla\Delta_N^{-1}(\partial_t\phi))$ belongs to $\mathcal{C}_0(I \times U)$. The identity above therefore holds for arbitrary $\phi \in H^1_{\mathrm{par},0}(I \times U)$, and we thus deduce that $\partial_t w \in L^2(I, H^{-1}(U))$. We can then integrate by parts in time and obtain that

$$\text{for all } \phi \in H^1_{\mathrm{par},0}(I \times U), \quad \int_{I\times U} (\nabla\phi \cdot (\boldsymbol{s}\nabla v + \boldsymbol{m}\nabla w) + \phi\, \partial_t w) = 0. \tag{2-33}$$

This property can be extended to arbitrary $\phi \in L^2(I; H^1_0(U))$ by density. The additional requirement that $(\nabla v, \boldsymbol{h}) \in \mathcal{C}(I \times U)$ gives

$$\text{for all } \psi \in L^2(I; H^1_0(U)), \quad \int_{I\times U} (\nabla\psi \cdot (\boldsymbol{s}\nabla w + \boldsymbol{m}\nabla v) + \psi\, \partial_t v) = 0. \tag{2-34}$$

Setting

$$u := \tfrac{1}{2}(v + w), \quad u^* := \tfrac{1}{2}(v - w), \tag{2-35}$$

we deduce that $u \in \mathcal{A}(I \times U)$, $u^* \in \mathcal{A}^*(I \times U)$, with

$$v = \tfrac{1}{2}(u + u^*), \quad \boldsymbol{h} = \boldsymbol{a}\nabla u - \boldsymbol{a}^{\mathrm{t}}\nabla u^*,$$

and this completes the proof that $\mathcal{S}(I \times U) \subseteq \mathcal{S}'(I \times U)$.

Conversely, given $u \in \mathcal{A}(I \times U)$ and $u^* \in \mathcal{A}^*(I \times U)$, we set

$$v = u + u^*, \quad w := u - u^*, \quad \boldsymbol{h} := \boldsymbol{a}\nabla u - \boldsymbol{a}^{\mathrm{t}}\nabla u^* = \boldsymbol{s}\nabla w + \boldsymbol{m}\nabla v,$$

and observe that

$$\boldsymbol{a}\nabla u + \boldsymbol{a}^{\mathrm{t}}\nabla u^* = \boldsymbol{s}\nabla v + \boldsymbol{m}\nabla w.$$

The identities (2-33) and (2-34) follow. This implies that the condition (2-32) is satisfied for every $(\nabla\phi,\, \boldsymbol{f}) \in \mathcal{C}_0(I\times U)$, and hence that $(\nabla v,\, \boldsymbol{h}) \in \mathcal{S}(I\times U)$. We have thus shown that $\mathcal{S}'(I\times U) \subseteq \mathcal{S}(I\times U)$, which completes the proof. □

Remark 2.7. Note that for $S = (\nabla u + \nabla u^*,\, \boldsymbol{a}\nabla u - \boldsymbol{a}^{\mathrm{t}}\nabla u^*) \in \mathcal{S}(V)$, we have

$$\boldsymbol{A}S = (\boldsymbol{a}\nabla u + \boldsymbol{a}^{\mathrm{t}}\nabla u^*,\, \nabla u - \nabla u^*). \tag{2-36}$$

Indeed, (2-36) is implicit in the proof of Lemma 2.6 above and can also be checked by a direct computation. In particular, ∇u can be written as one half of the sum of the first component of S and second component of $\boldsymbol{A}S$, and $\boldsymbol{a}\nabla u$ can be recovered similarly. This observation is needed in Section 5 in the construction of (approximate) correctors.

3. Functional inequalities

We collect here some functional inequalities which will be useful in the rest of the paper. The two main results are a "multiscale" version of the Poincaré inequality, and a Caccioppoli-type inequality for elements of $\mathcal{S}(\square_n)$. The proof of the latter is based on a parabolic version of the Helmholtz–Hodge decomposition of vector fields, which is of independent interest.

We first recall a useful version of the Poincaré inequality, for functions of the space variable only.

Lemma 3.1. *Let $\psi \in L^2(\square_n)$ satisfy*

$$\fint_{\square_n} \psi = 1.$$

There exists $C(d) < \infty$ such that, for every $u \in H^1(\square_n)$,

$$\left\| u - \fint_{\square_n} u\,\psi \right\|_{\underline{L}^2(\square_n)} \le C\|\psi\|_{\underline{L}^2(\square_n)}\|\nabla u\|_{\underline{\widehat{H}}^{-1}(\square_n)}. \tag{3-1}$$

Proof. By the usual Poincaré inequality, all we need to show is that

$$\left| \fint_{\square_n} u(1-\psi) \right| \le C\|\psi\|_{\underline{L}^2(\square_n)}\|\nabla u\|_{\underline{H}^{-1}(\square_n)}. \tag{3-2}$$

Let w be the solution of the Neumann problem

$$\begin{cases} -\Delta w = 1-\psi & \text{in } \square_n, \\ \boldsymbol{n}\cdot\nabla w = 0 & \text{on } \partial\square_n. \end{cases}$$

Notice that this has a solution because $\int_{\square_n}(1-\psi) = 0$, and we have the H^2 estimate, see for instance [Armstrong et al. 2017b, Lemma B.18],

$$\|\nabla w\|_{\underline{H}^1(\square_n)} \le C\|1-\psi\|_{\underline{L}^2(\square_n)} \le C(1+\|\psi\|_{\underline{L}^2(\square_n)}) \le C\|\psi\|_{\underline{L}^2(\square_n)}.$$

Testing the equation for w by u thus yields

$$\left| \fint_{\square_n} u(1-\psi) \right| = \left| \fint_{\square_n} \nabla u \cdot \nabla w \right| \le C \|\nabla u\|_{\underline{\widehat{H}}^{-1}(\square_n)} \|\nabla w\|_{\underline{H}^1(\square_n)} \le C \|\psi\|_{\underline{L}^2(\square_n)} \|\nabla u\|_{\underline{\widehat{H}}^{-1}(\square_n)}. \qquad \square$$

For every parabolic cylinder V and $f \in L^1(V)$, we recall that we use the following shorthand notation for the spatial average of f over V:

$$(f)_V := \fint_V f. \tag{3-3}$$

By the standard Poincaré inequality in $1+d$ coordinates, we have

$$\|u - (u)_{\boxdot_n}\|_{\underline{L}^2(\boxdot_n)} \le C 3^n \|\nabla u\|_{\underline{L}^2(\boxdot_n)} + C 3^{2n} \|\partial_t u\|_{\underline{L}^2(\boxdot_n)}. \tag{3-4}$$

In the context of parabolic equations, it is natural to try to preserve a matching between the number of times a function is differentiated in space and *half* the number of times it is differentiated in time. The estimate (3-4) is not consistent with this scaling. The purpose of the next proposition is to obtain such a bound — see also Corollary 3.4 below.

Proposition 3.2. *There exists $C(d) < \infty$ such that, for every $u \in H^1_{\mathrm{par}}(\boxdot_n)$ and $\boldsymbol{g} \in L^2(\boxdot_n; \mathbb{R}^d)$ satisfying $\partial_t u = \nabla \cdot \boldsymbol{g}$, we have*

$$\|u - (u)_{\boxdot_n}\|_{\underline{L}^2(\boxdot_n)} \le C\big(\|\nabla u\|_{\underline{L}^2(I_n; \underline{\widehat{H}}^{-1}(\square_n))} + \|\boldsymbol{g}\|_{\underline{L}^1(I_n; \underline{H}^{-1}(\square_n))}\big).$$

Remark 3.3. In the statement of Proposition 3.2 (and similarly for Corollary 3.4 and Proposition 3.7 below), the condition $\partial_t u = \nabla \cdot \boldsymbol{g}$ is interpreted as

$$\text{for all } \phi \in L^2(I_n; H^1_0(\square_n)), \quad \int_{\boxdot_n} \nabla \phi \cdot \boldsymbol{g} = -\int_{\boxdot_n} \phi \, \partial_t u.$$

Equivalently, this amounts to saying that $(\nabla u, \boldsymbol{g}) \in \mathcal{C}(\boxdot_n)$. As an example, we can always take $\boldsymbol{g} = \nabla \Delta_{\square_n}^{-1} \partial_t u$, where $\Delta_{\square_n}^{-1}$ is the solution operator for the Laplacian in $\square_n$ with null Dirichlet boundary condition.

Proof of Proposition 3.2. Let $\psi \in C_c^\infty(\square_n)$ be a smooth function of compact support in $\square_n$ such that $\fint_{\square_n} \psi(x)\,dx = 1$, $0 \le \psi \le 2$, and

$$3^{-n} \|\nabla \psi\|_{L^\infty(\square_n)} + \|\nabla^2 \psi\|_{L^\infty(\square_n)} \le C 3^{-2n}. \tag{3-5}$$

We write $(u)_{\square_n, \psi} := \fint_{\square_n} u(x)\,\psi(x)\,dx$ and $(u)_{\square_n} := \fint_{\square_n} u(x)\,dx$. Using the Poincaré inequality (in the form given by Lemma 3.1) in time slices gives, for every $t \in I_n$,

$$\|u(t,\cdot) - (u(t,\cdot))_{\square_n,\psi}\|_{\underline{L}^2(\square_n)} \le C \|\nabla u(t,\cdot)\|_{\underline{\widehat{H}}^{-1}(\square_n)}.$$

Thus

$$\fint_{I_n} \fint_{\square_n} |u(t,x) - (u(t,\cdot))_{\square_n,\psi}|^2 \,dx\,dt \le C \fint_{I_n} \|\nabla u(t,\cdot)\|^2_{\underline{\widehat{H}}^{-1}(\square_n)} \,dt. \tag{3-6}$$

Since $\psi \in H_0^1(U)$ and $\nabla\psi \in H_0^1(U;\mathbb{R}^d)$, we have, for every $t \in I_n$,

$$|\partial_t (u(t,\cdot))_{\square_n,\psi}| = \left| \fint_{\square_n} \psi(x)\, \partial_t u(t,x)\, dx \right| = \left| \fint_{\square_n} \nabla\psi(x) \cdot \boldsymbol{g}(t,\cdot)(x)\, dx \right|$$
$$\le C \|\nabla\psi\|_{\underline{H}^1(\square_n)} \|\boldsymbol{g}(t,\cdot)\|_{\underline{H}^{-1}(\square_n)} \le C 3^{-2n} \|\boldsymbol{g}(t,\cdot)\|_{\underline{H}^{-1}(\square_n)}.$$

Thus

$$\sup_{t\in I_n} \left| (u(t,\cdot))_{\square_n,\psi} - \fint_{I_n\times\square_n} \psi(x) u(t,x)\, dx\, dt \right| \le \int_{I_n} |\partial_t (u(t,\cdot))_{\square_n,\psi}|\, dt$$
$$\le C 3^{-2n} \int_{I_n} \|\boldsymbol{g}(t,\cdot)\|_{\underline{H}^{-1}(\square_n)}\, dt = C \fint_{I_n} \|\boldsymbol{g}(t,\cdot)\|_{\underline{H}^{-1}(\square_n)}\, dt.$$

Combining this with (3-6), we obtain

$$\fint_{I_n} \fint_{\square_n} \left| u(t,x) - \fint_{I_n\times\square_n} \psi(y) u(s,y)\, ds\, dy \right|^2 dx\, dt$$
$$\le C \fint_{I_n} \|\nabla u(t,\cdot)\|^2_{\underline{\widehat{H}}^{-1}(\square_n)}\, dt + \left(\fint_{I_n} \|\boldsymbol{g}(t,\cdot)\|_{\underline{H}^{-1}(\square_n)}\, dt \right)^2.$$

Since

$$\|u - (u)_{\square_n}\|_{\underline{L}^2(\square_n)} = \inf_{c\in\mathbb{R}} \|u - c\|_{\underline{L}^2(\square_n)},$$

this yields the announced result. □

Corollary 3.4. *There exists $C(d) < \infty$ such that, for every $u \in H^1_{\mathrm{par}}(\square_n)$ and $\boldsymbol{g} \in L^2(\square_n,\mathbb{R}^d)$ satisfying $\partial_t u = \nabla\cdot\boldsymbol{g}$,*

$$\|u - (u)_{\square_n}\|_{\underline{L}^2(\square_n)} \le C 3^n (\|\nabla u\|_{\underline{L}^2(\square_n)} + \|\boldsymbol{g}\|_{\underline{L}^2(\square_n)}). \tag{3-7}$$

Proof. This follows from Proposition 3.2 and the inequalities

$$\|v\|_{\underline{L}^2(I_n;\underline{H}^{-1}(\square_n))} \le \|v\|_{\underline{L}^2(I_n;\underline{\widehat{H}}^{-1}(\square_n))} \le C 3^n \|v\|_{\underline{L}^2(\square_n)}.$$ □

Remark 3.5. Recall from Remark 3.3 that in the statement of Corollary 3.4, we can take $\boldsymbol{g} = \nabla\Delta^{-1}_{\square_n}\partial_t u$. Moreover, there exists $C(d) < \infty$ such that, for every $f \in L^2(I_n; H^{-1}(\square_n))$,

$$C^{-1} \|f\|_{\underline{L}^2(I_n;\underline{H}^{-1}(\square_n))} \le \|\nabla\Delta^{-1}_{\square_n} f\|_{\underline{L}^2(\square_n)} \le C \|f\|_{\underline{L}^2(I_n;\underline{H}^{-1}(\square_n))}.$$

Indeed, by the standard Poincaré inequality, the norm $\|\cdot\|_{\underline{H}^1(\square_n)}$ is equivalent to the norm $u \mapsto \|\nabla u\|_{\underline{L}^2(\square_n)}$ on $H_0^1(\square_n)$, and moreover,

$$\sup\left\{ \fint_{\square_n} f\phi \,:\, \phi \in L^2(I_n; H_0^1(\square_n)),\ \|\nabla\phi\|_{L^2(\square_n)} \le 1 \right\}$$
$$= \sup\left\{ \fint_{\square_n} \nabla\phi\cdot\nabla\Delta^{-1}_{\square_n} f \,:\, \phi \in L^2(I_n; H_0^1(\square_n)),\ \|\nabla\phi\|_{L^2(\square_n)} \le 1 \right\}$$
$$= \|\nabla\Delta^{-1}_{\square_n} f\|_{\underline{L}^2(\square_n)}.$$

In particular, on the right side of (3-7), we can replace the term $\|\boldsymbol{g}\|_{\underline{L}^2(\square_n)}$ with $\|\partial_t u\|_{\underline{L}^2(I_n;\underline{H}^{-1}(\square_n))}$.

The next proposition allows us to obtain a control of the $H^{-1}_{\mathrm{par}}(V)$ norm of a function from knowledge of its spatial averages over large scales. For each $m, n \in \mathbb{N}$, $m \le n$, we set

$$\mathcal{Z}_m := [(3^{2m}\mathbb{Z}) \times (3^m\mathbb{Z}^d)] \cap \boxdot_n. \tag{3-8}$$

Although $\mathcal{Z}_m$ depends on n, we keep this dependence implicit in the notation, since its identity will be clear from the context. This is a parabolic version of the inequality which first appeared in [Armstrong et al. 2016, Proposition 6.1].

Proposition 3.6 (multiscale Poincaré inequality). *There exists $C(d, \Lambda) < \infty$ such that, for every $n \ge 1$ and $f \in L^2(\boxdot_n)$,*

$$\|f\|_{\underline{\widehat{H}}^{-1}_{\mathrm{par}}(\boxdot_n)} \le C\|f\|_{\underline{L}^2(\boxdot_n)} + C\sum_{m=0}^{n-1} 3^m \left(|\mathcal{Z}_m|^{-1} \sum_{z\in\mathcal{Z}_m} |(f)_{z+\boxdot_m}|^2 \right)^{\frac{1}{2}}.$$

Proof of Proposition 3.6. Recalling (1-15), we fix $g \in H^1_{\mathrm{par}}(\boxdot_n)$ such that

$$\|g\|_{\underline{H}^1_{\mathrm{par}}(\boxdot_n)} \le 1, \tag{3-9}$$

and decompose the proof into two steps.

Step 1. In this step, we show that there exists a constant $C(d) < \infty$ such that, for every $m \in \{0, \dots, n\}$,

$$\sum_{z\in\mathcal{Z}_m} \int_{z+\boxdot_m} |g - (g)_{z+\boxdot_m}|^2 \le C|\boxdot_n|\, 3^{2m}. \tag{3-10}$$

By Corollary 3.4 and Remark 3.5, the left side above is bounded by

$$C3^{2m}\,|\boxdot_m| \sum_{z\in\mathcal{Z}_m} (\|\nabla g\|^2_{\underline{L}^2(z+\boxdot_m)} + \|\partial_t g\|^2_{\underline{L}^2(z_0+I_m, \underline{H}^{-1}(z'+\square_m))}),$$

where we write $z = (z_0, z') \in \mathbb{Z} \times \mathbb{Z}^d$. The contribution of the first term is easily estimated, since by (3-9),

$$|\boxdot_m| \sum_{z\in\mathcal{Z}_m} \|\nabla g\|^2_{\underline{L}^2(z+\boxdot_m)} = \int_{\boxdot_n} |\nabla g|^2 \le |\boxdot_n|.$$

For the second term, we write

$$\sum_{z\in\mathcal{Z}_m} \|\partial_t g\|^2_{\underline{L}^2(z_0+I_m, \underline{H}^{-1}(z'+\square_m))} = \sup\left\{ \sum_{z\in\mathcal{Z}_m} \left(\fint_{z+\boxdot_m} \phi_z\, \partial_t g \right)^2 : \phi_z \in L^2(z_0+I_m; H^1_0(z'+\square_m)),\ \|\nabla\phi_z\|_{\underline{L}^2(z+\boxdot_m)} \le 1 \right\}.$$

For ϕ_z satisfying the conditions in the supremum above, we have

$$\begin{aligned} |\boxdot_n|^{-1} \sum_{z\in\mathcal{Z}_m} \left(\int_{z+\boxdot_m} \phi_z\, \partial_t g \right)^2 &= \fint_{\boxdot_n} \partial_t g \left(\sum_{z\in\mathcal{Z}_m} \phi_z \int_{z+\boxdot_n} \phi_z\, \partial_t g \right) \\ &\le \|\partial_t g\|_{\underline{L}^2(I_n; \underline{H}^{-1}(\square_n))} \left\| \sum_{z\in\mathcal{Z}_m} \phi_z \int_{z+\boxdot_n} \phi_z\, \partial_t g \right\|_{\underline{L}^2(I_n; \underline{H}^1(\square_n))}. \end{aligned}$$

Notice that, by (3-9), the first term on the right side of the previous inequality is bounded by 1. Moreover, by the normalization of the functions ϕ_z,

$$\left\| \sum_{z\in\mathcal{Z}_m} \phi_z \int_{z+\square_n} \phi_z \, \partial_t g \right\|^2_{\underline{L}^2(I_n;\underline{H}^1(\square_n))} \le C \left\| \sum_{z\in\mathcal{Z}_m} \nabla\phi_z \int_{z+\square_n} \phi_z \, \partial_t g \right\|^2_{\underline{L}^2(\square_n)} \le C \frac{|\square_m|}{|\square_n|} \sum_{z\in\mathcal{Z}_m} \left(\int_{z+\square_m} \phi_z \, \partial_t g \right)^2.$$

Combining the last three displays, we arrive at

$$\sum_{z\in\mathcal{Z}_m} \|\partial_t g\|^2_{\underline{L}^2(z_0+I_m,\underline{H}^{-1}(z'+\square_m))} \le C \frac{|\square_n|}{|\square_m|},$$

and this completes the proof of (3-10).

Step 2. We aim to control $⨍_{\square_n} fg$, which we decompose into

$$⨍_{\square_n} fg = ⨍_{\square_n} f(g-(g)_{\square_n}) + (f)_{\square_n}(g)_{\square_n}. \tag{3-11}$$

By the definition of the H^1_{par} norm in (1-10), we have $(g)_{\square_n} \le 3^n$, and therefore, by Jensen's inequality,

$$(f)_{\square_n}(g)_{\square_n} \le 3^n |(f)_{\square_n}| \le C3^{n-1} \left(|\mathcal{Z}_{n-1}|^{-1} \sum_{z\in\mathcal{Z}_{n-1}} |(f)_{z+\square_{n-1}}|^2 \right)^{\frac12}. \tag{3-12}$$

We then proceed to decompose the first integral on the right side of (3-11) recursively. For every $m \in \{0,\dots,n-1\}$ and $z \in \mathcal{Z}_{m+1}$, we have

$$\begin{aligned} \int_{z+\square_{m+1}} f\cdot(g-(g)_{z+\square_{m+1}}) = {} & \sum_{y\in\mathcal{Z}_m\cap(z+\square_{m+1})} \int_{y+\square_m} f\cdot(g-(g)_{y+\square_m}) \\ & + |\square_m| \sum_{y\in\mathcal{Z}_m\cap(z+\square_{m+1})} ((g)_{y+\square_m}-(g)_{z+\square_{m+1}})\cdot(f)_{y+\square_m}. \end{aligned} \tag{3-13}$$

Summing over $z \in \mathcal{Z}_{m+1}$ and using Hölder's inequality, we get

$$\begin{aligned} \sum_{z\in\mathcal{Z}_{m+1}} \int_{z+\square_{m+1}} f\cdot(g-(g)_{z+\square_{m+1}}) \le {} & \sum_{y\in\mathcal{Z}_m} \int_{y+\square_m} f\cdot(g-(g)_{y+\square_m}) \\ & + |\square_m| \left(\sum_{\substack{z\in\mathcal{Z}_{m+1}\\ y\in\mathcal{Z}_m\cap(z+\square_{m+1})}} |(g)_{y+\square_m}-(g)_{z+\square_{m+1}}|^2 \right)^{\frac12} \left(\sum_{y\in\mathcal{Z}_m} |(f)_{y+\square_m}|^2 \right)^{\frac12}. \end{aligned}$$

By Jensen's inequality, we have, for each $z \in \mathcal{Z}_{m+1}$,

$$\sum_{y\in\mathcal{Z}_m\cap(z+\square_{m+1})} |(g)_{y+\square_m}-(g)_{z+\square_{m+1}}|^2 \le \sum_{y\in\mathcal{Z}_m\cap(z+\square_{m+1})} ⨍_{y+\square_m} |g-(g)_{z+\square_{m+1}}|^2,$$

and thus, by (3-10),

$$\sum_{\substack{z\in\mathcal{Z}_{m+1}\\ y\in\mathcal{Z}_m\cap(z+\square_{m+1})}} |(g)_{y+\square_m}-(g)_{z+\square_{m+1}}|^2 \le C \frac{|\square_n|}{|\square_m|} 3^{2m}.$$

Using also that $|\mathcal{Z}_m| = |\square_n|/|\square_m|$ and combining with (3-13), we obtain

$$\sum_{z\in\mathcal{Z}_{m+1}} \int_{z+\square_{m+1}} f\cdot(g-(g)_{z+\square_{m+1}}) \le \sum_{y\in\mathcal{Z}_m} \int_{y+\square_m} f\cdot(g-(g)_{y+\square_m}) + C\,|\square_n|\,3^m \left(|\mathcal{Z}_m|^{-1}\sum_{y\in\mathcal{Z}_m} |(f)_{y+\square_m}|^2\right)^{\frac12}.$$

Summing over $m\in\{0,\dots,n-1\}$ yields

$$\int_{\square_n} f(g-(g)_{\square_n}) \le \sum_{z\in\mathcal{Z}_0} \int_{z+\square_0} f\cdot(g-(g)_{z+\square_0}) + C\,|\square_m| \sum_{m=0}^{n-1} 3^m \left(|\mathcal{Z}_m|^{-1}\sum_{y\in\mathcal{Z}_m} |(f)_{y+\square_m}|^2\right)^{\frac12}.$$

Hence, by Hölder's inequality and (3-10),

$$\int_{\square_n} f(g-(g)_{\square_n}) \le C|\square_n|^{\frac12}\left(\int_{\square_n}|f|^2\right)^{\frac12} + C\,|\square_n| \sum_{m=0}^{n-1} 3^m \left(|\mathcal{Z}_m|^{-1}\sum_{y\in\mathcal{Z}_m} |(f)_{y+\square_m}|^2\right)^{\frac12}.$$

Dividing by $|\square_n|$ and combining with (3-11)–(3-12), we obtain

$$\fint_{\square_n} fg \le C\|f\|_{\underline{L}^2(\square_n)} + C\sum_{m=0}^{n} 3^m \left(|\mathcal{Z}_m|^{-1}\sum_{y\in\mathcal{Z}_m} |(f)_{y+\square_m}|^2\right)^{\frac12}.$$

Taking the supremum over all g satisfying (3-9) yields the result. □

The name "multiscale Poincaré inequality" for Proposition 3.6 is best understood in conjunction with the following statement.

Proposition 3.7. *There exists a constant $C(d)<\infty$ such that, for every $n\in\mathbb{N}$, $u\in H^1_{\mathrm{par}}(\square_{n+1})$ and $\boldsymbol{g}\in L^2(\square_{n+1};\mathbb{R}^d)$ satisfying $\partial_t u = \nabla\cdot\boldsymbol{g}$, we have*

$$\|u-(u)_{\square_n}\|_{\underline{L}^2(\square_n)} \le C\|\nabla u\|_{\underline{H}^{-1}_{\mathrm{par}}(\square_{n+1})} + C\|\boldsymbol{g}\|_{\underline{H}^{-1}_{\mathrm{par}}(\square_{n+1})}.$$

Remark 3.8. It is clear that the proof of Proposition 3.7 can be adapted to show that for every $r>0$, there exists a constant $C(r,d)<\infty$ such that, for every $n\in\mathbb{N}$, $u\in H^1_{\mathrm{par}}(\square_n)$ and $\boldsymbol{g}\in L^2(\square_n;\mathbb{R}^d)$ satisfying $\partial_t u = \nabla\cdot\boldsymbol{g}$, we have

$$\|u-(u)_{\square_{n-r}}\|_{\underline{L}^2(\square_{n-r})} \le C\|\nabla u\|_{\underline{H}^{-1}_{\mathrm{par}}(\square_n)} + C\|\boldsymbol{g}\|_{\underline{H}^{-1}_{\mathrm{par}}(\square_n)}.$$

Although one can expect that the estimate above still holds for $r=0$, we leave it as an open question here, and content ourselves with an interior estimate.

Combining Propositions 3.6 and 3.7 allows to estimate the (interior) L^2 oscillation of u in terms of spatial averages of ∇u and $\boldsymbol{g}$; see also [Armstrong et al. 2016, Proposition 6.1]. The estimate yields better interior information than the "single-scale" Poincaré inequality provided by Proposition 3.2 as soon as the spatial averages of ∇u and $\boldsymbol{g}$ display nontrivial cancellations over large scales. This feature will be crucial to our subsequent arguments.

Before turning to the proof of Proposition 3.7, we recall the classical H^2 estimate for solutions of the heat equation. For simplicity, we state it using periodic boundary conditions in the space variable. We denote the corresponding function spaces by $H^1_{\#}(U)$, $H^1_{\mathrm{par},\#}(I\times U)$, etc.

Lemma 3.9 (H^2 estimate for the Cauchy problem). *There exists $C(d) < \infty$ such that, for every $n \in \mathbb{N}$, if $f \in L^2(\boxdot_n)$ and $u \in H^1_{\mathrm{par},\#}(\boxdot_n)$ solves the Cauchy problem*

$$\begin{cases} (\partial_t - \Delta)u = f & \text{in } \boxdot_n, \\ u = 0 & \text{on } \{-3^{2n}/2\} \times \square_n, \end{cases} \tag{3-14}$$

then $\nabla u \in H^1_{\mathrm{par},\#}(\boxdot_n)$ and

$$3^{-n}\|u\|_{\underline{H}^1_{\mathrm{par},\#}(\boxdot_n)} + \|\nabla u\|_{\underline{H}^1_{\mathrm{par},\#}(\boxdot_n)} \le C\|f\|_{\underline{L}^2(\boxdot_n)}. \tag{3-15}$$

Proof. By scaling, it suffices to prove the result for $n = 0$. For each $s \in I_0$, we test (3-14) against the function $(t, x) \mapsto \mathbb{1}_{t<s}u(t, x)$ and get

$$\int_{(-\frac{1}{2},s)\times\square_0} (u\,\partial_t u + |\nabla u|^2) = \int_{(-\frac{1}{2},s)\times\square_0} uf,$$

which implies

$$\tfrac{1}{2}\|u(s,\cdot)\|^2_{L^2(\square_0)} + \int_{(-\frac{1}{2},s)\times\square_0} |\nabla u|^2 \le \|u\|_{L^2(\boxdot_0)}\|f\|_{L^2(\boxdot_0)}.$$

Taking the supremum over $s \in I_0$, observing that

$$\sup_{s\in I_0} \|u(s,\cdot)\|_{L^2(\square_0)} \le \|u\|_{L^2(\boxdot_0)}$$

and using Young's inequality, we obtain

$$\|u\|_{L^2(\boxdot_0)} + \|\nabla u\|_{L^2(\boxdot_0)} \le C\|f\|_{L^2(\boxdot_0)}.$$

We now turn to the estimation of $\|\nabla^2 u\|_{L^2(\boxdot_0)}$. We first observe that by integration by parts, we have $\|\nabla^2 u\|_{L^2(\boxdot_0)} = \|\Delta u\|_{L^2(\boxdot_0)}$. Moreover, using (3-14), we get

$$\|\Delta u\|^2_{L^2(\boxdot_0)} = \int_{\boxdot_0} \Delta u(f + \partial_t u) = \int_{\boxdot_0} f\,\Delta u - \int_{\boxdot_0} \nabla u \cdot \partial_t \nabla u,$$

with

$$\int_{\boxdot_0} \nabla u \cdot \partial_t \nabla u = \tfrac{1}{2}\big\|\nabla u\big(\tfrac{1}{2},\cdot\big)\big\|^2_{L^2(\square_0)} \ge 0,$$

and therefore

$$\|\Delta u\|_{L^2(\square_0)} \le \|f\|_{L^2(\square_0)}. \tag{3-16}$$

We also need bounds on the time derivatives of u and ∇u. Note that

$$\begin{aligned} \|\partial_t \nabla u\|_{L^2(I_0;H^{-1}_\#(\square_0))} &= \sup\left\{ \int_{\boxdot_0} \partial_t \nabla u \cdot F : \|F\|_{L^2(I_0;H^1_\#(\square_0))} \le 1 \right\} \\ &= \sup\left\{ \int_{\boxdot_0} \partial_t u\, \nabla \cdot F : \|F\|_{L^2(I_0;H^1_\#(\square_0))} \le 1 \right\} \le \|\partial_t u\|_{L^2(\boxdot_0)}, \end{aligned}$$

and we can estimate the L^2-norm of $\partial_t u$ using (3-14) and (3-16):

$$\|\partial_t u\|_{L^2(\boxdot_0)} = \|f + \Delta u\|_{L^2(\boxdot_0)} \le 2\|f\|_{L^2(\boxdot_0)}.$$

The obvious bound

$$\|\partial_t u\|_{L^2(I_0;H^{-1}_\#(\square_0))} \le C\|\partial_t u\|_{L^2(\boxdot_0)}$$

completes the proof of (3-15). □

Proof of Proposition 3.7. By homogeneity, it suffices to show the result for $n=0$. Let $\psi \in C_c^\infty(\mathbb{R}^{1+d})$ be a smooth function with compact support in $\boxdot_{-2}$ and such that $\int_{\mathbb{R}^{1+d}} \psi = 1$. We decompose the proof into three steps.

Step 1. Let $\eta \in C_c^\infty(\boxdot_1)$ be a smooth function with compact support in $\boxdot_{3/4}$ and such that $\eta \equiv 1$ on $\boxdot_{1/2}$. In this step, we show that there exists a constant $C(d,\psi,\eta) < \infty$ such that, for every $u \in L^2(\boxdot_1)$,

$$\|\eta(u - u\star\psi)\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)} \le C(\|\nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)} + \|\boldsymbol{g}\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)}). \tag{3-17}$$

For each $n \in \mathbb{N}$, we set

$$\psi_n := 3^{n(2+d)}\psi(3^{2n}t, 3^n x).$$

Since the function $\psi_{-1} - \psi_0$ is compactly supported and of mean zero, we can use, e.g., [Armstrong et al. 2017b, Lemma 5.7] (in $1+d$ dimensions) to rewrite it in the form

$$\psi_{-1} - \psi_0 = \partial_t \boldsymbol{h}^\circ + \nabla\cdot\boldsymbol{h}',$$

where $\boldsymbol{h} = (\boldsymbol{h}^\circ, \boldsymbol{h}') \in C_c^\infty(\mathbb{R}^{1+d}; \mathbb{R}\times\mathbb{R}^d)$ is supported in $\boxdot_{-2}$ (with $\boldsymbol{h}^\circ$ taking values in $\mathbb{R}$ and $\boldsymbol{h}'$ in $\mathbb{R}^d$). For each $n \in \mathbb{N}$, we define

$$\boldsymbol{h}_n(t,x) = (\boldsymbol{h}_n^\circ, \boldsymbol{h}_n')(t,x) := 3^{n(2+d)}\boldsymbol{h}(3^{2n}t, 3^n x),$$

so that

$$\psi_{n-1} - \psi_n = 3^{-2n}\,\partial_t \boldsymbol{h}_n^\circ + 3^{-n}\,\nabla\cdot\boldsymbol{h}_n'.$$

Since $u\star\psi_n \to u$ in $L^2(\boxdot_{3/4})$, we can use the triangle inequality to bound

$$\|\eta(u - u\star\psi)\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)} \le \sum_{n=0}^{+\infty} \|\eta(u\star\psi_{n-1} - u\star\psi_n)\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)}. \tag{3-18}$$

We next observe that, for every $z \in \boxdot_{3/4}$,

$$\begin{aligned}
(u\star\psi_{n-1} - u\star\psi_n)(z) &= \int_{\boxdot_1} u(y)(3^{-2n}\partial_t \boldsymbol{h}_n^\circ + 3^{-n}\,\nabla\cdot\boldsymbol{h}_n')(z-y)\,dy \\
&= \int_{\boxdot_1} (3^{-2n}\partial_t u(y)\boldsymbol{h}_n^\circ(z-y) + 3^{-n}\,\nabla u(y)\cdot\boldsymbol{h}_n'(z-y))\,dy \\
&= 3^{-n}\int_{\boxdot_1} \nabla u(y)\cdot\boldsymbol{h}_n'(z-y)\,dy - 3^{-2n}\int_{\boxdot_1} \boldsymbol{g}(y)\cdot\nabla\boldsymbol{h}_n^\circ(z-y)\,dy.
\end{aligned}$$

We fix $f \in H^1_{\mathrm{par}}(\boxdot_1)$, set $\tilde f := \eta f$, and compute

$$\int_{\boxdot_1} \eta(u\star\psi_{n-1} - u\star\psi_n) f = 3^{-n}\int_{\boxdot_1} \nabla u\cdot(\tilde f * \boldsymbol{h}_n') - 3^{-2n}\int_{\boxdot_1} \boldsymbol{g}\cdot(\tilde f * \nabla\boldsymbol{h}_n^\circ).$$

One can check that there exists a constant $C(d, \psi, \boldsymbol{h}) < \infty$ such that

$$\|\tilde{f} * \boldsymbol{h}_n'\|_{H^1_{\mathrm{par}}(\boxdot_1)} + 3^{-n}\|\tilde{f} * \nabla \boldsymbol{h}_n^\circ\|_{H^1_{\mathrm{par}}(\boxdot_1)} \le C\|f\|_{H^1_{\mathrm{par}}(\boxdot_1)}.$$

Summarizing, we have thus shown that

$$\|\eta(u \star \psi_{n-1} - u \star \psi_n)\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)} \le C3^{-n}(\|\nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)} + \|\boldsymbol{g}\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)}).$$

Summing over $n \in \mathbb{N}$ in (3-18), we obtain (3-17).

Step 2. Define

$$c(u) := u \star \psi(0).$$

In this step, we show that there exists a constant $C(d, \psi, \eta) < \infty$ such that

$$\|\eta(u - c(u))\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)} \le C(\|\nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)} + \|\boldsymbol{g}\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)}). \tag{3-19}$$

This is an immediate consequence of the fact that there exists a constant $C(d, \psi) < \infty$ such that, for every $y, z \in \boxdot_{3/4}$,

$$|u \star \psi(z) - u \star \psi(y)| \le C(\|\nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)} + \|\boldsymbol{g}\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)}). \tag{3-20}$$

The proof of (3-20) is very similar to the previous step, only simpler: we represent the function $\psi(\cdot - z) - \psi(\cdot - y)$ in the form

$$\partial_t \tilde{\boldsymbol{h}}^\circ + \nabla \cdot \tilde{\boldsymbol{h}}',$$

with $(\tilde{\boldsymbol{h}}^\circ, \tilde{\boldsymbol{h}}') \in C_c^\infty(\boxdot_1; \mathbb{R} \times \mathbb{R}^d)$, and then obtain (3-20) thanks to an integration by parts.

Step 3. For concision, we write

$$\tilde{u} := u - c(u).$$

Let $\chi \in C_c^\infty(\boxdot_1)$ be a smooth function with compact support in $\boxdot_{1/2}$ and such that $\chi \equiv 1$ on $\boxdot_0$. In this step, we show that there exists a constant $C(d, \psi, \eta, \chi) < \infty$ such that

$$\|\chi\tilde{u}\|_{L^2(\boxdot_1)} \le C(\|\nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)} + \|\boldsymbol{g}\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)}). \tag{3-21}$$

Let $w \in H^1_{\mathrm{par},\#}(\boxdot_1)$ solve the Cauchy problem

$$\begin{cases} (\partial_t - \Delta)w = \chi\tilde{u} & \text{in } \boxdot_1, \\ w = 0 & \text{on } \left\{-\frac{9}{2}\right\} \times \square_1. \end{cases} \tag{3-22}$$

By Lemma 3.9, there exists a constant $C(d) < \infty$ such that

$$\|w\|_{H^1_{\mathrm{par},\#}(\boxdot_1)} + \|\nabla w\|_{H^1_{\mathrm{par},\#}(\boxdot_1)} \le C\|\chi\tilde{u}\|_{L^2(\boxdot_1)}. \tag{3-23}$$

Testing (3-22) against $\chi\tilde{u}$ and integrating by parts gives

$$\begin{aligned}\|\chi\tilde{u}\|^2_{L^2(\boxdot_1)} &= \int_{\boxdot_1} (\nabla w\cdot\nabla(\chi\tilde{u}) - w\,\partial_t(\chi\tilde{u}))\\ &= \int_{\boxdot_1} (\nabla w\cdot\nabla(\chi\tilde{u}) + \nabla(\chi w)\cdot\boldsymbol{g} - \partial_t\chi\, w\tilde{u})\\ &\le \|\nabla w\|_{H^1_{\#,\mathrm{par}(\boxdot_1)}}(\|\nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)} + \|\tilde{u}\nabla\chi\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)} + \|\boldsymbol{g}\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)})\\ &\qquad + \|w\|_{H^1_{\#,\mathrm{par}(\boxdot_1)}}(\|\boldsymbol{g}\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)} + \|\tilde{u}\nabla\chi\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)}).\end{aligned}$$

Using the result of the previous step and (3-23), we obtain (3-21). This completes the proof of Proposition 3.7, since

$$\|u-(u)_{\boxdot_0}\|_{L^2(\boxdot_0)} \le \|u-c(u)\|_{L^2(\boxdot_0)} \le \|\chi(u-c(u))\|_{L^2(\boxdot_0)}. \qquad \square$$

Finally, we prove a Caccioppoli-type inequality for elements of $\mathcal{S}(\boxdot_n)$.

Proposition 3.10. *There exists a constant $C(d,\Lambda)<\infty$ such that, for every $n\in\mathbb{N}$ and $S\in\mathcal{S}(\boxdot_{n+1})$,*

$$\|S\|_{\underline{L}^2(\boxdot_n)} \le C3^{-n}\|S\|_{\underline{\widehat{H}}^{-1}_{\mathrm{par}}(\boxdot_{n+1})}.$$

In order to prove this result, we first describe more explicitly the structure of elements of $\widehat{H}^{-1}_{\mathrm{par}}(I\times U)$.

Lemma 3.11 (identification of $\widehat{H}^{-1}_{\mathrm{par}}$). *There exists a constant $C(I,U,d)<\infty$ and, for each $u^*\in L^2(I\times U)$, a pair $(w,w^*)\in L^2(I;H^1_0(U))\times L^2(I;H^{-1}(U))$ such that*

$$u^* = \partial_t w + w^*, \tag{3-24}$$

with

$$\|w\|_{L^2(I;H^1(U))} + \|w^*\|_{L^2(I;H^{-1}(U))} \le C\|u^*\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)}. \tag{3-25}$$

Let $V\subseteq V'\subseteq I\times U$ be subdomains of $I\times U$ such that $\overline{V}\subseteq V'$. If u^* vanishes outside of V, then there exists a pair $(w,w^*)\in L^2(I;H^1_0(U))\times L^2(I;H^{-1}(U))$ satisfying (3-24)–(3-25) for a constant $C(V,V',I,U,d)<\infty$, and such that w and w^* vanish outside of V'.

Proof. Denote by $\Delta^{-1}_{\square_n}$ the solution operator for the Laplacian in $\square_n$ with null Dirichlet boundary condition. We observe that

$$(u,v)\mapsto \fint_{I\times U}(|U|^{-\frac{2}{d}}uv + \nabla u\cdot\nabla v + \nabla\Delta_U^{-1}\partial_t u\cdot\nabla\Delta_U^{-1}\partial_t v)$$

is a scalar product for the Hilbert space $H^1_{\mathrm{par}}(I\times U)$. By the Riesz representation theorem, there exists a unique $u\in H^1_{\mathrm{par}}(I\times U)$ such that, for every $v\in H^1_{\mathrm{par}}(I\times U)$,

$$\fint_{I\times U} u^*v = \fint_{I\times U}(|U|^{-\frac{2}{d}}uv + \nabla u\cdot\nabla v + \nabla\Delta_U^{-1}\partial_t u\cdot\nabla\Delta_U^{-1}\partial_t v),$$

and moreover, by testing this identity with $v=u$, we obtain

$$\|u\|_{H^1_{\mathrm{par}}(I\times U)} \le C\|u^*\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)}. \tag{3-26}$$

We set

$$w := \Delta_U^{-1}\partial_t u \quad \text{and} \quad w^* := |U|^{-\frac{2}{d}}u - \Delta u.$$

The estimate (3-25) follows from (3-26). For $v \in H^1_{\mathrm{par}}(I \times U)$ with compact support in $I \times U$, we have

$$\fint_{I\times U} u^* v = \fint_{I\times U} (w^* v + \partial_t w\, v).$$

Since $u^* \in L^2(I \times U)$, we can argue by density to infer that $w^* + \partial_t w \in L^2(I \times U)$. The identity above then implies (3-24).

If u^* vanishes outside of $V \subseteq I \times U$, then we select a smooth cutoff function η such that $\eta \equiv 1$ on V and $\eta \equiv 0$ outside of V', and we write

$$u^* = \eta u^* = \eta(\partial_t w + w^*) = \partial_t(\eta w) + \eta w^* - w\partial_t \eta.$$

This decomposition yields the second part of the statement, by standard comparisons of norms. □

Proof of Proposition 3.10. By scaling, we may fix $n = 0$. In order to localize an element $S \in \mathcal{S}(\boxdot_1)$ into an element of $\mathcal{C}_0(\boxdot_1)$ and thus be able to use the orthogonality property in the definition of the set $\mathcal{S}(\boxdot_1)$, see (2-15), we introduce a version of the Helmholtz–Hodge decomposition of S which is adapted to the parabolic setting. In order to minimize difficulties due to boundary conditions, we work with functions which are periodic in the space variable. In the course of the proof, we will use the elementary variant of Proposition 2.1 for the standard heat operator with space-periodic boundary condition.

We decompose the proof into four steps. The first two steps are devoted to the construction of the Helmholtz–Hodge decomposition of S, and its estimation in relevant norms. The last step uses this representation to localize S to an element of $\mathcal{C}_0(\boxdot_1)$ and concludes the proof.

Step 1. We write $S = (\nabla u, \boldsymbol{g}) \in \mathcal{S}(\boxdot_1)$. We recall that since $\mathcal{S}(\boxdot_1) \subseteq \mathcal{C}(\boxdot_1)$, we have $\partial_t u = \nabla \cdot \boldsymbol{g}$; see (2-9). The function u is determined up to an additive constant, which we fix so that

$$\fint_{\boxdot_{3/4}} u = 0.$$

Let $\eta \in C_c^\infty(\boxdot_1)$ be a smooth function with compact support in $\boxdot_{3/4}$ such that $0 \le \eta \le 1$ and $\eta \equiv 1$ on $\boxdot_{1/2}$. We set

$$\tilde{u} := \eta u, \qquad \text{and} \qquad \text{for all } j \in \{1, \dots, d\}, \quad \tilde{\boldsymbol{g}}_j := \eta \boldsymbol{g}_j.$$

For each $j \in \{1, \dots, d\}$, let $T_{0j} \in H^1_{\mathrm{par},\#}(\boxdot_1)$ be the unique solution of

$$\begin{cases} (\partial_t - \Delta) T_{0j} = \tilde{\boldsymbol{g}}_j - \partial_{x_j}\tilde{u} & \text{in } \boxdot_1, \\ T_{0j} = 0 & \text{on } \left\{-\frac{9}{2}\right\} \times \square_1. \end{cases} \tag{3-27}$$

By Lemma 3.11 there exist $(w_j, w_j^*) \in L^2(I_1; H^1_0(\square_1)) \times L^2(I_1; H^{-1}(\square_1))$ which vanish in a neighborhood of $\partial\boxdot_1$ and satisfy

$$\tilde{\boldsymbol{g}}_j - \partial_{x_j}\tilde{u} = \partial_t w_j + w_j^*,$$

with

$$\|w_j\|_{L^2(I_1;H^1(\square_1))} + \|w_j^*\|_{L^2(I_1;H^{-1}(\square_1))} \le C\|\tilde{\boldsymbol{g}}_j - \partial_{x_j}\tilde{u}\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)}.$$

Since (w_j, w_j^*) vanish in a neighborhood of $\partial\boxdot_1$, we can interpret this pair as an element of

$$L^2(I_1; H^1_{\#}(\square_1)) \times L^2(I_1; H^{-1}_{\#}(\square_1)),$$

with the estimate

$$\|w_j\|_{L^2(I_1;H^1_{\#}(\square_1))} + \|w_j^*\|_{L^2(I_1;H^{-1}_{\#}(\square_1))} \le C\|\tilde{\boldsymbol{g}}_j - \partial_{x_j}\tilde{u}\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)}.$$

Since $\tilde{\boldsymbol{g}}_j - \partial_{x_j}\tilde{u} \in L^2(\boxdot_1)$, it is clear that $\partial_t w_j \in L^2(I_1; H^{-1}_{\#}(\boxdot_1))$, and we therefore deduce that $w_j \in H^1_{\#,\mathrm{par},\sqcup}(\boxdot_1)$. Moreover, by Proposition 2.1 applied to the standard heat operator, there exist a constant $C(d) < \infty$ and $T'_{0j} \in H^1_{\#,\mathrm{par},\sqcup}(\boxdot_1)$ such that

$$(\partial_t - \Delta)T'_{0j} = -\Delta w_j + w_j^*,$$

with

$$\begin{aligned}\|T'_{0j}\|_{L^2(\boxdot_1)} + \|\nabla T'_{0j}\|_{L^2(\boxdot_1)} &\le C(\|w_j\|_{L^2(I_1;H^1(\square_1))} + \|w_j^*\|_{L^2(I_1;H^{-1}(\square_1))}) \\ &\le C\|\tilde{\boldsymbol{g}}_j - \partial_{x_j}\tilde{u}\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)}.\end{aligned}$$

We thus have

$$(\partial_t - \Delta)(w_j + T'_{0j} - T_{0j}) = 0,$$

with $w_j + T'_{0j} - T_{0j} \in H^1_{\#,\mathrm{par},\sqcup}(\boxdot_1)$. Therefore,

$$T_{0j} = w_j + T'_{0j},$$

and

$$\|T_{0j}\|_{L^2(\boxdot_1)} + \|\nabla T_{0j}\|_{L^2(\boxdot_1)} \le C\|\tilde{\boldsymbol{g}}_j - \partial_{x_j}\tilde{u}\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)}.$$

It is clear that $\|\tilde{\boldsymbol{g}}_j\|_{H^{-1}_{\mathrm{par}}(\boxdot_1)} \le C\|\boldsymbol{g}_j\|_{H^{-1}_{\mathrm{par}}(\boxdot_1)}$. By Proposition 3.7, Remark 3.8 and the comparison

$$\|\tilde{u}\|_{H^{-1}_{\mathrm{par}}(\boxdot_1)} \le \|\tilde{u}\|_{L^2(\boxdot_1)} \le \|u\|_{L^2(\boxdot_{3/4})} = \|u - (u)_{\boxdot_{3/4}}\|_{L^2(\boxdot_{3/4})},$$

we obtain

$$\|T_{0j}\|_{L^2(\boxdot_1)} + \|\nabla T_{0j}\|_{L^2(\boxdot_1)} \le C(\|\nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)} + \|\boldsymbol{g}\|_{\widehat{H}^{-1}_{\mathrm{par}}(\boxdot_1)}). \tag{3-28}$$

Step 2. For each $i, j \in \{1, \dots, d\}$, we define $T_{ij} \in H^1_{\mathrm{par},\#}(\boxdot_1)$ as the solution of

$$\begin{cases}(\partial_t - \Delta)T_{ij} = \partial_{x_i}\tilde{\boldsymbol{g}}_j - \partial_{x_j}\tilde{\boldsymbol{g}}_i & \text{in } \boxdot_1, \\ T_{ij} = 0 & \text{on } \left\{-\frac{9}{2}\right\} \times \square_1.\end{cases}$$

The solution T_{ij} is well-defined since the right-hand side belongs to $L^2(I_1; H^{-1}_{\#}(\square_1))$. We now estimate the L^2 norm of T_{ij} using Lemma 3.9 and duality. We define $\phi_{ij} \in H^1_{\mathrm{par},\#}(\boxdot_1)$ to be the solution of the backwards heat equation

$$\begin{cases}-(\partial_t + \Delta)\phi_{ij} = T_{ij} & \text{in } \boxdot_1, \\ \phi_{ij} = 0 & \text{on } \left\{\frac{9}{2}\right\} \times \square_1.\end{cases} \tag{3-29}$$

By Lemma 3.9, we have

$$\|\nabla\phi_{ij}\|_{H^1_{\mathrm{par},\#}(\square_1)} \le C\|T_{ij}\|_{L^2(\square_1)}.$$

Testing (3-29) against T_{ij}, we get

$$\|T_{ij}\|^2_{L^2(\square_1)} = \int_{\square_1} (\tilde{\boldsymbol{g}}_i\,\partial_{x_j}\phi_{ij} - \tilde{\boldsymbol{g}}_j\,\partial_{x_i}\phi_{ij}) \le \|\nabla\phi_{ij}\|_{H^1_{\mathrm{par},\#}(\square_1)}\,\|\tilde{\boldsymbol{g}}\|_{\widehat{H}^{-1}_{\mathrm{par}}(\square_1)}.$$

Combining the two previous displays yields

$$\|T_{ij}\|_{L^2(\square_1)} \le C\|\tilde{\boldsymbol{g}}\|_{\widehat{H}^{-1}_{\mathrm{par}}(\square_1)} \le \|\boldsymbol{g}\|_{\widehat{H}^{-1}_{\mathrm{par}}(\square_1)}. \tag{3-30}$$

Step 3. For notational convenience, for each $j \in \{1,\dots,d\}$, we set $T_{j0} := -T_{0j}$, $\partial_0 := \partial_t$, $\partial_j := \partial_{x_j}$,

$$R_0 := \tilde{u} - \sum_{j=1}^d \partial_j T_{0j},$$

and, for each $i \in \{1,\dots,d\}$,

$$R_i := \tilde{\boldsymbol{g}}_i - \sum_{j=0}^d \partial_j T_{ij}.$$

In this step, we show that

$$\|R_0\|_{L^2(\square_1)} + \|\nabla R_0\|_{L^2(\square_1)} \le C\|\nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(\square_1)} + C\|\boldsymbol{g}\|_{\widehat{H}^{-1}_{\mathrm{par}}(\square_1)}, \tag{3-31}$$

and that for every $i \in \{1,\dots,d\}$,

$$\|R_i\|_{L^2(\square_1)} \le C\|\nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(\square_1)} + C\|\boldsymbol{g}\|_{\widehat{H}^{-1}_{\mathrm{par}}(\square_1)}. \tag{3-32}$$

Recalling that $\partial_t u = \nabla\cdot\boldsymbol{g}$, we note that

$$(\partial_t - \Delta)R_0 = (\partial_t - \Delta)\tilde{u} - \sum_{j=1}^d \partial_j(\tilde{\boldsymbol{g}}_j - \partial_j\tilde{u}) = u\,\partial_t\eta + \boldsymbol{g}\cdot\nabla\eta,$$

and, for each $i \in \{1,\dots,d\}$,

$$\begin{aligned}(\partial_t - \Delta)R_i &= (\partial_t - \Delta)\tilde{\boldsymbol{g}}_i - \partial_t(\tilde{\boldsymbol{g}}_i - \partial_i\tilde{u}) - \sum_{j=1}^d \partial_j(\partial_i\tilde{\boldsymbol{g}}_j - \partial_j\tilde{\boldsymbol{g}}_i)\\ &= (\partial_i u)(\partial_t\eta) + (\partial_t u)(\partial_i\eta) + (\nabla\cdot\boldsymbol{g})(\partial_i\eta) + (\partial_i\boldsymbol{g})\cdot\nabla\eta.\end{aligned}$$

Moreover, it is clear from their definitions that T_{0j} and T_{ij} vanish in a neighborhood of the initial time slice $\{-\frac{9}{2}\}\times\square_1$. The estimates (3-31) and (3-32) are thus obtained by following the steps to the derivations of (3-28) and (3-30) respectively.

Step 4. We now select a cutoff function $\chi \in C^\infty_c(\square_1)$ such that $\chi \equiv 1$ on $\square_0$ and $\chi \equiv 0$ outside of $\square_{1/2}$, and observe that

$$\begin{pmatrix} \sum_{j=1}^d \nabla\partial_j(\chi^4 T_{0j}) \\ \sum_{j=0}^d \partial_j(\chi^4 T_{ij}) \end{pmatrix} \in \mathcal{C}_0(\square_1),$$

where we understand that the second component above denotes a d-dimensional vector field with components indexed by $i \in \{1, \dots, d\}$. By the definition of $\mathcal{S}(\square_1)$, we deduce that

$$\int_{\square_1} \begin{pmatrix} \sum_{j=1}^d \nabla \partial_j (\chi^4 T_{0j}) \\ \sum_{j=0}^d \partial_j (\chi^4 T_{ij}) \end{pmatrix} \cdot \boldsymbol{A} \begin{pmatrix} \nabla u \\ \boldsymbol{g} \end{pmatrix} = 0. \tag{3-33}$$

In the display above, the first vector is of dimension $2d$: the gradient appearing on the first row carries the first d components, while the other d components are represented by the second row and indexed by $i \in \{1, \dots, d\}$. Applying the chain rule in the identity (3-33) yields a number of terms, one of which is

$$\begin{aligned} \int_{\square_1} \chi^4 \begin{pmatrix} \sum_{j=1}^d \nabla \partial_j T_{0j} \\ \sum_{j=0}^d \partial_j T_{ij} \end{pmatrix} \cdot \boldsymbol{A} \begin{pmatrix} \nabla u \\ \boldsymbol{g} \end{pmatrix} &= \int_{\square_1} \chi^4 \begin{pmatrix} \nabla \tilde{u} \\ \tilde{\boldsymbol{g}} \end{pmatrix} \cdot \boldsymbol{A} \begin{pmatrix} \nabla u \\ \boldsymbol{g} \end{pmatrix} + \int_{\square_1} \chi^4 \begin{pmatrix} \nabla R_0 \\ R_i \end{pmatrix} \cdot \boldsymbol{A} \begin{pmatrix} \nabla u \\ \boldsymbol{g} \end{pmatrix} \\ &= \int_{\square_1} \chi^4 \begin{pmatrix} \nabla u \\ \boldsymbol{g} \end{pmatrix} \cdot \boldsymbol{A} \begin{pmatrix} \nabla u \\ \boldsymbol{g} \end{pmatrix} + \int_{\square_1} \chi^4 \begin{pmatrix} \nabla R_0 \\ R_i \end{pmatrix} \cdot \boldsymbol{A} \begin{pmatrix} \nabla u \\ \boldsymbol{g} \end{pmatrix}. \end{aligned}$$

We are interested in estimating the first term in this sum. By the uniform boundedness of $\boldsymbol{A}$, the absolute value of the second term in this sum is bounded by a constant times

$$\left(\|\nabla R_0\|_{L^2(\square_1)} + \sum_{i=1}^d \|R_i\|_{L^2(\square_1)} \right) \left(\int_{\square_1} \chi^4 \begin{pmatrix} \nabla u \\ \boldsymbol{g} \end{pmatrix} \cdot \boldsymbol{A} \begin{pmatrix} \nabla u \\ \boldsymbol{g} \end{pmatrix} \right)^{\frac{1}{2}}.$$

When applying the chain rule in the identity (3-33), the leftover terms are

$$4 \int_{\square_1} \chi^2 \begin{pmatrix} \sum_{j=1}^d T_{0j} (3 \nabla \chi \, \partial_j \chi + \chi \, \nabla \partial_j \chi) + 2 \nabla T_{0j} \chi \, \partial_j \chi \\ \sum_{j=0}^d T_{ij} \chi \, \partial_j \chi \end{pmatrix} \cdot \boldsymbol{A} \begin{pmatrix} \nabla u \\ \boldsymbol{g} \end{pmatrix}.$$

Using once more the uniform boundedness of $\boldsymbol{A}$, we obtain that the absolute value of the quantity above is bounded by a constant times

$$\left(\sum_{j=1}^d (\|T_{0j}\|_{L^2(\square_1)} + \|\nabla T_{0j}\|_{L^2(\square_1)}) + \sum_{j=0}^d \|T_{ij}\|_{L^2(\square_1)} \right) \left(\int_{\square_1} \chi^4 \begin{pmatrix} \nabla u \\ \boldsymbol{g} \end{pmatrix} \cdot \boldsymbol{A} \begin{pmatrix} \nabla u \\ \boldsymbol{g} \end{pmatrix} \right)^{\frac{1}{2}}.$$

Combining the previous displays with the estimates (3-28), (3-30), (3-31) and (3-32), we arrive at

$$\left(\int_{\square_1} \chi^4 \begin{pmatrix} \nabla u \\ \boldsymbol{g} \end{pmatrix} \cdot \boldsymbol{A} \begin{pmatrix} \nabla u \\ \boldsymbol{g} \end{pmatrix} \right)^{\frac{1}{2}} \leq C (\|\nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(\square_1)} + \|\boldsymbol{g}\|_{\widehat{H}^{-1}_{\mathrm{par}}(\square_1)}).$$

By the uniform ellipticity of $\boldsymbol{A}$, the left side is an upper bound for $\|S\|_{L^2(\square_0)}$, up to a multiplicative constant, and therefore the proof is complete. □

4. Convergence of subadditive quantities

We obtain an algebraic rate of convergence for the limits of the subadditive quantities by adapting the approach of [Armstrong and Smart 2016; Armstrong and Mourrat 2016], following the presentation of [Armstrong et al. 2017b, Chapter 2].

We let $\bar{\boldsymbol{A}}$ be the $2d$-by-$2d$ matrix characterized by the limit

$$\lim_{n\to\infty} \mathbb{E}[\mu(\boxdot_n, X)] = \tfrac{1}{2} X \cdot \bar{\boldsymbol{A}} X. \tag{4-1}$$

Note that the existence of the limit on the left side follows from the subadditivity of $\mu(\cdot, X) = J(\cdot, X, 0)$ and stationarity, which together ensure that $\mathbb{E}[\mu(\boxdot_n, X)]$ is a nonincreasing sequence. The fact that $X \mapsto \mu(\boxdot_n, X)$ is quadratic ensures that the limit is also quadratic in X and can therefore be represented by a matrix. Moreover, by Lemma 2.4, there exists $C(\Lambda) < \infty$ such that

$$\frac{1}{C} I_{2d} \le \bar{\boldsymbol{A}} \le C I_{2d}, \tag{4-2}$$

where I_{2d} is the $2d$-by-$2d$ identity matrix. It is convenient to define

$$\bar{J}(X, X^*) := \tfrac{1}{2} X \cdot \bar{\boldsymbol{A}} X + \tfrac{1}{2} X^* \cdot \bar{\boldsymbol{A}}^{-1} X^* - X \cdot X^*.$$

The goal of this section is to prove the following theorem.

Theorem 4.1 (convergence of J). *There exist an exponent $\beta(d, \Lambda) > 0$ and, for each $s \in (0, 2+d)$, a constant $C(s, d, \Lambda) < \infty$ such that, for every $X, X^* \in B_1$ and $n \in \mathbb{N}$, we have*

$$|J(\boxdot_n, X, X^*) - \bar{J}(X, X^*)| \le C 3^{-n\beta(2+d-s)} + \mathcal{O}_1(C 3^{-ns}). \tag{4-3}$$

The next lemma, which should be compared to [Armstrong et al. 2017b, Lemma 2.7], allows us to reduce Theorem 4.1 to an estimate on the quantity $J(\boxdot_n, X, \bar{\boldsymbol{A}} X)$. Note that, in view of Lemma 2.3, a control on the size of $\inf_{X^*} J(V, X, X^*)$ can be interpreted as information on the "convex duality defect" between the quantities μ and μ^*, quantifying how close these functions are to a convex dual pair.

Lemma 4.2 (reduction to minimal set). *For each $\Gamma \ge 1$, there exists a constant $C(\Gamma, d, \Lambda) < \infty$ such that, for every $2d$-by-$2d$ symmetric matrix $\tilde{\boldsymbol{A}}$ satisfying*

$$\Gamma^{-1} I_{2d} \le \tilde{\boldsymbol{A}} \le \Gamma I_{2d} \tag{4-4}$$

and every parabolic cylinder $V \subseteq \mathbb{R}^{d+1}$, we have

$$\sup_{X, X^* \in B_1} \left| J(V, X, X^*) - \left(\tfrac{1}{2} X \cdot \tilde{\boldsymbol{A}} X + \tfrac{1}{2} X^* \cdot \tilde{\boldsymbol{A}}^{-1} X^* - X \cdot X^* \right) \right| \le C \sup_{X \in B_1} (J(V, X, \tilde{\boldsymbol{A}} X))^{\frac{1}{2}}. \tag{4-5}$$

Proof. Since the domain V plays no role in the argument, we drop the explicit dependence on V. Define

$$\delta^2 := \sup_{X \in B_1} J(X, \tilde{\boldsymbol{A}} X).$$

To avoid a conflict in the notation, we denote the Legendre–Fenchel transform (convex dual function) of μ by

$$H(X^*) := \sup_{X \in \mathbb{R}^{2d}} (X \cdot X^* - \mu(X)).$$

It is clear from (2-14) that

$$H(X^*) \le \mu^*(X^*).$$

Thus, by (2-18), for every $X \in B_1$,

$$0 \le \mu(X) + H(\tilde{\boldsymbol{A}} X) - X \cdot \tilde{\boldsymbol{A}} X \le \mu(X) + \mu^*(\tilde{\boldsymbol{A}} X) - X \cdot \tilde{\boldsymbol{A}} X = J(X, \tilde{\boldsymbol{A}} X) \le \delta^2. \tag{4-6}$$

This implies that, for every $X \in B_1$,

$$|\mu^*(\tilde{\boldsymbol{A}}X) - H(\tilde{\boldsymbol{A}}X)| \le \delta^2. \tag{4-7}$$

For each $X \in \mathbb{R}^{2d}$, the minimum of the map $X^* \mapsto \mu(X) + H(X^*) - X \cdot X^*$ is zero and it is achieved at X^* for which $X^* = \nabla\mu(X)$. By uniform convexity (quadratic response) and (4-6), we deduce that, for every $X \in B_1$,

$$|\tilde{\boldsymbol{A}}X - \nabla\mu(X)|^2 \le C\delta^2. \tag{4-8}$$

Using the expression

$$\mu(X) = \tfrac{1}{2}X \cdot \nabla\mu(X)$$

we obtain, for every $X \in B_1$,

$$\left|\tfrac{1}{2}X \cdot \tilde{\boldsymbol{A}}X - \mu(X)\right| \le C\delta.$$

From this, uniform convexity and (4-4), we obtain, for every $X^* \in B_1$,

$$\left|\tfrac{1}{2}X^* \cdot \tilde{\boldsymbol{A}}^{-1}X^* - H(X^*)\right| \le C\delta.$$

Hence by (4-7), (4-4) again,

$$\left|\tfrac{1}{2}X^* \cdot \tilde{\boldsymbol{A}}^{-1}X^* - \mu^*(X^*)\right| \le C\delta.$$

The formula (2-18) now yields the lemma. □

We decompose the estimate for $J(\square_n, X, \bar{\boldsymbol{A}}X)$ into three steps. In the first step, we identify a convenient finite-volume approximation of the homogenized matrix $\bar{\boldsymbol{A}}$. We next control the expectation of $J(\square_n, X, \bar{\boldsymbol{A}}X)$ in Section 4B. We finally use the subadditivity of J in Section 4C to deduce a control of the fluctuations of $J(\square_n, X, \bar{\boldsymbol{A}}X)$, and complete the proof of Theorem 4.1.

4A. ***The coarsened mapping.*** Recall that $S(\cdot, V, X, X^*)$ denotes the unique maximizer in the definition of $J(V, X, X^*)$; see (2-16). We let $\bar{\boldsymbol{A}}_V \in \mathbb{R}^{2d\times 2d}$ be the symmetric matrix such that, for every $X^* \in \mathbb{R}^{2d}$,

$$\mathbb{E}[J(V, 0, X^*)] = \tfrac{1}{2}X^* \cdot \bar{\boldsymbol{A}}_V^{-1}X^*.$$

By (2-24), there exists $C(d, \Lambda) < \infty$ such that

$$\frac{1}{C}\, I_{2d} \le \bar{\boldsymbol{A}}_V \le C\, I_{2d},$$

and by (2-29),

$$\mathbb{E}\left[\fint_V S(\cdot, V, 0, X^*)\right] = \bar{\boldsymbol{A}}_V^{-1}X^*.$$

Recalling also (2-30) and the linearity of the mapping $(X, X^*) \mapsto S(\cdot, V, X, X^*)$, we thus see that the matrix $\bar{\boldsymbol{A}}_V$ is such that, for every $X \in \mathbb{R}^{2d}$,

$$\mathbb{E}\left[\fint_V S(\cdot, V, X, \bar{\boldsymbol{A}}_V X)\right] = 0. \tag{4-9}$$

We note that by Lemmas 2.3 and 2.4, for each $X \in \mathbb{R}^{2d}$, the mapping

$$X^* \mapsto \mathbb{E}[J(V, X, X^*)] = \mathbb{E}[\mu(V, X)] + \mathbb{E}[\mu^*(V, X^*)] - X \cdot X^*$$

is uniformly convex, and achieves its unique minimum at X^* satisfying

$$\mathbb{E}[\nabla_{X^*}\mu^*(V, X^*)] = X.$$

Moreover, the latter condition is equivalent to $X^* = \bar{\boldsymbol{A}}_V X$. We thus deduce that for every $X, X^* \in \mathbb{R}^{2d}$,

$$\mathbb{E}[J(V, X, \bar{\boldsymbol{A}}_V X)] \le \mathbb{E}[J(V, X, X^*)] \le \mathbb{E}[J(V, X, \bar{\boldsymbol{A}}_V X)] + C|X^* - \bar{\boldsymbol{A}}_V X|^2. \tag{4-10}$$

We use the shorthand notation

$$\bar{\boldsymbol{A}}_n := \bar{\boldsymbol{A}}_{\square_n}. \tag{4-11}$$

4B. ***Control of the expectation of J.*** The goal of this subsection is to prove the following proposition.

Proposition 4.3 (decay of $\mathbb{E}[J]$). *There exist $\beta(d, \Lambda) > 0$ and $C(d, \Lambda) < \infty$ such that, for every $n \in \mathbb{N}$ and $X \in B_1(\mathbb{R}^{2d})$,*

$$\mathbb{E}[J(\square_n, X, \bar{\boldsymbol{A}}X)] \le C3^{-\beta n}. \tag{4-12}$$

The main step to prove this result is to control the size of J near $(X, \bar{\boldsymbol{A}}X)$ in terms of the expected "additivity defect" of J between successive triadic scales. We measure the latter using the quantity

$$\tau_n := \sup_{X, X^* \in B_1} \big(\mathbb{E}[J(\square_n, X, X^*)] - \mathbb{E}[J(\square_{n+1}, X, X^*)]\big). \tag{4-13}$$

Proposition 4.4. *There exist $\alpha(d) < \infty$ and $C(d, \Lambda) < \infty$ such that, for every $n \in \mathbb{N}$ and $X \in B_1(\mathbb{R}^{2d})$,*

$$\mathbb{E}[J(\square_n, X, \bar{\boldsymbol{A}}_n X)] \le C\, 3^{-\alpha n}\left(1 + \sum_{k=0}^{n-1} 3^{\alpha k}\tau_k\right).$$

As will be explained below, Proposition 4.3 follows from Proposition 4.4 by iteration, in analogy with an ODE argument. We focus for now on the proof of Proposition 4.4, and start by rewriting the quadratic response (2-27) in a more convenient form.

Lemma 4.5 (quadratic response). *There exists a constant $C(d, \Lambda) < \infty$ such that the following holds. Let $V, V_1, \ldots, V_k$ be parabolic cylinders such that $\{V_1, \ldots, V_k\}$ forms a partition of V, up to a set of null measure. For every $X, X^* \in \mathbb{R}^{2d}$, we have*

$$\sum_{j=1}^{k} \frac{|V_j|}{|V|} \|S(\cdot, V, X, X^*) - S(\cdot, V_j, X, X^*)\|^2_{\underline{L}^2(V_j)} \le C \sum_{j=1}^{k} \frac{|V_j|}{|V|}(J(V_j, X, X^*) - J(V, X, X^*)).$$

Proof. Define $T := S(\cdot, V, X, X^*)$. Applying (2-27) on the subdomain V_j for each $j \in \{1, \ldots, k\}$, we get

$$\frac{1}{C}\int_{V_j} |T - S(\cdot, V_j, X, X^*)|^2 \le |V_j|\, J(V_j, X, X^*) - \int_{V_j} \big(-\tfrac{1}{2}T \cdot \boldsymbol{A}T - X \cdot \boldsymbol{A}T + X^* \cdot T\big).$$

Summing over j and recalling (2-16) yields the result. □

We next show that the spatial averages of S can be controlled by an expression involving the additivity defects of J on all smaller length scales. We define

$$\bar{S}_n(X, X^*) := \mathbb{E}\left[\fint_{\square_n} S(\cdot, \square_n, X, X^*)\right].$$

Lemma 4.6. *There exist $\alpha(d) < \infty$ and $C(d, \Lambda) < \infty$ such that, for every $n \in \mathbb{N}$ and $X, X^* \in B_1(\mathbb{R}^{2d})$, we have*

$$\mathbb{E}\left[\left|\fint_{\boxdot_n} S(\cdot, \boxdot_n, X, X^*) - \bar{S}_n(X, X^*)\right|^2\right] \le C3^{-\alpha n}\left(1 + \sum_{k=0}^{n-1} 3^{\alpha k}\tau_k\right). \tag{4-14}$$

Proof. For any $X' \in B_1$, $m \le n$ and $z \in \mathcal{Z}_m$, the first variation (2-26) gives

$$X' \cdot \fint_{z+\boxdot_m} (S(\cdot, \boxdot_n, X, X^*) - S(\cdot, z+\boxdot_m, X, X^*)) = \fint_{z+\boxdot_m} S(\cdot, z+\boxdot_m, 0, X') \cdot \boldsymbol{A}(S(\cdot, \boxdot_n, X, X^*) - S(\cdot, z+\boxdot_m, X, X^*)).$$

Averaging over $z \in \mathcal{Z}_m$ and using the Cauchy–Schwarz inequality yields

$$\begin{aligned} &|\mathcal{Z}_m|^{-1} \cdot \left|X' \cdot \sum_{z\in\mathcal{Z}_m} \fint_{z+\boxdot_m} (S(\cdot, \boxdot_n, X, X^*) - S(\cdot, z+\boxdot_m, X, X^*))\right| \\ &\quad \le \left(|\mathcal{Z}_m|^{-1} \sum_{z\in\mathcal{Z}_m} \fint_{z+\boxdot_m} |\boldsymbol{A}S(\cdot, z+\boxdot_m, 0, X')|^2\right)^{\frac{1}{2}} \\ &\qquad \cdot \left(|\mathcal{Z}_m|^{-1} \sum_{z\in\mathcal{Z}_m} \fint_{z+\boxdot_m} |S(\cdot, \boxdot_n, X, X^*) - S(\cdot, z+\boxdot_m, X, X^*)|^2\right)^{\frac{1}{2}}. \end{aligned} \tag{4-15}$$

The first term on the right side is bounded by a constant $C(d, \Lambda) < \infty$. We use Lemma 4.5 to bound the second term and obtain

$$\left||\mathcal{Z}_m|^{-1} \sum_{z\in\mathcal{Z}_m} \fint_{z+\boxdot_m} |S(\cdot, \boxdot_n, X, X^*) - S(\cdot, z+\boxdot_m, X, X^*)|^2\right| \le \frac{C}{|\mathcal{Z}_m|} \sum_{z\in\mathcal{Z}_m} (J(z+\boxdot_m, X, X^*) - J(\boxdot_n, X, X^*)).$$

Now, we can estimate the variance of $\fint_{\boxdot_n} S(\cdot, \boxdot_n, X, X^*)$ using those at scale m:

$$\operatorname{var}\left[\fint_{\boxdot_n} S(\cdot, \boxdot_n, X, X^*)\right] \le 2\operatorname{var}\left[3^{-(d+2)(n-m)} \sum_{z\in\mathcal{Z}_m} \fint_{z+\boxdot_m} S(\cdot, z+\boxdot_m, X, X^*)\right] + C\mathbb{E}[J(\boxdot_m, X, X^*) - J(\boxdot_n, X, X^*)].$$

For $m, n \in \mathbb{N}$, $m \le n$, we can decompose $\mathcal{Z}_m$ into a union of 2^{d+1} "checkerboard" subsets $\mathcal{Z}^{(1)}, \dots, \mathcal{Z}^{(2^{1+d})}$ to ensure that for each $i \in \{1, \dots, 2^{1+d}\}$,

$$(z, z') \in \mathcal{Z}^{(i)} \quad \Longrightarrow \quad \operatorname{dist}(z+\boxdot_m, z'+\boxdot_m) \ge 1.$$

For example, to any $i \in \{1, \dots, 2^{d+1}\}$ we can associate $(i_0, \dots, i_d) \in \{0, 1\}^{d+1}$, and then set

$$Z^{(i)} := \left((i_0 3^{2m}, i_1 3^m, \dots, i_d 3^m) + 2((3^{2m}\mathbb{Z}) \times (3^m\mathbb{Z}^d))\right) \cap \boxdot_n. \tag{4-16}$$

Thus, we obtain the bound

$$\operatorname{var}\left[\sum_{z\in\mathcal{Z}_m} \fint_{z+\boxdot_m} S(\cdot, z+\boxdot_m, X, X^*)\right] \le C(d) \sum_{i=1}^{2^{d+1}} \operatorname{var}\left[\sum_{z\in\mathcal{Z}^{(i)}} \fint_{z+\boxdot_m} S(\cdot, z+\boxdot_m, X, X^*)\right],$$

and by independence at distance larger than 1 and stationarity,

$$\operatorname{var}\Big[\sum_{z\in\mathcal{Z}_m} \fint_{z+\square_m} S(\cdot, z+\square_m, X, X^*)\Big] \le C3^{(d+2)(n-m)} \operatorname{var}\Big[\fint_{\square_m} S(\cdot, \square_m, X, X^*)\Big].$$

We can now estimate the variance of the spatial average of S at scale n by the variance at smaller scales:

$$\operatorname{var}\Big[\fint_{\square_n} S(\cdot, \square_n, X, X^*)\Big] \le C3^{-(d+2)(n-m)} \operatorname{var}\Big[\fint_{\square_m} S(\cdot, \square_m, X, X^*)\Big] + C\mathbb{E}[J(\square_m, X, X^*) - J(\square_n, X, X^*)]. \quad (4\text{-}17)$$

Selecting ℓ to be the smallest integer such that $C3^{-(d+2)\ell} \le \frac{1}{3}$, we get

$$\operatorname{var}\Big[\fint_{\square_{m+\ell}} S(\cdot, \square_{m+\ell}, X, X^*)\Big] \le \frac{1}{3} \operatorname{var}\Big[\fint_{\square_m} S(\cdot, \square_m, X, X^*)\Big] + C\mathbb{E}[J(\square_m, X, X^*) - J(\square_{m+\ell}, X, X^*)].$$

We introduce

$$u_n := \operatorname{var}\Big[\fint_{\square_n} S(\cdot, \square_n, X, X^*)\Big],$$

and $v_n := u_{n\ell}$. We have

$$v_n \le \tfrac{1}{3} v_{n-1} + C \sum_{k=(n-1)\ell}^{n\ell-1} \tau_k,$$

and, by induction,

$$v_n \le 3^{-n} u_0 + C \sum_{i=1}^{n} 3^{-i} \sum_{k=(n-i)\ell}^{(n-i+1)\ell-1} \tau_k.$$

Defining $\alpha := 1/\ell$ (recall that ℓ only depends on d), we get

$$v_n \le C\Big(3^{-n} + \sum_{k=0}^{n\ell-1} 3^{-\alpha(n\ell-k)} \tau_k\Big).$$

Thus, if n is a multiple of ℓ, we have

$$u_n \le C\Big(3^{-\alpha n} + \sum_{k=0}^{n-1} 3^{-\alpha(n-k)} \tau_k\Big),$$

and for $n = \ell n' + m$, with $0 \le m < \ell$, another application of (4-17) gives the same estimate, so finally

$$\operatorname{var}\Big[\fint_{\square_n} S(\cdot, \square_n, X, X^*)\Big] \le C\Big(3^{-\alpha n} + \sum_{k=0}^{n-1} 3^{-\alpha(n-k)} \tau_k\Big),$$

which is (4-14). □

We can now sum the scales and deduce that $S(\cdot, \square_n, X, X^*)$ is close to a constant in a weak sense, provided that a weighted norm of $(\tau_k)_{k<n}$ is small.

Lemma 4.7 (weak control of S). *There exist $\alpha(d) < \infty$ and $C(d, \Lambda) < \infty$ such that, for every $n \in \mathbb{N}$ and $X, X^* \in B_1(\mathbb{R}^{2d})$,*

$$\mathbb{E}\Big[\|S(\cdot, \boxdot_n, X, X^*) - \bar{S}_n(X, X^*)\|^2_{\underline{\widehat{H}}^{-1}_{\mathrm{par}}(\boxdot_n)}\Big] \le C 3^{(2-\alpha)n}\bigg(1 + \sum_{k=0}^{n-1} 3^{\alpha k}\tau_k\bigg).$$

Proof. We decompose the proof into three steps.

Step 1. To begin, we show that there exists a constant $C(d, \Lambda) < \infty$ such that, for every $m, n \in \mathbb{N}$, $m \le n$, and $X, X^* \in B^1(\mathbb{R}^{2d})$,

$$|\bar{S}_n(X, X^*) - \bar{S}_m(X, X^*)|^2 \le C \sum_{k=m}^{n-1} \tau_k. \tag{4-18}$$

Indeed, recalling the definition of $\mathcal{Z}_m$ in (3-8) (which depends implicitly on n), we use Jensen's inequality, Lemma 4.5 and stationarity to get

$$\begin{aligned}
&\bigg|\mathbb{E}\bigg[\fint_{\boxdot_n} S(\cdot, \boxdot_n, X, X^*) - |\mathcal{Z}_m|^{-1} \sum_{z\in\mathcal{Z}_m} \fint_{z+\boxdot_m} S(\cdot, z+\boxdot_m, X, X^*)\bigg]\bigg|^2 \\
&\qquad \le \mathbb{E}\bigg[|\mathcal{Z}_m|^{-1} \sum_{z\in\mathcal{Z}_m} \fint_{z+\boxdot_m} \Big|S(\cdot, \boxdot_n, X, X^*) - S(\cdot, z+\boxdot_m, X, X^*)\Big|^2\bigg] \\
&\qquad \le C\big(J(\boxdot_n, X, X^*) - J(\boxdot_m, X, X^*)\big) \le C \sum_{k=m}^{n-1} \tau_k,
\end{aligned}$$

and this implies (4-18).

Step 2. In this step, we show that there exists a constant $C(d, \Lambda) < \infty$ such that, for every $m \in \mathbb{N}$, $m \le n$,

$$\mathbb{E}\bigg[|\mathcal{Z}_m|^{-1} \sum_{z\in\mathcal{Z}_m} \Big|\big(S(\cdot, \boxdot_n, X, X^*)\big)_{z+\boxdot_m} - \bar{S}_n(X, X^*)\Big|^2\bigg] \le C\bigg(3^{-\alpha m} + \sum_{k=0}^{m} 3^{\alpha(k-m)}\tau_k + \sum_{k=m}^{n-1}\tau_k\bigg). \tag{4-19}$$

By Lemma 4.5, we have

$$\begin{aligned}
&\sum_{z\in\mathcal{Z}_m} \|S(\cdot, \boxdot_n, X, X^*) - S(\cdot, z+\boxdot_m, X, X^*)\|^2_{\underline{L}^2(z+\boxdot_m)} \\
&\qquad\qquad \le C \sum_{z\in\mathcal{Z}_m} \big(J(\boxdot_n, X, X^*) - J(z+\boxdot_m, X, X^*)\big).
\end{aligned} \tag{4-20}$$

Taking expectations, using stationarity and Jensen's inequality, we deduce that

$$\mathbb{E}\bigg[\sum_{z\in\mathcal{Z}_m} \Big|\big(S(\cdot, \boxdot_n, X, X^*)\big)_{z+\boxdot_m} - \big(S(\cdot, z+\boxdot_m, X, X^*)\big)_{z+\boxdot_m}\Big|^2\bigg] \le C|\mathcal{Z}_m| \sum_{k=m}^{n-1}\tau_k. \tag{4-21}$$

Moreover, by stationarity and Lemma 4.6, we have

$$\mathbb{E}\bigg[|\mathcal{Z}_m|^{-1} \sum_{z\in\mathcal{Z}_m} \Big|\big(S(\cdot, z+\boxdot_m, X, X^*)\big)_{z+\boxdot_m} - \bar{S}_m(X, X^*)\Big|^2\bigg] \le C\bigg(3^{-\alpha m} + \sum_{k=0}^{m} 3^{\alpha(k-m)}\tau_k\bigg). \tag{4-22}$$

Since

$$\begin{aligned}\big|(S(\cdot,\boxdot_n,X,X^*))_{z+\boxdot_m}-\bar S_n(X,X^*)\big|^2 &\le 3\big|(S(\cdot,\boxdot_n,X,X^*))_{z+\boxdot_m}-(S(\cdot,z+\boxdot_m,X,X^*))_{z+\boxdot_m}\big|^2\\ &\quad+3\big|(S(\cdot,z+\boxdot_m,X,X^*))_{z+\boxdot_m}-\bar S_m(X,X^*)\big|^2\\ &\quad+3|\bar S_n(X,X^*)-\bar S_m(X,X^*)|^2,\end{aligned}$$

we obtain (4-19) by combining (4-21), (4-22) and (4-18).

Step 3. We now combine Proposition 3.6 with the result of the previous step to obtain

$$\|S(\cdot,\boxdot_n,X,X^*)-\bar S_n(X,X^*)\|^2_{\underline{H}^{-1}_{\mathrm{par}}(\boxdot_n)}\le C\bigg(1+\bigg(\sum_{m=0}^{n-1}3^m Z_m^{\frac12}\bigg)^2\bigg),\tag{4-23}$$

where Z_m is a random variable satisfying

$$\mathbb{E}[Z_m]\le C\bigg(3^{-\alpha m}+\sum_{k=0}^{m}3^{\alpha(k-m)}\tau_k+\sum_{k=m}^{n-1}\tau_k\bigg).\tag{4-24}$$

By Hölder's inequality, we have

$$\bigg(\sum_{m=0}^{n-1}3^m Z_m^{\frac12}\bigg)^2\le\bigg(\sum_{m=0}^{n-1}3^m\bigg)\bigg(\sum_{m=0}^{n-1}3^m Z_m\bigg)\le C3^n\sum_{m=0}^{n-1}3^m Z_m.\tag{4-25}$$

Taking expectations and using (4-24), we get

$$\mathbb{E}\bigg[\bigg(\sum_{m=0}^{n-1}3^m Z_m^{\frac12}\bigg)^2\bigg]\le C3^n\sum_{m=0}^{n-1}3^m\bigg(3^{-\alpha m}+\sum_{k=0}^{m}3^{\alpha(k-m)}\tau_k+\sum_{k=m}^{n-1}\tau_k\bigg).$$

For the last two terms, we reverse the order of the sums to find

$$\sum_{m=0}^{n-1}3^m\sum_{k=0}^{m}3^{\alpha(k-m)}\tau_k=\sum_{k=0}^{n-1}3^{\alpha k}\tau_k\sum_{m=k}^{n-1}3^{(1-\alpha m)}\le C3^{(1-\alpha)n}\sum_{k=0}^{n-1}3^{\alpha k}\tau_k,$$

and

$$\sum_{m=0}^{n-1}3^m\sum_{k=m}^{n-1}\tau_k=\sum_{k=0}^{n-1}\sum_{m=0}^{k}3^m\tau_k\le C\sum_{k=0}^{n-1}3^k\tau_k.\tag{4-26}$$

The second sum is bounded by the first; thus combining the above displays yields

$$\mathbb{E}\bigg[\bigg(\sum_{m=0}^{n-1}3^m Z_m^{\frac12}\bigg)^2\bigg]\le C3^{(2-\alpha)n}\bigg(1+\sum_{k=0}^{n-1}3^{\alpha k}\tau_k\bigg).$$

□

We next complete the proof of Proposition 4.4 and then of Proposition 4.3.

Proof of Proposition 4.4. According to Lemma 4.7, Proposition 3.10 and (4-9), we have

$$\mathbb{E}\big[\|S(\cdot,\boxdot_n,X,\bar{\boldsymbol{A}}_nX)\|^2_{\underline{L}^2(\boxdot_{n-1})}\big]\le C3^{-\alpha n}\bigg(1+\sum_{k=0}^{n-1}3^{\alpha k}\tau_k\bigg).$$

By Lemma 4.5, we deduce

$$\mathbb{E}\Big[\|S(\cdot,\boxdot_{n-1},X,\bar{\boldsymbol{A}}_nX)\|^2_{\underline{L}^2(\boxdot_{n-1})}\Big]\le C3^{-\alpha n}\bigg(1+\sum_{k=0}^{n-1}3^{\alpha k}\tau_k\bigg).$$

Recall that $(z+\boxdot_{n-1})_{z\in\mathcal{Z}_{n-1}}$ is a partition of $\boxdot_n$, up to a set of null measure. Moreover, by stationarity, the previous display implies that, for every $z\in\mathcal{Z}_{n-1}$,

$$\mathbb{E}\Big[\|S(\cdot,z+\boxdot_{n-1},X,\bar{\boldsymbol{A}}_nX)\|^2_{\underline{L}^2(z+\boxdot_{n-1})}\Big]\le C3^{-\alpha n}\bigg(1+\sum_{k=0}^{n-1}3^{\alpha k}\tau_k\bigg).$$

Applying Lemma 4.5 once more and summing over $z\in\mathcal{Z}_{n-1}$, we obtain the result. □

Proof of Proposition 4.3. We denote by $\mathcal{B}$ the set of canonical basis elements of $\mathbb{R}^{2d}$, and observe that there exists a constant $C(d)<\infty$ such that if $X\mapsto B(X)$ is a nonnegative quadratic form over $\mathbb{R}^{2d}$, then

$$\sup_{X\in B_1}B(X)\le C\sum_{X\in\mathcal{B}}B(X). \tag{4-27}$$

Indeed, a quadratic form is associated to a nonnegative symmetric matrix with largest eigenvalue bounded by its trace; this trace is equal to the right side above.

By the definition of τ_n, see (4-13), and Lemma 2.3, we have

$$\tau_n\le\sup_{X\in B_1}\Big(\mathbb{E}[\mu(\boxdot_n,X)]-\mathbb{E}[\mu(\boxdot_{n+1},X)]\Big)+\sup_{X^*\in B_1}\Big(\mathbb{E}[\mu^*(\boxdot_n,X^*)]-\mathbb{E}[\mu^*(\boxdot_{n+1},X^*)]\Big).$$

Since $X\mapsto\mathbb{E}[\mu(V,X)]$ and $X^*\mapsto\mathbb{E}[\mu(V,X^*)]$ are nonnegative quadratic forms, and since this property is stable under linear changes of coordinates, it follows from (4-27) that

$$\tau_n\le C\sum_{X\in\mathcal{B}}\Big(\mathbb{E}[\mu(\boxdot_n,X)]-\mathbb{E}[\mu(\boxdot_{n+1},X)]\Big)+C\sum_{X\in\mathcal{B}}\Big(\mathbb{E}[\mu^*(\boxdot_n,\bar{\boldsymbol{A}}_nX)]-\mathbb{E}[\mu^*(\boxdot_{n+1},\bar{\boldsymbol{A}}_nX)]\Big),$$

and thus by Lemma 2.3,

$$\tau_n\le C\sum_{X\in\mathcal{B}}\Big(\mathbb{E}[J(\boxdot_n,X,\bar{\boldsymbol{A}}_nX)]-\mathbb{E}[J(\boxdot_{n+1},X,\bar{\boldsymbol{A}}_nX)]\Big).$$

By (4-10), we have

$$\mathbb{E}[J(\boxdot_{n+1},X,\bar{\boldsymbol{A}}_{n+1}X)]\le\mathbb{E}[J(\boxdot_{n+1},X,\bar{\boldsymbol{A}}_nX)],$$

and therefore

$$\tau_n\le C\sum_{X\in\mathcal{B}}\Big(\mathbb{E}[J(\boxdot_n,X,\bar{\boldsymbol{A}}_nX)]-\mathbb{E}[J(\boxdot_{n+1},X,\bar{\boldsymbol{A}}_{n+1}X)]\Big). \tag{4-28}$$

This motivates the definition of

$$D_n:=\sum_{X\in\mathcal{B}}\mathbb{E}[J(\boxdot_n,X,\bar{\boldsymbol{A}}_nX)].$$

Proposition 4.4 asserts that

$$D_n\le C3^{-\alpha n}\bigg(1+\sum_{k=0}^{n-1}3^{\alpha k}\tau_k\bigg).$$

Setting

$$\widetilde{D}_n := 3^{-\frac{\alpha}{2}n} \sum_{k=0}^{n} 3^{\frac{\alpha}{2}k} D_k,$$

we deduce that

$$\begin{aligned} \widetilde{D}_n &\le C3^{-\frac{\alpha}{2}n} \sum_{m=0}^{n} 3^{-\frac{\alpha}{2}m} \Big(1 + \sum_{k=0}^{m} 3^{\alpha k} \tau_k\Big) \\ &\le C3^{-\frac{\alpha}{2}n} + C3^{-\frac{\alpha}{2}n} \sum_{k=0}^{n} \sum_{m=k}^{n} 3^{-\frac{\alpha}{2}m}\, 3^{\alpha k} \tau_k \le C3^{-\frac{\alpha}{2}n} \Big(1 + \sum_{k=0}^{n} 3^{\frac{\alpha}{2}k} \tau_k\Big). \end{aligned} \tag{4-29}$$

Since $D_0 \le C$, we also have

$$\widetilde{D}_n - \widetilde{D}_{n+1} \ge 3^{-\frac{\alpha}{2}n} \sum_{k=0}^{n} 3^{\frac{\alpha}{2}k} (D_k - D_{k+1}) - C3^{-\frac{\alpha}{2}n}.$$

Combining this with (4-28) yields

$$\widetilde{D}_n - \widetilde{D}_{n+1} \ge C^{-1}\, 3^{-\frac{\alpha}{2}n} \sum_{k=0}^{n} 3^{\frac{\alpha}{2}} \tau_k - C3^{-\frac{\alpha}{2}n}.$$

From this and (4-29), we obtain that there exists an exponent $\beta(d, \Lambda) \in \big(0, \frac{\alpha}{2}\big)$ such that

$$\widetilde{D}_{n+1} \le 3^{-\beta} \widetilde{D}_n + C3^{-\frac{\alpha}{2}n};$$

introducing $v_n := 3^{\beta n} \widetilde{D}_n$ and multiplying the previous identity by $3^{(\beta+1)n}$ gives

$$v_{n+1} \le v_n + C3^{(\beta - \frac{\alpha}{2})n}.$$

Summing this inequality over n yields

$$v_n \le v_0 + \frac{C}{1 - (3^{\beta - \frac{\alpha}{2}})}.$$

That is, v is bounded; i.e.,

$$\widetilde{D}_n \le C3^{-\beta n}.$$

By (4-28), we also obtain

$$\tau_n \le C3^{-\beta n}.$$

By the definition of $\bar{A}_n$, we have

$$|\bar{A}_n - \bar{A}_{n+1}| \le C\tau_n,$$

so that, setting

$$\bar{A} := \lim_{n \to \infty} \bar{A}_n, \tag{4-30}$$

we get

$$|\bar{A}_n - \bar{A}| \le \sum_{m=n}^{\infty} |\bar{A}_m - \bar{A}_{m+1}| \le C \sum_{m=n}^{\infty} \tau_m \le C3^{-\beta n}.$$

Combining the last displays with (4-10) yields

$$\sup_{X\in B_1} \mathbb{E}[J(\square_n, X, \bar{\boldsymbol{A}}X)] \le C3^{-\beta n}.$$

By an application of Lemma 4.2, we can verify that the matrix $\bar{\boldsymbol{A}}$ defined in (4-30) coincides with that defined in (4-1). □

4C. ***Control of the fluctuations of J.*** In this subsection, we prove Theorem 4.1. In view of Lemma 4.2, the main point is to obtain a control on the fluctuations of $J(\square_n, X, \bar{\boldsymbol{A}}X)$, which we do using subadditivity.

Proof of Theorem 4.1. Step 1. In this first step, we show that there exists an exponent $\beta(d, \Lambda) > 0$ and a constant $C(d, \Lambda) > 0$ such that, for every $X \in B_1(\mathbb{R}^{2d})$ and $m, n \in \mathbb{N},\ m \le n$, we have

$$3^{-(2+d)(n-m)} \log \mathbb{E}[\exp(C^{-1}3^{(2+d)(n-m)} J(\square_n, X, \bar{\boldsymbol{A}}X))] \le C3^{-\beta m}. \tag{4-31}$$

For $m, n \in \mathbb{N},\ m \le n$, recall that the cube $\square_n$ is partitioned into a union of 2^{d+1} "checkerboard" subsets, see (4-16), to ensure that for each $i \in \{1, \dots, 2^{1+d}\}$,

$$z, z' \in \mathcal{Z}^{(i)} \quad \Longrightarrow \quad \operatorname{dist}(z + \square_n, z' + \square_n) \ge 1.$$

In particular, for each fixed $i \in \{1, \dots, 2^{1+d}\}$, the random variables $(z + \square_n)_{z\in\mathcal{Z}_m^{(i)}}$ are independent. By subadditivity, for each $X \in B_1(\mathbb{R}^{2d})$ and $t > 0$, we have

$$\log \mathbb{E}[\exp(t3^{(2+d)(n-m)} J(\square_n, X, \bar{\boldsymbol{A}}X))] \le \log \mathbb{E}\left[\exp\left(t \sum_{z\in\mathcal{Z}_m} J(z + \square_m, X, \bar{\boldsymbol{A}}X)\right)\right],$$

and by Hölder's inequality and independence, the latter is bounded by

$$\begin{aligned} &\le 2^{-(1+d)} \sum_{i=1}^{2^{1+d}} \log \mathbb{E}\left[\exp\left(t2^{1+d} \sum_{z\in\mathcal{Z}_m^{(i)}} J(z + \square_m, X, \bar{\boldsymbol{A}}X)\right)\right] \\ &\le 2^{-(1+d)} \sum_{z\in\mathcal{Z}_m} \log \mathbb{E}[\exp(t2^{1+d} J(z + \square_m, X, \bar{\boldsymbol{A}}X))]. \end{aligned}$$

By stationarity, the summands above do not depend on $z \in \mathcal{Z}_m$. Since

$$J(\square_m, X, \bar{\boldsymbol{A}}X) \le C(d, \Lambda),$$

we can choose $t(d, \Lambda) > 0$ sufficiently small and use the elementary inequalities

$$\begin{aligned} &\exp(s) \le 1 + 2s \quad \text{for all } 0 \le s \le 1, \\ &\log(1 + s) \le s \quad \text{for all } s \ge 0 \end{aligned}$$

to obtain

$$\log \mathbb{E}[\exp(C^{-1}3^{(2+d)(n-m)} J(\square_n, X, \bar{\boldsymbol{A}}X))] \le C3^{(2+d)(n-m)} \mathbb{E}[J(\square_m, X, \bar{\boldsymbol{A}}X)].$$

Inequality (4-31) then follows by an application of Proposition 4.3.

Step 2. Set

$$\rho_n := \sup_{X \in B_1} J(\square_n, X, \bar{\boldsymbol{A}} X).$$

In this step, we show that there exists an exponent $\beta(d, \Lambda) > 0$ and, for every $s \in (0, 2+d)$, a constant $C(s, d, \Lambda) < \infty$ such that, for every $n \in \mathbb{N}$,

$$\rho_n \le C 3^{-\beta(2+d-s)n} + \mathcal{O}_1(C 3^{-sn}). \tag{4-32}$$

By (4-27) and Hölder's inequality, the relation (4-31) can be improved to

$$3^{-(2+d)(n-m)} \log \mathbb{E}[\exp(C^{-1} 3^{(2+d)(n-m)} \rho_n)] \le C 3^{-\beta m}. \tag{4-33}$$

By Chebyshev's inequality, for every $t \ge 0$,

$$\begin{aligned} \mathbb{P}[\rho_n \ge t] &\le \exp(-C^{-1} 3^{(2+d)(n-m)} t)\, \mathbb{E}[\exp(C^{-1} 3^{(2+d)(n-m)} \rho_n)] \\ &\le \exp(-C^{-1} 3^{(2+d)(n-m)} t + C 3^{(2+d)(n-m)-\beta m}). \end{aligned}$$

Replacing t by $C 3^{-\beta m} + t$ gives

$$\mathbb{P}[\rho_n \ge C 3^{-\beta m} + t] \le \exp(-C^{-1} 3^{(2+d)(n-m)} t).$$

Choosing

$$m := \left\lfloor \frac{2+d-s}{2+d} n \right\rfloor$$

yields

$$\mathbb{P}[\rho_n \ge C 3^{-\beta \frac{2+d-s}{2+d} n} + t] \le \exp(-C^{-1} 3^{sn} t).$$

By (1-19), this is (4-32), up to a redefinition of $\beta(d, \Lambda) > 0$.

Step 3. We now combine Lemma 4.2, (4-33) and the elementary inequality

$$\text{for all } a, b > 0, \quad (a+b)^{\frac{1}{2}} \le a^{\frac{1}{2}} + \tfrac{1}{2} a^{-\frac{1}{2}} b \tag{4-34}$$

to get

$$\begin{aligned} \sup_{X, X^* \in B_1} \left| J(\square_n, X, X^*) - \left(\tfrac{1}{2} X \cdot \bar{\boldsymbol{A}} X + \tfrac{1}{2} X^* \cdot \bar{\boldsymbol{A}}^{-1} X^* - X \cdot X^* \right) \right| \\ \le 3^{-\frac{\beta}{2}(2+d-s)n} + \mathcal{O}_1(C 3^{-\left(s - \frac{\beta}{2}(2+d-s)\right)n}). \end{aligned} \tag{4-35}$$

For every $s' \in (0, 2+d)$, if we set

$$s := \frac{2s' + \beta(2+d)}{2+\beta} \in (0, 2+d),$$

then the right side of (4-35) can be rewritten as

$$3^{-\frac{\beta}{2+\beta}(2+d-s')n} + \mathcal{O}_1(C 3^{-s'n}).$$

We have thus obtained (4-3), up to a redefinition of $\beta(d, \Lambda) > 0$. □

Proposition 4.8. *There exist $C(d,\Lambda)<\infty$ and a matrix $\bar{\boldsymbol{a}}\in\mathbb{R}^{d\times d}$ satisfying*

$$\text{for all } \xi\in\mathbb{R}^d,\quad \xi\cdot\bar{\boldsymbol{a}}\xi\ge\frac{1}{C}|\xi|^2 \quad \text{and} \quad |\bar{\boldsymbol{a}}\xi|\le C|\xi| \tag{4-36}$$

such that, for every $p,q\in\mathbb{R}^d$, we have the equivalence

$$\frac{1}{2}\begin{pmatrix}p\\q\end{pmatrix}\cdot\bar{\boldsymbol{A}}\begin{pmatrix}p\\q\end{pmatrix}-p\cdot q=0 \quad\Longleftrightarrow\quad q=\bar{\boldsymbol{a}}p. \tag{4-37}$$

Proof. Step 1. We show that, for every $p,q\in\mathbb{R}^d$,

$$\bar{A}(p,q):=\frac{1}{2}\begin{pmatrix}p\\q\end{pmatrix}\cdot\bar{\boldsymbol{A}}\begin{pmatrix}p\\q\end{pmatrix}\ge p\cdot q. \tag{4-38}$$

By Lemma 2.2, we have for every $S=(\nabla u,\boldsymbol{g})\in\mathcal{C}_0(I\times U)$ and $p,q\in\mathbb{R}^d$ that

$$\fint A(p+\nabla u,q+\boldsymbol{g},\cdot)\ge\fint(p+\nabla u)\cdot(q+\boldsymbol{g})=p\cdot q.$$

By the definition of μ in (2-5), we deduce that

$$\mu(I\times U,p,q)\ge p\cdot q,$$

and thus (4-38) follows from (4-1).

Step 2. We show that, for every $q^*,p^*\in\mathbb{R}^d$,

$$\frac{1}{2}\begin{pmatrix}q^*\\p^*\end{pmatrix}\cdot\bar{\boldsymbol{A}}^{-1}\begin{pmatrix}q^*\\p^*\end{pmatrix}\ge q^*\cdot p^*. \tag{4-39}$$

Fix $p^*\in\mathbb{R}^d$. For every $u\in\ell_{p^*}+H^1_{\mathrm{par},\sqcup}(I\times U)$ and $\boldsymbol{g}\in L^2(I\times U;\mathbb{R}^d)$ satisfying $-\nabla\cdot\boldsymbol{g}=-\partial_t u$, we have $(\nabla u,\boldsymbol{g})\in\mathcal{C}(I\times U)$, as well as

$$\fint_{I\times U}\nabla u=p^*$$

and

$$\fint_{I\times U}(p^*-\nabla u)\cdot\boldsymbol{g}=\frac{1}{|I|}\fint_U(u-\ell_p)^2\ge 0.$$

Therefore, for every $q^*\in\mathbb{R}^d$,

$$\begin{aligned}\mu^*(I\times U,q^*,p^*)&\ge\fint_{I\times U}(-A(\nabla u,\boldsymbol{g},\cdot)+q^*\cdot\nabla u+p^*\cdot\boldsymbol{g})\\&\ge\fint_{I\times U}(-A(\nabla u,\boldsymbol{g},\cdot)+q^*\cdot p^*+\nabla u\cdot\boldsymbol{g})\\&=q^*\cdot p^*-\fint_{I\times U}(A(\nabla u,\boldsymbol{g},\cdot)-\nabla u\cdot\boldsymbol{g}).\end{aligned}$$

By the solvability of the Cauchy–Dirichlet problem (Proposition A.1), for every $p^*\in\mathbb{R}^d$,

$$0=\inf\left\{\int_{I\times U}(A(\nabla u,\boldsymbol{g},\cdot)-\nabla u\cdot\boldsymbol{g}):u\in\ell_{p^*}+H^1_{\mathrm{par},\sqcup}(I\times U),\ \boldsymbol{g}\in L^2(I\times U;\mathbb{R}^d),\ -\nabla\cdot\boldsymbol{g}=-\partial_t u\right\}.$$

Combining the above yields

$$\mu^*(I \times U, q^*, p^*) \geq q^* \cdot p^*.$$

According to Theorem 4.1, we have the $\mathbb{P}$-a.s. limit

$$\lim_{n\to\infty} \mu^*(\square_n, X^*) = \tfrac{1}{2} X^* \cdot \bar{\boldsymbol{A}}^{-1} X^*.$$

We therefore obtain (4-39).

Step 3. We argue that, for every $p \in \mathbb{R}^d$,

$$\inf_{q\in\mathbb{R}^d} (\bar{A}(p,q) - p\cdot q) = 0. \tag{4-40}$$

We have already shown in (4-38) that the infimum on the left is nonnegative. The infimum is attained, by the quadratic growth of $q \mapsto \bar{A}(p,q)$. To see that it is equal to zero, we fix $p \in \mathbb{R}^d$ and select $q \in \mathbb{R}^d$ achieving the infimum. Then

$$\bar{\boldsymbol{A}} \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} * \\ p \end{pmatrix}.$$

Let $q^* \in \mathbb{R}^d$ denote the "$*$" in the previous line, so that

$$\bar{\boldsymbol{A}} \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} q^* \\ p \end{pmatrix}. \tag{4-41}$$

Then using (4-41), we find that

$$\begin{aligned} \frac{1}{2} \begin{pmatrix} q^* \\ p \end{pmatrix} \bar{\boldsymbol{A}}^{-1} \begin{pmatrix} q^* \\ p \end{pmatrix} &= \sup_{p',q'\in\mathbb{R}^d} \left(\begin{pmatrix} q^* \\ p \end{pmatrix} \cdot \begin{pmatrix} p' \\ q' \end{pmatrix} - \frac{1}{2} \begin{pmatrix} p' \\ q' \end{pmatrix} \cdot \bar{\boldsymbol{A}} \begin{pmatrix} p' \\ q' \end{pmatrix} \right) \\ &= \begin{pmatrix} q^* \\ p \end{pmatrix} \cdot \begin{pmatrix} p \\ q \end{pmatrix} - \frac{1}{2} \begin{pmatrix} p \\ q \end{pmatrix} \cdot \bar{\boldsymbol{A}} \begin{pmatrix} p \\ q \end{pmatrix}. \end{aligned}$$

By the previous inequality and (4-39), we discover that

$$p\cdot q^* \leq \begin{pmatrix} q^* \\ p \end{pmatrix} \cdot \begin{pmatrix} p \\ q \end{pmatrix} - \frac{1}{2} \begin{pmatrix} p \\ q \end{pmatrix} \cdot \bar{\boldsymbol{A}} \begin{pmatrix} p \\ q \end{pmatrix} = p\cdot q^* + p\cdot q - \frac{1}{2} \begin{pmatrix} p \\ q \end{pmatrix} \cdot \bar{\boldsymbol{A}} \begin{pmatrix} p \\ q \end{pmatrix}.$$

Rearranging, this yields $\bar{A}(p,q) \leq p\cdot q$, which in view of (4-38) allows us to deduce that $\bar{A}(p,q) = p\cdot q$ and completes the proof of (4-40).

Step 4. We define $\bar{\boldsymbol{a}}$ to be the matrix associated to the linear mapping taking p to the q achieving the infimum in (4-40). That the infimum is achieved at a unique minimum point is a consequence of the uniform convexity of $q \mapsto \bar{A}(p,q)$. That this mapping is linear is due to the fact that $q \mapsto \bar{A}(p,q)$ is quadratic. The bounds (4-36) are a consequence of (4-2). This completes the proof of the proposition. □

5. Quantitative homogenization of the Cauchy–Dirichlet problem

In this section, we demonstrate the passage from the convergence of J to the homogenization of the parabolic operator. In particular, we complete the proof of Theorem 1.1 on the quantitative homogenization

of the Cauchy–Dirichlet problem. The argument is completely deterministic in the sense that the only probabilistic ingredient is the appeal to Theorem 4.1. The argument proceeds in four steps: (i) we show that convergence of J implies convergence of S and $\boldsymbol{A}S$ in H^{-1}; (ii) we use Remark 2.7 to show that there are "finite-volume correctors" which can be found hiding in S and $\boldsymbol{A}S$ and we obtain estimates on them; (iii) we use the finite-volume correctors and a quantitative version of the standard two-scale expansion argument to pass from estimates on the correctors to estimates on the homogenization error for a general Cauchy–Dirichlet problem.

5A. ***Convergence of J maximizers.*** We use the multiscale Poincaré inequality (Proposition 3.6) to obtain information about the weak convergence of $S(\boxdot_n, X, X^*)$ as $n \to \infty$. It is useful to define the quantity

$$\mathcal{E}(V) := \sup_{X, X^* \in B_1} |J(V, X, X^*) - \bar{J}(X, X^*)|,$$

which keeps track of the convergence of J. We also define, given $X, X^* \in \mathbb{R}^{2d}$,

$$\bar{S}(X, X^*) := \bar{\boldsymbol{A}}^{-1} X^* - X.$$

Note that $\bar{S} = \nabla_{X^*} \bar{J}$ and therefore, by (2-29) and the fact that J and $\bar{J}$ are quadratic, we have, for every $X, X^* \in \mathbb{R}^{2d}$,

$$\begin{aligned} \left| \bar{S}(X, X^*) - \fint_V S(\cdot, V, X, X^*) \right| &= |\nabla_{X^*} \bar{J}(X, X^*) - \nabla_{X^*} J(V, X, X^*)| \\ &\le C(|X| + |X^*|)\mathcal{E}(V). \end{aligned} \tag{5-1}$$

Similarly,

$$\begin{aligned} \left| \bar{\boldsymbol{A}}\bar{S}(X, X^*) - \fint_V \boldsymbol{A}S(\cdot, V, X, X^*) \right| &= |\nabla_X \bar{J}(X, X^*) - \nabla_X J(V, X, X^*)| \\ &\le C(|X| + |X^*|)\mathcal{E}(V). \end{aligned} \tag{5-2}$$

That is, we can control the spatial averages of $S(\cdot, V, X, X^*)$ and $\boldsymbol{A}S(\cdot, V, X, X^*)$ in terms of the random variable $\mathcal{E}(V)$. The combination of this observation and Proposition 3.6 yields the following result.

Proposition 5.1 (weak convergence of $(S, \boldsymbol{A}S)$). *There exists $C(d, \Lambda) < \infty$ such that, for every $X, X^* \in B_1$ and $n \in \mathbb{N}$,*

$$3^{-n} \|S(\cdot, \boxdot_n, X, X^*) - \bar{S}(X, X^*)\|_{\underline{\widehat{H}}^{-1}_{\mathrm{par}}(\boxdot_n)} \le C3^{-n} + C \sum_{m=0}^{n-1} 3^{m-n} \left(|\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} (\mathcal{E}(z + \boxdot_m)) \right)^{\frac{1}{2}} \tag{5-3}$$

and

$$3^{-n} \|\boldsymbol{A}S(\cdot, \boxdot_n, X, X^*) - \bar{\boldsymbol{A}}\bar{S}(X, X^*)\|_{\underline{\widehat{H}}^{-1}_{\mathrm{par}}(\boxdot_n)} \le C3^{-n} + C \sum_{m=0}^{n-1} 3^{m-n} \left(|\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} (\mathcal{E}(z + \boxdot_m)) \right)^{\frac{1}{2}}. \tag{5-4}$$

Proof. We fix $X, X^* \in B_1$ and, since it plays no role in the argument, we drop explicit display of the dependence on (X, X^*). According to Proposition 3.6,

$$\|S(\cdot, \boxdot_n) - \bar{S}\|_{\underline{\widehat{H}}^{-1}_{\mathrm{par}}(\boxdot_n)} \le C \|S(\cdot, \boxdot_n) - \bar{S}\|_{\underline{L}^2(\boxdot_n)} + C \sum_{m=0}^{n-1} 3^m \left(|\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} |(S(\cdot, \boxdot_n))_{z + \boxdot_m} - \bar{S}|^2 \right)^{\frac{1}{2}}.$$

To estimate the first term on the right side, we just observe that

$$\|S(\cdot,\square_n)-\bar S\|_{\underline{L}^2(\square_n)} \le \|S(\cdot,\square_n)\|_{\underline{L}^2(\square_n)} + |\bar S| \le C.$$

We next estimate the second term. By the triangle inequality, (5-1) and Lemma 4.5,

$$\begin{aligned}\sum_{z\in\mathcal{Z}_m} |(S(\cdot,\square_n))_{z+\square_m}-\bar S|^2 &\le 2\sum_{z\in\mathcal{Z}_m}\left(|(S(\cdot,z+\square_m))_{z+\square_m}-\bar S|^2+\|S(\cdot,\square_n)-S(\cdot,z+\square_m)\|^2_{\underline{L}^2(z+\square_m)}\right)\\ &\le C\sum_{z\in\mathcal{Z}_m}\mathcal{E}(z+\square_m)+C\sum_{z\in\mathcal{Z}_m}(J(z+\square_m)-J(\square_n))\\ &\le C\sum_{z\in\mathcal{Z}_m}(\mathcal{E}(z+\square_m)+\mathcal{E}(\square_n)).\end{aligned}$$

Thus

$$\begin{aligned}\sum_{m=0}^{n-1} 3^m\left(|\mathcal{Z}_m|^{-1}\sum_{z\in\mathcal{Z}_m}|(S(\cdot,\square_n))_{z+\square_m}-\bar S|^2\right)^{\frac12} &\le C\sum_{m=0}^{n-1}3^m\left(C\mathcal{E}(\square_n)+|\mathcal{Z}_m|^{-1}\sum_{z\in\mathcal{Z}_m}(\mathcal{E}(z+\square_m))\right)^{\frac12}\\ &\le C\sum_{m=0}^{n-1}3^m\left(|\mathcal{Z}_m|^{-1}\sum_{z\in\mathcal{Z}_m}(\mathcal{E}(z+\square_m))\right)^{\frac12}.\end{aligned}$$

Combining the above yields (5-3). The estimate (5-4) is obtained similarly; we just need to use (5-2) instead of (5-1). □

We next give an estimate of the random variable appearing on the right side of (5-3) and (5-4), which is a straightforward consequence of Theorem 4.1. This is the only place in this section where Theorem 4.1 or any other stochastic ingredient is used.

Proposition 5.2. *There exists $\beta(d,\Lambda)$ and, for every $s\in(0,2+d)$, a constant $C(s,d,\Lambda)<\infty$ such that, for every $n\in\mathbb{N}$,*

$$\sum_{m=0}^{n-1}3^{m-n}\left(|\mathcal{Z}_m|^{-1}\sum_{z\in\mathcal{Z}_m}(\mathcal{E}(z+\square_m))\right)^{\frac12} \le C3^{-n\beta(2+d-s)}+\mathcal{O}_1(C3^{-ns}). \tag{5-5}$$

Proof. Fix $s':=\frac13(2s+2+d)$ and $s'':=\frac13(s+2(2+d))$ so that $s<s'<s''<2+d$ with equally sized gaps between these numbers. By Theorem 4.1 and (1-20), we have

$$|\mathcal{Z}_m|^{-1}\sum_{z\in\mathcal{Z}_m}(\mathcal{E}(z+\square_m)) \le C3^{-m\beta(2+d-s'')}+\mathcal{O}_1(C3^{-ms''}).$$

Using the elementary inequality (4-34), we deduce that

$$\left(|\mathcal{Z}_m|^{-1}\sum_{z\in\mathcal{Z}_m}(\mathcal{E}(z+\square_m))\right)^{\frac12} \le C3^{-m\beta\frac{2+d-s''}{2}}+\mathcal{O}_1(C3^{m\beta\frac{2+d-s''}{2}-ms''}).$$

Redefining β to be smaller if necessary, we get

$$\left(|\mathcal{Z}_m|^{-1}\sum_{z\in\mathcal{Z}_m}(\mathcal{E}(z+\square_m))\right)^{\frac12} \le C3^{-m\beta(2+d-s')}+\mathcal{O}_1(C3^{-ms'}).$$

As the left side of the previous line is bounded by C, we can apply [Armstrong et al. 2017b, Lemma A.3] to obtain

$$\left(|\mathcal{Z}_m|^{-1}\sum_{z\in\mathcal{Z}_m}(\mathcal{E}(z+\square_m))\right)^{\frac{1}{2}} \le C3^{-\frac{m\beta}{2+d-s'}} + \mathcal{O}_{d+2}(C3^{-\frac{ms'}{d+2}}).$$

Since $s'/(d+2) \le 1-c$, we may apply (1-20) again to obtain

$$\sum_{m=0}^{n-1} 3^{m-n}\left(|\mathcal{Z}_m|^{-1}\sum_{z\in\mathcal{Z}_m}(\mathcal{E}(z+\square_m))\right)^{\frac{1}{2}} \le C3^{-\frac{n\beta}{2+d-s'}} + \mathcal{O}_{d+2}(C3^{-\frac{ns'}{d+2}}).$$

We conclude by observing that, for any nonnegative random variable X,

$$X \le C3^{-n\beta(2+d-s')} + \mathcal{O}_{d+2}(C3^{-\frac{ns'}{d+2}}) \quad\Longrightarrow\quad X \le C3^{-n\beta(2+d-s)} + \mathcal{O}_1(C3^{-ns}), \tag{5-6}$$

where $\beta(d,\Lambda)$ in the second statement may be smaller than in the first. To see this, we compute

$$\begin{aligned} X \le X + 3^{-n\beta(2+d-s)(1+d)} &\le 3^{n\beta(2+d-s)(1+d)}(X+3^{-n\beta(2+d-s)})^{2+d} \\ &\le C3^{n\beta(2+d-s)(1+d)}(X^{2+d}+3^{-n\beta(2+d-s')(2+d)}) \\ &\le C3^{n\beta(2+d-s)(1+d)}(C3^{-n\beta(2+d-s)(2+d)}+\mathcal{O}_1(C3^{-ns'})) \\ &\le C3^{-n\beta(2+d-s)}+\mathcal{O}_1(C3^{-ns}), \end{aligned}$$

provided that β is small enough that $\beta(2+d-s)(2+d) \le s'-s$. It suffices to require $\beta \le 1/(3(2+d))$. This completes the proof of (5-6) and of the proposition. □

5B. ***Construction of finite-volume correctors.*** We next give the construction of the (finite-volume) correctors. The usage of the term "corrector" in stochastic homogenization is typically reserved for a function with stationary, mean-zero gradient which is the difference of a solution of the equation in the full space and an affine function. For our purposes, it is more convenient to work with a finite-volume approximation of the corrector which will be defined on a large cylinder $\square_n$, because this is what comes most easily and naturally out of the estimates we have already proved above. These correctors will be obtained in a simple way from $S(\cdot, X, X^*)$ and $\boldsymbol{A}S(\cdot, X, X^*)$ and Remark 2.7; the estimates we need for them will be easy consequences of (5-3) and (5-4). The fact that these correctors are not stationary functions defined in the whole space does not create any complication in the proof of Theorem 1.1.

The corrector with slope $e \in \mathbb{R}^d$ on the cylinder $\square_n$ with $n \in \mathbb{N}$ will be denoted by $\phi_{e,n}$. We define it from the maximizers of $J(\square_m, X, 0)$, studied in the previous section. We first must make an appropriate choice of X, depending on e. This is a linear algebra exercise using Proposition 4.8. We set

$$X_e := -\begin{pmatrix} e \\ \bar{\boldsymbol{a}}e \end{pmatrix}$$

and observe from (4-37) that we have

$$\bar{\boldsymbol{A}}X_e = -\begin{pmatrix} \bar{\boldsymbol{a}}e \\ e \end{pmatrix}. \tag{5-7}$$

To check the previous line, we note (see Proposition 4.8) that the map

$$q \mapsto \frac{1}{2}\begin{pmatrix} e \\ q \end{pmatrix} \cdot \bar{\boldsymbol{A}} \begin{pmatrix} e \\ q \end{pmatrix} - e \cdot q \quad \text{attains its minimum at } q = \bar{\boldsymbol{a}} e$$

and the map

$$p \mapsto \frac{1}{2}\begin{pmatrix} p \\ \bar{\boldsymbol{a}} e \end{pmatrix} \cdot \bar{\boldsymbol{A}} \begin{pmatrix} p \\ \bar{\boldsymbol{a}} e \end{pmatrix} - p \cdot \bar{\boldsymbol{a}} e \quad \text{attains its minimum at } p = e.$$

Differentiating in p and q, respectively, gives (5-7).

We next take $u_{e,n} \in H^1_{\mathrm{par}}(\square_{n+1})$ to be the element $u \in \mathcal{A}(\square_{n+1})$ in the representation of $S(\,\cdot\,, \square_{n+1}, X_e, 0)$ given in Lemma 2.6, with additive constant chosen so that $(u_{e,n})_{\square_n} = 0$. Equivalently, in view of Remark 2.7, we can define $u_{e,n}$ to be the function on $\square_{n+1}$ with mean zero on $\square_n$ with gradient given by

$$\nabla u_{e,n} = \tfrac{1}{2}\big(\pi_1 S(\,\cdot\,, \square_{n+1}, X_e, 0) + \pi_2 \boldsymbol{A} S(\,\cdot\,, \square_{n+1}, X_e, 0)\big), \tag{5-8}$$

where π_1 and π_2 denote the projections $\mathbb{R}^{2d} \to \mathbb{R}^d$ onto the first and second d-variables, respectively (that is, $\pi_1(x, y) = x$ and $\pi_2(x, y) = y$ for $x, y \in \mathbb{R}^d$). Note that, by Remark 2.7, we also have the formula

$$\boldsymbol{a} \nabla u_{e,n} = \tfrac{1}{2}\big(\pi_2 S(\,\cdot\,, \square_{n+1}, X_e, 0) + \pi_1 \boldsymbol{A} S(\,\cdot\,, \square_{n+1}, X_e, 0)\big). \tag{5-9}$$

By Proposition 5.1, (5-7), (5-8) and (5-9), we have

$$\begin{aligned} 3^{-n}\|\nabla u_{e,n} - e\|_{\underline{\widehat{H}}^{-1}_{\mathrm{par}}(\square_{n+1})} + 3^{-n}\|\boldsymbol{a}\nabla u_{e,n} - \bar{\boldsymbol{a}} e\|_{\underline{\widehat{H}}^{-1}_{\mathrm{par}}(\square_{n+1})} \\ \le C 3^{-n} + C \sum_{m=0}^{n} 3^{m-n} \left(|\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} (\mathcal{E}(z + \square_m)) \right)^{\frac{1}{2}}. \end{aligned} \tag{5-10}$$

Since $u_{e,n} \in \mathcal{A}(\square_{n+1})$, we have that $u_{e,n}$ is a solution of

$$\partial_t u_{e,n} - \nabla \cdot (\boldsymbol{a} \nabla u_{e,n}) = 0 \quad \text{in } \square_{n+1}. \tag{5-11}$$

The approximate first-order corrector $\phi_{e,n}$ is defined by subtracting the affine function $x \mapsto e \cdot x$ from $u_{e,n}$:

$$\phi_{e,n}(x) := u_{e,n}(x) - e \cdot x.$$

Summarizing, we therefore have that $\phi_{e,n}$ is a solution of

$$\partial_t \phi_{e,n} - \nabla \cdot (\boldsymbol{a}(e + \nabla \phi_{e,n})) = 0 \quad \text{in } \square_{n+1}, \tag{5-12}$$

and satisfies the estimates

$$\begin{aligned} 3^{-n}\big(\|\nabla \phi_{e,n}\|_{\underline{\widehat{H}}^{-1}_{\mathrm{par}}(\square_{n+1})} + \|\boldsymbol{a}(e + \nabla \phi_{e,n}) - \bar{\boldsymbol{a}} e\|_{\underline{\widehat{H}}^{-1}_{\mathrm{par}}(\square_{n+1})}\big) \\ \le C 3^{-n} + C \sum_{m=0}^{n} 3^{m-n} \left(|\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} (\mathcal{E}(z + \square_m)) \right)^{\frac{1}{2}}. \end{aligned} \tag{5-13}$$

By the previous two displays, Proposition 3.7 and $(\phi_{e,n})_{\square_n} = 0$, we also have

$$3^{-n}\|\phi_{e,n}\|_{\underline{L}^2(\square_n)} \le C 3^{-n} + C \sum_{m=0}^{n} 3^{m-n} \left(|\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} (\mathcal{E}(z + \square_m)) \right)^{\frac{1}{2}}. \tag{5-14}$$

5C. ***The proof of Theorem 1.1.*** The main step in the proof of Theorem 1.1 is the following proposition. It is a deterministic estimate of the homogenization error in terms of the error in the convergence of the correctors defined in the previous subsection. Since we have already estimated the latter in (5-5), (5-13) and (5-14), this is sufficient to imply the theorem. It is convenient to define, for every $m \in \mathbb{N}$,

$$\mathcal{E}'(m) := 3^{-m} \sum_{k=1}^{d} \Big(\|\phi_{e_k,m}\|_{\underline{L}^2(\square_m)} + \|\nabla \phi_{e_k,m}\|_{\underline{\widehat{H}}^{-1}_{\mathrm{par}}(\square_m)} + \|\boldsymbol{a}(e_k + \nabla \phi_{e_k,m}) - \bar{\boldsymbol{a}} e_k\|_{\underline{\widehat{H}}^{-1}_{\mathrm{par}}(\square_m)} \Big).$$

We also set, for each $\varepsilon > 0$,

$$\boldsymbol{a}^\varepsilon(t,x) := \boldsymbol{a}\left(\frac{t}{\varepsilon^2}, \frac{x}{\varepsilon}\right) \quad \text{and} \quad \phi^\varepsilon_{e,n}(t,x) := \varepsilon \phi_{e,n}\left(\frac{t}{\varepsilon^2}, \frac{x}{\varepsilon}\right). \tag{5-15}$$

Proposition 5.3. *Fix a bounded interval $I := (I_-, 0) \subseteq \left(-\frac{1}{4}, 0\right)$, a bounded Lipschitz domain $U \subseteq \square_0$, a small parameter $\varepsilon \in \left(0, \frac{1}{2}\right]$, an exponent $\delta > 0$ and an initial-boundary condition $f \in W^{1,2+\delta}_{\mathrm{par}}(I \times U)$. Let*

$$u^\varepsilon, u \in f + H^1_{\mathrm{par},\sqcup}(I \times U)$$

respectively denote the solutions of

$$\begin{cases} \partial_t u^\varepsilon - \nabla \cdot (\boldsymbol{a}^\varepsilon \nabla u^\varepsilon) = 0 & \text{in } I \times U, \\ u^\varepsilon = f & \text{on } \partial_\sqcup (I \times U), \end{cases} \tag{5-16}$$

and

$$\begin{cases} \partial_t u - \nabla \cdot (\bar{\boldsymbol{a}} \nabla u) = 0 & \text{in } I \times U, \\ u = f & \text{on } \partial_\sqcup (I \times U). \end{cases} \tag{5-17}$$

Let $n \in \mathbb{N}$ be such that $3^{-n} \le \varepsilon < 3^{-(n+1)}$. Then there exist $\beta(\delta, d, \Lambda) > 0$ and $C(I, U, \delta, d, \Lambda) < \infty$ such that, for every $r \in (0,1)$, we have the estimate

$$\|\nabla u^\varepsilon - \nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)} + \|\boldsymbol{a}^\varepsilon \nabla u^\varepsilon - \bar{\boldsymbol{a}} \nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)} + \|u^\varepsilon - u\|_{L^2(I\times U)} \le C \|f\|_{W^{1,2+\delta}_{\mathrm{par}}(I\times U)} \left(r^\beta + \frac{1}{r^{3+(2+d)/2}} \mathcal{E}'(n) \right). \tag{5-18}$$

Proof. With n fixed as in the statement of the proposition, we let $\phi_e = \phi_{e,n}$ denote, for each $e \in \mathbb{R}^d$, the (finite-volume) corrector defined in the previous subsection (we will not display its dependence on n). We also use the notation $\phi^\varepsilon_e = \phi^\varepsilon_{e,n}$ as in (5-15).

We will argue that u^ε is close to its modified two-scale expansion suitably cut off near the boundary. The latter is defined by

$$\begin{aligned} w^\varepsilon(t,x) &:= u(t,x) + \varepsilon \zeta_r(t,x) \sum_{k=1}^{d} \partial_{x_k} u(t,x) \phi_{e_k}\left(\frac{t}{\varepsilon^2}, \frac{x}{\varepsilon}\right) \\ &= u(t,x) + \zeta_r(t,x) \sum_{k=1}^{d} \partial_{x_k} u(t,x) \phi^\varepsilon_{e_k}(t,x), \end{aligned} \tag{5-19}$$

where $r \in (0,1)$ is the free parameter (representing a mesoscopic scale) given in the proposition, and we define

$$U_r := \{x \in U : \operatorname{dist}(x, \partial U) > r\} \quad \text{and} \quad I_r := (I_- + r^2, I_+),$$

where the cutoff function ζ_r is selected so that

$$\begin{aligned} &0 \le \zeta_r \le 1, \quad \zeta_r = 1 \quad \text{in } I_{2r} \times U_{2r}, \\ &\zeta_r \equiv 0 \quad \text{in } (I \times U) \setminus (I_r \times U_r), \\ &\text{for all } k, l \in \mathbb{N}, \quad |\nabla^k \partial_t^l \zeta_r| \le C_{k+2l} r^{-(k+2l)}. \end{aligned} \tag{5-20}$$

Note that the constant C_m here depends on (I, U, d) in addition to $m \in \mathbb{N}$.

Step 0. We record some standard estimates from the deterministic regularity theory for uniformly parabolic equations that are needed below. The global Meyers estimate (see Proposition B.2) gives us $\delta_0(U, d, \Lambda) > 0$ such that $\delta \le \delta_0$ implies

$$\|\nabla u^\varepsilon\|_{L^{2+\delta}(I\times U)} + \|\nabla u\|_{L^{2+\delta}(I\times U)} \le C \|f\|_{W^{1,2+\delta}_{\mathrm{par}}(I\times U)}. \tag{5-21}$$

We henceforth assume without loss of generality that $\delta \le \delta_0$ so that (5-21) holds. We also need pointwise derivative estimates for constant-coefficient parabolic equations. These can be found for instance in [Evans 2010, Section 2.3.3.c] (note that estimates for the operator $\partial_t - \nabla \cdot \bar{\boldsymbol{a}} \nabla$ are implied by estimates for the heat equation by a simple affine change of variables), and they yield, for every $m, l \in \mathbb{N}$,

$$\begin{aligned} \|\partial_t^l \nabla^m u\|_{L^\infty(I_r\times U_r)} &\le C_{m+2l} r^{-m-2l} r^{-\frac{2+d}{2}} \|u\|_{L^2(I\times U)} \\ &\le C_{m+2l} r^{-m-2l} r^{1-\frac{2+d}{2}} \|\nabla u\|_{L^2(I\times U)}. \end{aligned} \tag{5-22}$$

Here C_k depends only on (d, Λ) in addition to $k \in \mathbb{N}$.

The main step in the proof is to obtain an estimate on $\|u^\varepsilon - w^\varepsilon\|_{H^1_{\mathrm{par}}(I\times U)}$, which is stated below in (5-25).

Step 1. We plug w^ε into the heterogeneous equation and estimate the error. The claim is that we can write $(\partial_t - \nabla \cdot \boldsymbol{a}^\varepsilon \nabla) w^\varepsilon$ in the form

$$(\partial_t - \nabla \cdot \boldsymbol{a}^\varepsilon \nabla) w^\varepsilon = \partial_t F + G,$$

where $F \in H^1_{\mathrm{par},\sqcup}(I \times U)$ and $G \in L^2(I; H^{-1}(U))$ satisfy the estimates

$$\|F\|_{L^2(I;H^1(U))} + \|G\|_{L^2(I;H^{-1}(U))} \le C\big(r^{\frac{\delta}{4+2\delta}} + r^{-3-\frac{2+d}{2}} \mathcal{E}'(n)\big) \|f\|_{W^{1,2+\delta}_{\mathrm{par}}(I\times U)}. \tag{5-23}$$

We begin by computing

$$\begin{aligned} \nabla w^\varepsilon &= \zeta_r \sum_{k=1}^d (e_k + \nabla \phi^\varepsilon_{e_k})\, \partial_{x_k} u + \sum_{k=1}^d \phi^\varepsilon_{e_k} \nabla(\zeta_r\, \partial_{x_k} u) + (1-\zeta_r)\nabla u, \\ \partial_t w^\varepsilon &= \partial_t u + \zeta_r \sum_{k=1}^d \partial_t \phi^\varepsilon_{e_k}\, \partial_{x_k} u + \sum_{k=1}^d \phi^\varepsilon_{e_k}\, \partial_t(\zeta_r\, \partial_{x_k} u). \end{aligned}$$

According to (5-11), the map $\hat{u}^\varepsilon_e(t,x) := e \cdot x + \phi^\varepsilon_e(t,x)$ is a solution of the equation

$$\partial_t \hat{u}^\varepsilon_e - \nabla \cdot (\boldsymbol{a}^\varepsilon \nabla \hat{u}^\varepsilon_e) = 0 \quad \text{in } I \times U.$$

Therefore we find that

$$\partial_t w^\varepsilon - \nabla\cdot(\boldsymbol{a}^\varepsilon \nabla w^\varepsilon) = \partial_t u + \sum_{k=1}^d \phi_{e_k}^\varepsilon\, \partial_t(\zeta_r\, \partial_{x_k} u) - \sum_{k=1}^d \nabla(\zeta_r\, \partial_{x_k} u)\cdot \boldsymbol{a}^\varepsilon(e_k + \nabla\phi_{e_k}^\varepsilon) \\ -\nabla\cdot\left(\boldsymbol{a}^\varepsilon\left(\sum_{k=1}^d \phi_{e_k}^\varepsilon \nabla(\zeta_r\, \partial_{x_k} u) + (1-\zeta_r)\nabla u\right)\right).$$

Since u satisfies the homogenized equation, we have furthermore that

$$\partial_t u = \nabla\cdot\bar{\boldsymbol{a}}\nabla u = \sum_{k=1}^d \nabla(\zeta_r\, \partial_{x_k} u)\cdot \bar{\boldsymbol{a}} e_k + \nabla\cdot((1-\zeta_r)\,\bar{\boldsymbol{a}}\nabla u),$$

and this gives us the identity

$$\partial_t w^\varepsilon - \nabla\cdot(\boldsymbol{a}^\varepsilon \nabla w^\varepsilon) = \sum_{k=1}^d \phi_{e_k}^\varepsilon\, \partial_t(\zeta_r\, \partial_{x_k} u) - \sum_{k=1}^d \nabla(\zeta_r\, \partial_{x_k} u)\cdot(\boldsymbol{a}^\varepsilon(e_k + \nabla\phi_{e_k}^\varepsilon) - \bar{\boldsymbol{a}} e_k) \\ -\nabla\cdot\left(\boldsymbol{a}^\varepsilon \sum_{k=1}^d \phi_{e_k}^\varepsilon \nabla(\zeta_r\, \partial_{x_k} u)\right) - \nabla\cdot((\boldsymbol{a}^\varepsilon - \bar{\boldsymbol{a}})(1-\zeta_r)\nabla u).$$

According to Lemma 3.11, we can find $v \in L^2(I; H_0^1(U))$ and $v^* \in L^2(I; H^{-1}(U))$ such that

$$f^* := -\sum_{k=1}^d \nabla(\zeta_r\, \partial_{x_k} u)\cdot(\boldsymbol{a}^\varepsilon(e_k + \nabla\phi_{e_k}^\varepsilon) - \bar{\boldsymbol{a}} e_k) = \partial_t v + v^* \tag{5-24}$$

and

$$\|v\|_{L^2(I;H^1(U))} + \|v^*\|_{L^2(I;H^{-1}(U))} \le C\|f^*\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)}.$$

The lemma allows us to take v and v^* to vanish in a neighborhood of the parabolic boundary of $I\times U$. Since the left side of (5-24) belongs to $L^2(I; H^{-1}(U))$, we have also that $v \in H^1_{\mathrm{par},\sqcup}(I\times U)$. Therefore we obtain

$$\partial_t w^\varepsilon - \nabla\cdot(\boldsymbol{a}^\varepsilon \nabla w^\varepsilon) = \partial_t F + G,$$

where

$$F := v$$

and

$$G := v^* + \sum_{k=1}^d \phi_{e_k}^\varepsilon \partial_t(\zeta_r\, \partial_{x_k} u) - \nabla\cdot\left(\boldsymbol{a}^\varepsilon \sum_{k=1}^d \phi_{e_k}^\varepsilon \nabla(\zeta_r\, \partial_{x_k} u)\right) - \nabla\cdot((\boldsymbol{a}^\varepsilon - \bar{\boldsymbol{a}})(1-\zeta_r)\nabla u).$$

It is clear that

$$\begin{aligned}
&\|F\|_{L^2(I;H^1(U))} + \|G\|_{L^2(I;H^{-1}(U))} \\
&\quad\le C\sum_{k=1}^d \|\phi_{e_k}^\varepsilon\, \partial_t(\zeta_r\, \partial_{x_k} u)\|_{L^2(I\times U)} + C\sum_{k=1}^d \|\nabla(\zeta_r\, \partial_{x_k} u)\cdot(\boldsymbol{a}^\varepsilon(e_k + \nabla\phi_{e_k}^\varepsilon) - \bar{\boldsymbol{a}} e_k)\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)} \\
&\qquad + C\left\|\boldsymbol{a}^\varepsilon \sum_{k=1}^d \phi_{e_k}^\varepsilon \nabla(\zeta_r\, \partial_{x_k} u)\right\|_{L^2(I\times U)} + C\|(\boldsymbol{a}^\varepsilon - \bar{\boldsymbol{a}})(1-\zeta_r)\nabla u\|_{L^2(I\times U)} \\
&\quad = T_1 + T_2 + T_3 + T_4.
\end{aligned}$$

We will now show that each of the four terms T_i can be estimated by the right side of (5-23), using the definition of $\mathcal{E}'(n)$ and the bounds (5-20), (5-21) and (5-22). For T_1, we use (5-20) and (5-22) to find that, for each $e \in \partial B_1$,

$$\begin{aligned}\|\phi_{e_k}^\varepsilon \partial_t(\zeta_r\, \partial_{x_k} u)\|_{L^2(I\times U)} &\le C\|\partial_t(\zeta_r\, \partial_{x_k} u)\|_{L^\infty(I\times U)}\|\phi_e^\varepsilon\|_{L^2(I\times U)}\\ &\le Cr^{-3-\frac{2+d}{2}}\mathcal{E}'(n)\|f\|_{H^1_{\mathrm{par}}(I\times U)}.\end{aligned}$$

For T_2, we have

$$\begin{aligned}\|\nabla(\zeta_r\, \partial_{x_k} u)\cdot(\boldsymbol{a}^\varepsilon(e+\nabla\phi_e^\varepsilon)-\bar{\boldsymbol{a}}e)\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)} &\le C\|\nabla(\zeta_r\, \partial_{x_k} u)\|_{W^{1,\infty}(I\times U)}\|\boldsymbol{a}^\varepsilon(e+\nabla\phi_e^\varepsilon)-\bar{\boldsymbol{a}}e\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)}\\ &\le Cr^{-3-\frac{2+d}{2}}\mathcal{E}'(n)\|f\|_{H^1_{\mathrm{par}}(I\times U)}.\end{aligned}$$

For T_3, we use (5-20) and (5-22) again to get

$$\begin{aligned}\|\boldsymbol{a}^\varepsilon\phi_e^\varepsilon\nabla(\zeta_r\, \partial_{x_k} u)\|_{L^2(I\times U)} &\le C\|\nabla(\zeta_r\, \partial_{x_k} u)\|_{L^\infty(I\times U)}\|\phi_e^\varepsilon\|_{L^2(I\times U)}\\ &\le Cr^{-2-\frac{2+d}{2}}\mathcal{E}'(n)\|f\|_{H^1_{\mathrm{par}}(I\times U)}.\end{aligned}$$

Finally, for T_4, we use (5-20), (5-21) and Hölder's inequality to get

$$\begin{aligned}\|(\boldsymbol{a}^\varepsilon-\bar{\boldsymbol{a}})(1-\zeta_r)\nabla u\|_{L^2(I\times U)} &\le C|\{x\in I\times U : \zeta_r(x)\neq 1\}|^{\frac{\delta}{4+2\delta}}\|\nabla u\|_{L^{2+\delta}(I\times U)}\\ &\le Cr^{\frac{\delta}{4+2\delta}}\|f\|_{W^{1,2+\delta}_{\mathrm{par}}(I\times U)}.\end{aligned}$$

This completes the proof of (5-23).

Step 2. We deduce that

$$\|u^\varepsilon - w^\varepsilon\|_{L^2(I;H^1(U))} \le C\big(r^{\frac{\delta}{4+2\delta}} + r^{-3-\frac{2+d}{2}}\mathcal{E}'(n)\big)\|f\|_{W^{1,2+\delta}_{\mathrm{par}}(I\times U)}. \tag{5-25}$$

This is an immediate consequence of the estimate (5-23) proved in the previous step, the fact that $u^\varepsilon - w^\varepsilon$, $F \in H^1_{\mathrm{par},\sqcup}(I\times U)$ and the estimate (A-12) proved in Appendix A.

At this point, we have succeeded in comparing u^ε to w^ε. What is left is to compare w^ε to u by showing that the second term on the right side of (5-19) is small. This is relatively straightforward to obtain from (5-13) and (5-14).

Step 3. We show that

$$\begin{aligned}&\|u-w^\varepsilon\|_{L^2(I\times U)} + \|\nabla u - \nabla w^\varepsilon\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)} + \|\bar{\boldsymbol{a}}\nabla u - \boldsymbol{a}^\varepsilon\nabla w^\varepsilon\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)}\\ &\qquad\le C(r^{\frac{\delta}{4+2\delta}} + r^{-3-\frac{2+d}{2}}\mathcal{E}'(n))\|f\|_{W^{1,2+\delta}_{\mathrm{par}}(I\times U)}.\end{aligned} \tag{5-26}$$

We use the formula

$$\nabla w^\varepsilon(t,x) - \nabla u(t,x) = \varepsilon\sum_{k=1}^d \nabla(\zeta_r\, \partial_{x_k} u)(t,x)\phi_{e_k}\left(\frac{t}{\varepsilon^2},\frac{x}{\varepsilon}\right) + \zeta_r(t,x)\sum_{k=1}^d \partial_{x_k} u(t,x)\nabla\phi_{e_k}\left(\frac{t}{\varepsilon^2},\frac{x}{\varepsilon}\right)$$

to get

$$\begin{aligned}&\|\nabla u - \nabla w^\varepsilon\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)}\\ &\qquad\le \|\nabla(\zeta_r\nabla u)\|_{L^\infty(I\times U)}\,\varepsilon\sum_{k=1}^d\|\phi_{e_k}\|_{\underline{L}^2(Q_{\varepsilon^{-1}})} + C\|\zeta_r\nabla u\|_{W^{1,\infty}(I\times U)}\left\|\nabla\phi_{e_k}\left(\frac{\cdot}{\varepsilon^2},\frac{\cdot}{\varepsilon}\right)\right\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)}\\ &\qquad\le Cr^{-2}\|f\|_{H^1_{\mathrm{par}}(I\times U)}\,\mathcal{E}'(n).\end{aligned}$$

For the fluxes, we find it convenient to use coordinates. We have

$$(\boldsymbol{a}^\varepsilon \nabla w^\varepsilon)_i(t,x) = \sum_{j,k=1}^d \zeta_r(t,x)\, \boldsymbol{a}_{ij}^\varepsilon(t,x)\, \partial_{x_k} u(t,x) \left(\delta_{jk} + \partial_{x_j}\phi_{e_k}\left(\frac{t}{\varepsilon^2}, \frac{x}{\varepsilon}\right)\right) + \varepsilon \sum_{j,k=1}^d \boldsymbol{a}_{ij}^\varepsilon(t,x)\, \partial_{x_j}(\zeta_r\, \partial_{x_k} u)(t,x)\, \phi_{e_k}\left(\frac{t}{\varepsilon^2}, \frac{x}{\varepsilon}\right).$$

Thus

$$\begin{aligned}(\boldsymbol{a}^\varepsilon \nabla w^\varepsilon)_i(t,x) - (\bar{\boldsymbol{a}}\nabla u)_i(t,x) &= \sum_{j,k=1}^d \zeta_r(t,x)\, \partial_{x_k} u(t,x) \left(\boldsymbol{a}_{ij}^\varepsilon(t,x)\left(\delta_{jk} + \partial_{x_j}\phi_{e_k}\left(\frac{t}{\varepsilon^2}, \frac{x}{\varepsilon}\right)\right) - \bar{\boldsymbol{a}}_{ik}\right) \\ &\quad + \sum_{j,k=1}^d (1-\zeta_r(t,x))\, \bar{\boldsymbol{a}}_{ik}\, \partial_{x_k} u(t,x) + \varepsilon \sum_{j,k=1}^d \boldsymbol{a}_{ij}^\varepsilon(t,x)\, \partial_{x_j}(\zeta_r\, \partial_{x_k} u)(t,x)\, \phi_{e_k}\left(\frac{t}{\varepsilon^2}, \frac{x}{\varepsilon}\right).\end{aligned}$$

We can easily estimate the last two terms on the right side using (5-20), (5-21), (5-22) and the Hölder inequality. We have

$$\begin{aligned}\sum_{j,k=1}^d \|(1-\zeta_r)\bar{\boldsymbol{a}}_{ik}\, \partial_{x_k} u\|_{L^2(I\times U)} &\le C\|\nabla u\|_{L^2((I\times U)\setminus(I_r\times U_r))} \\ &\le C r^{\frac{\delta}{4+2\delta}} \|\nabla u\|_{L^{2+\delta}(I\times U)} \le C r^{\frac{\delta}{4+2\delta}} \|f\|_{W^{1,2+\delta}_{\mathrm{par}}(I\times U)}\end{aligned}$$

and

$$\begin{aligned}\varepsilon \sum_{j,k=1}^d \left\| \boldsymbol{a}_{ij}^\varepsilon\, \partial_{x_j}(\zeta_r\, \partial_{x_k} u)\phi_{e_k}\left(\frac{\cdot}{\varepsilon^2}, \frac{\cdot}{\varepsilon}\right)\right\|_{L^2(I\times U)} &\le C\|\nabla(\zeta_r \nabla u)\|_{L^\infty(I\times U)}\, \varepsilon \sum_{k=1}^d \|\phi_{e_k}\|_{\underline{L}^2(Q_{1/2})} \\ &\le C r^{-2-\frac{2+d}{2}} \|f\|_{H^1_{\mathrm{par}}(I\times U)}\, \mathcal{E}'(n).\end{aligned}$$

For the first term, we have

$$\begin{aligned}&\sum_{j,k=1}^d \left\| \zeta_r\, \partial_{x_k} u \left(\boldsymbol{a}_{ij}^\varepsilon\left(\delta_{jk} + \partial_{x_j}\phi_{e_k}\left(\frac{\cdot}{\varepsilon^2}, \frac{\cdot}{\varepsilon}\right)\right) - \bar{\boldsymbol{a}}_{ik}\right)\right\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)} \\ &\qquad \le C \sum_{j,k=1}^d \|\zeta_r\, \partial_{x_k} u\|_{W^{1,\infty}(I\times U)} \left\| \boldsymbol{a}_{ij}^\varepsilon\left(\delta_{jk} + \partial_{x_j}\phi_{e_k}\left(\frac{\cdot}{\varepsilon^2}, \frac{\cdot}{\varepsilon}\right)\right) - \bar{\boldsymbol{a}}_{ik}\right\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)} \\ &\qquad \le C r^{-3} \|f\|_{H^1_{\mathrm{par}}(I\times U)} \sum_{k=1}^d \|\boldsymbol{a}^\varepsilon(e_k + \nabla\phi_{e_k}) - \bar{\boldsymbol{a}} e_k\|_{\widehat{H}^{-1}_{\mathrm{par}}(Q_{1/2})} \\ &\qquad \le C r^{-3-\frac{2+d}{2}} \|f\|_{H^1_{\mathrm{par}}(I\times U)} \mathcal{E}'(n).\end{aligned}$$

Combining the previous four displays, we obtain

$$\|\boldsymbol{a}^\varepsilon \nabla w^\varepsilon - \bar{\boldsymbol{a}}\nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)} \le C\left(r^{\frac{\delta}{4+2\delta}} + r^{-3-\frac{2+d}{2}} \mathcal{E}'(n)\right) \|f\|_{W^{1,2+\delta}_{\mathrm{par}}(I\times U)}.$$

Finally, for the estimate of $w^\varepsilon - u$, we have

$$\begin{aligned}\|w^\varepsilon - u\|_{L^2(I\times U)} &\le C\|\nabla u\|_{L^\infty((I\times U)\setminus(I_r\times U_r))}\varepsilon \sum_{k=1}^d \left\| \phi_{e_k}\left(\frac{\cdot}{\varepsilon^2},\frac{\cdot}{\varepsilon}\right)\right\|_{L^2(I\times U)} \\ &\le Cr^{-1-\frac{2+d}{2}}\|f\|_{H^1_{\mathrm{par}}(I\times U)}\mathcal{E}'(n).\end{aligned}$$

This completes the proof of (5-26).

Step 4. We summarize and conclude the argument. According to (5-25), (5-26) and the triangle inequality, we have

$$\begin{aligned}\|\nabla u^\varepsilon - \nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)} &\le \|\nabla u^\varepsilon - \nabla w^\varepsilon\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)} + \|\nabla w^\varepsilon - \nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)} \\ &\le C\|u^\varepsilon - w^\varepsilon\|_{L^2(I;H^1(U))} + \|\nabla w^\varepsilon - \nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)} \\ &\le C(r^{\frac{\delta}{4+2\delta}} + r^{-3-\frac{2+d}{2}}\mathcal{E}'(n))\|f\|_{W^{1,2+\delta}_{\mathrm{par}}(I\times U)}.\end{aligned}$$

Similarly, for the fluxes we have

$$\begin{aligned}\|\boldsymbol{a}^\varepsilon\nabla u^\varepsilon - \bar{\boldsymbol{a}}\nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)} &\le \|\boldsymbol{a}^\varepsilon\nabla u^\varepsilon - \boldsymbol{a}^\varepsilon\nabla w^\varepsilon\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)} + \|\boldsymbol{a}^\varepsilon\nabla w^\varepsilon - \bar{\boldsymbol{a}}\nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)} \\ &\le C\|u^\varepsilon - w^\varepsilon\|_{L^2(I;H^1(U))} + \|\boldsymbol{a}^\varepsilon\nabla w^\varepsilon - \bar{\boldsymbol{a}}\nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)} \\ &\le C(r^{\frac{\delta}{4+2\delta}} + r^{-3-\frac{2+d}{2}}\mathcal{E}'(n))\|f\|_{W^{1,2+\delta}_{\mathrm{par}}(I\times U)},\end{aligned}$$

and, for the homogenization error, we have

$$\begin{aligned}\|u^\varepsilon - u\|_{L^2(I\times U)} &\le \|u^\varepsilon - w^\varepsilon\|_{L^2(I\times U)} + \|w^\varepsilon - u\|_{L^2(I\times U)} \\ &\le C(r^{\frac{\delta}{4+2\delta}} + r^{-3-\frac{2+d}{2}}\mathcal{E}'(n))\|f\|_{W^{1,2+\delta}_{\mathrm{par}}(I\times U)}.\end{aligned}$$

□

To complete the proof of Theorem 1.1, we just need to estimate the random variables on the right side of (5-18) using Proposition 5.2 and the estimates (5-13) and (5-14) for the correctors.

Proof of Theorem 1.1. Fix $s \in (0, 2+d)$ and put $s' := \frac{1}{2}s + \frac{1}{2}(2+d)$ and $s'' := \frac{1}{2}s' + \frac{1}{2}(2+d)$. Thus $s < s' < s'' < 2+d$ and the gaps are at least of size $\frac{1}{4}(2+d-s)$. Observe that (5-13) and (5-14) imply

$$\mathcal{E}'(n) \le C3^{-n} + C\sum_{m=0}^{n-1} 3^{m-n}\left(|\mathcal{Z}_m|^{-1}\sum_{z\in\mathcal{Z}_m}(\mathcal{E}(z+\square_m))\right)^{\frac{1}{2}}.$$

Thus, by Proposition 5.2,

$$\mathcal{E}'(n) \le C3^{-n\beta(2+d-s'')} + \mathcal{O}_1(C3^{-ns''}).$$

Hence

$$3^{ns'}(\mathcal{E}'(n) - C3^{-n\beta(2+d-s'')})_+ \le \mathcal{O}_1(C3^{-n(s''-s')}).$$

By (1-20),

$$\mathcal{X} := \sum_{n\in\mathbb{N}} 3^{ns'}(\mathcal{E}'(n) - C3^{-n\beta(2+d-s'')})_+ \le \mathcal{O}_1(C).$$

Proposition 5.3 yields therefore that, for every $\varepsilon \in \left(0, \frac{1}{2}\right]$ and $r \in (0,1)$,

$$\|\nabla u^\varepsilon - \nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)} + \|\boldsymbol{a}^\varepsilon \nabla u^\varepsilon - \bar{\boldsymbol{a}}\nabla u\|_{\widehat{H}^{-1}_{\mathrm{par}}(I\times U)} + \|u^\varepsilon - u\|_{\underline{L}^2(I\times U)} \leq C\|f\|_{W^{1,p}_{\mathrm{par}}(I\times U)}\left(r^\beta + \frac{1}{r^{3+(2+d)/2}}(\varepsilon^{\beta(2+d-s'')} + \varepsilon^{s'}\mathcal{X}')\right).$$

We now select $r \in (0,1)$ as small as possible (it must be no larger than a positive power of ε) such that $r^{-3-(2+d)/2}\varepsilon^{s'} \leq \varepsilon^s$ and $r^{-3-(2+d)/2} \leq \varepsilon^{-\beta(2+d-s'')/2}$. We can take for example

$$r := \varepsilon^{\beta(2+d-s'')/(3+(2+d)/2)} \vee \varepsilon^{(s'-s)/(3+(2+d)/2)}.$$

Recalling that $2+d-s'' \geq \frac{1}{4}(2+d-s)$ and $s'-s \geq \frac{1}{4}(2+d-s)$, we obtain the theorem. □

6. Regularity theory

In this section, we sketch the proof of Theorem 1.2, following along the lines of the argument given in the proof of [Armstrong et al. 2017b, Theorem 3.6] in the elliptic case. We do not give full details, since this would involve an almost verbatim repetition of the proof of the latter.

We begin by reformulating Theorem 1.1 in a slightly different way in terms of *caloric approximation*, which is more convenient for its application in this section. The next statement can be compared to its elliptic analogue in [Armstrong et al. 2017b, Proposition 3.2].

Proposition 6.1 (caloric approximation). *Fix $s \in (0, 2+d)$. There exist an exponent $\alpha(d,\Lambda) > 0$, a constant $C(s,d,\Lambda) < \infty$, and a random variable $\mathcal{X}_s : \Omega \to [1,\infty]$ satisfying the estimate*

$$\mathcal{X}_s = \mathcal{O}_s(C) \tag{6-1}$$

such that the following holds: for every $R \geq \mathcal{X}_s$ and weak solution $u \in H^1_{\mathrm{par}}(Q_R)$ of

$$\partial_t u - \nabla\cdot(\boldsymbol{a}\nabla u) = 0 \quad \text{in } Q_R, \tag{6-2}$$

there exists a solution $\bar{u} \in H^1_{\mathrm{par}}(Q_{R/2})$ of the equation

$$\partial_t \bar{u} - \nabla\cdot(\bar{\boldsymbol{a}}\nabla \bar{u}) = 0 \quad \text{in } Q_{R/2}$$

such that

$$\|u - \bar{u}\|_{\underline{L}^2(Q_{R/2})} \leq C R^{-\alpha(2+d-s)}\|u - (u)_{Q_R}\|_{\underline{L}^2(Q_R)}. \tag{6-3}$$

Proof. This is a simple application of Theorem 1.1 combined with the parabolic Meyers estimate. The argument is almost the same as in the elliptic case presented in [Armstrong et al. 2017b, Proposition 3.2]; we just need to replace the elliptic interior Meyers estimate with its parabolic analogue proved in Proposition B.1 below. The latter gives us $\delta(d,\Lambda) > 0$ and $C(d,\Lambda) > 0$ such that, for every $u \in H^1_{\mathrm{par}}(Q_R)$ satisfying (6-2), we have $\nabla u \in L^{2+\delta}(Q_{R/2})$ and the estimate

$$\|\nabla u\|_{\underline{L}^{2+\delta}(Q_{R/2})} \leq \frac{C}{R}\|u - (u)_{Q_R}\|_{\underline{L}^2(Q_R)}. \tag{6-4}$$

Following the proof of [Armstrong et al. 2017b, Proposition 3.2], using (6-4), substituting Theorem 1.1 in place of Theorem 2.16 there and making obvious changes to the notation, we obtain the proposition. □

We next state a parabolic counterpart of [Armstrong et al. 2017b, Lemma 3.5].

Lemma 6.2. *Fix $\alpha \in [0,1]$, $K \geq 1$ and $X \geq 1$. Let $R \geq 2X$ and $u \in L^2(Q_R)$ have the property that, for every $r \in [X, R]$, there exists $w_r \in H^1_{\mathrm{par}}(Q_{r/2})$ which is a solution of*

$$\partial_t w_r - \nabla \cdot (\bar{\boldsymbol{a}} \nabla w_r) = 0 \quad \text{in } Q_{r/2}$$

and satisfies

$$\|u - w_r\|_{\underline{L}^2(Q_{r/2})} \leq K r^{-\alpha} \|u - (u)_{Q_r}\|_{\underline{L}^2(Q_r)}. \tag{6-5}$$

Then, for every $k \in \mathbb{N}$, there exists $\theta(\alpha, k, d, \Lambda) \in \left(0, \frac{1}{2}\right)$ and $C(\alpha, k, d, \Lambda) < \infty$ such that, for every $r \in [X, R]$,

$$\inf_{p \in \bar{\mathcal{A}}_k(Q_\infty)} \|u - p\|_{\underline{L}^2(Q_{\theta r})} \leq \tfrac{1}{4}\theta^{k+1-\frac{\alpha}{2}} \inf_{p \in \bar{\mathcal{A}}_k(Q_\infty)} \|u - p\|_{\underline{L}^2(Q_r)} + CKr^{-\alpha}\|u - (u)_{Q_r}\|_{\underline{L}^2(Q_r)}. \tag{6-6}$$

Proof. The proof is essentially the same as that of [Armstrong et al. 2017b, Lemma 3.5]. We just have to substitute balls for parabolic cylinders and use Proposition 6.1 in place of its elliptic version. These changes cause no additional complexity in the proof. □

With Lemma 6.2 in hand, the proof of Theorem 1.2 is now completed in the same way as the one of [Armstrong et al. 2017b, Theorem 3.6], by following the argument almost verbatim and making only obvious modifications. We refer to that book for the details.

Appendix A. Variational structure of uniformly parabolic equations

The aim of this appendix is to show that the solution of the parabolic equation (2-1) can be obtained as the minimizer of a uniformly convex functional. We will prove this result in the more general context of uniformly monotone operators, since this causes no modification to the proof. Although our statement differs in detail, it is close to the main result of [Ghoussoub and Tzou 2004]; see also [Ghoussoub 2009]. The proof we give is also relatively close to that of [Ghoussoub and Tzou 2004]; we hope that the reader will appreciate the short and self-contained presentation in this appendix. The fact that a parabolic equation can be cast as the first variation of a uniformly convex integral functional was first discovered in [Brezis and Ekeland 1976a; 1976b].

Let $I := (0, T) \subseteq \mathbb{R}$ and $U \subseteq \mathbb{R}^d$ be a bounded Lipschitz domain. For a given right-hand side w^* and boundary condition (both of which will be made precise below), we study the solvability of the parabolic equation

$$\partial_t u - \nabla \cdot (\boldsymbol{a}(\nabla u, \cdot)) = w^* \quad \text{in } I \times U, \tag{A-1}$$

where the dot represents the time-space variable in $I \times U \subseteq \mathbb{R}^{1+d}$, and $\boldsymbol{a} \in L^\infty_{\mathrm{loc}}(\mathbb{R}^d \times \mathbb{R}^{1+d}; \mathbb{R}^d)$ is Lipschitz and uniformly monotone in its first argument. That is, we assume that there exists a constant $\lambda < \infty$ such that, for every $p_1, p_2 \in \mathbb{R}^d$ and $z \in \mathbb{R}^{1+d}$,

$$\begin{aligned} |\boldsymbol{a}(p_1, z) - \boldsymbol{a}(p_2, z)| &\leq \lambda |p_1 - p_2|, \\ (\boldsymbol{a}(p_1, z) - \boldsymbol{a}(p_2, z)) \cdot (p_1 - p_2) &\geq \lambda^{-1} |p_1 - p_2|^2. \end{aligned} \tag{A-2}$$

As a first step, we introduce a variational representation of the mapping $p \mapsto \boldsymbol{a}(p,z)$, for each $z \in \mathbb{R}^{1+d}$. This idea is often attributed to Fitzpatrick [1988], although it actually appeared in [Krylov 1982] several years earlier.

By [Armstrong and Mourrat 2016, Theorem 2.9], there exists $A \in L^\infty_{\mathrm{loc}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{1+d})$ satisfying the following properties for $\Lambda := 2\lambda + 1$ and for each $z \in \mathbb{R}^{1+d}$:

- The mapping
$$(p,q) \mapsto A(p,q,z) - \frac{1}{2\Lambda}(|p|^2 + |q|^2) \quad \text{is convex.} \tag{A-3}$$
- The mapping
$$(p,q) \mapsto A(p,q,z) - \frac{\Lambda}{2}(|p|^2 + |q|^2) \quad \text{is concave.} \tag{A-4}$$
- For every $p, q \in \mathbb{R}^d$, we have
$$A(p,q,z) \geq p \cdot q, \tag{A-5}$$
and
$$A(p,q,z) = p \cdot q \quad \Longleftrightarrow \quad q = \boldsymbol{a}(p,z). \tag{A-6}$$

In the particular case when $p \mapsto \boldsymbol{a}(p,z)$ is linear, we can define the mapping $(p,q) \mapsto A(p,q,z)$ according to (2-3); see Lemma 2.2. Another familiar example is when $\boldsymbol{a}(\cdot\,,z)$ is the gradient of a uniformly convex Lagrangian $L(p,z)$; that is, $\boldsymbol{a}(\cdot\,,z) = \nabla_p L(\cdot\,,z)$, where $p \mapsto L(p,z)$ is uniformly convex. In this case, we can take

$$A(p,q,z) := L(p,z) + L^*(q,z),$$

where L^* is the Legendre–Fenchel transform of L. We remark that the choice of A is in general not unique.

We define the function space

$$Z(I \times U) := \{(u,u^*) : u \in L^2(I; H^1(U)) \text{ and } (u^* - \partial_t u) \in L^2(I; H^{-1}(U))\},$$

with norm

$$\|(u,u^*)\|_{Z(I\times U)} := \|u\|_{L^2(I;H^1(U))} + \|u^* - \partial_t u\|_{L^2(I;H^{-1}(U))}.$$

The function space $H^1_{\mathrm{par}}(I \times U)$ is defined in (1-9)–(1-10). We denote by $H^1_{\mathrm{par},\sqcup}(I \times U)$ the closure in $H^1_{\mathrm{par}}(I \times U)$ of the set of smooth functions with compact support in $(0,T] \times U$. For every $(u,u^*) \in Z(I \times U)$, we set

$$\mathcal{J}[u,u^*] := \inf\left\{ \int_{I\times U} (A(\nabla u, \boldsymbol{g}, \cdot) - \nabla u \cdot \boldsymbol{g}) : -\nabla \cdot \boldsymbol{g} = u^* - \partial_t u \right\}. \tag{A-7}$$

In the infimum above, we understand that $\boldsymbol{g} \in L^2(I \times U; \mathbb{R}^d)$, and the last condition is interpreted as

$$\text{for all } \phi \in L^2(I; H^1_0(U)), \quad \int_{I\times U} \nabla\phi \cdot \boldsymbol{g} = \int_{I\times U} \phi(u^* - \partial_t u). \tag{A-8}$$

Note that the set of candidates for $\boldsymbol{g}$ is not empty; indeed, denoting by Δ_U^{-1} the solution operator for the Laplacian in U with a null Dirichlet boundary condition, we verify that

$$\boldsymbol{g} = \nabla \Delta_U^{-1}(u^* - \partial_t u)$$

is a suitable candidate, by the assumption of $u^* - \partial_t u \in L^2(I; H^{-1}(U))$.

The goal of this appendix is to prove the following proposition.

Proposition A.1. *For each $(w, w^*) \in Z(I \times U)$, the mapping*

$$\begin{aligned} w + H^1_{\mathrm{par},\sqcup}(I \times U) &\to \mathbb{R}, \\ u &\mapsto \mathcal{J}[u, w^*], \end{aligned} \tag{A-9}$$

is uniformly convex. Moreover, its minimum is zero, and the associated minimizer is the unique $u \in w + H^1_{\mathrm{par},\sqcup}(I \times U)$ solution of (A-1), in the sense that

$$\text{for all } \phi \in L^2(I; H^1_0(U)), \quad \int_{I\times U} \nabla\phi \cdot \boldsymbol{a}(\nabla u, \cdot) = \int_{I\times U} \phi(w^* - \partial_t u).$$

Remark A.2. By the inclusion

$$H^1_{\mathrm{par}}(I \times U) \times L^2(I; H^{-1}(U)) \subseteq Z(I \times U), \tag{A-10}$$

Proposition A.1 ensures in particular the solvability of the parabolic equation (A-1) for every right-hand side $w^* \in L^2(I; H^{-1}(U))$ and every boundary condition $w \in H^1_{\mathrm{par}}(I \times U)$; the solution thus obtained then belongs to $H^1_{\mathrm{par}}(I \times U)$.

More generally, for every w^* of the form

$$w^* = \partial_t f + v, \quad f \in L^2(I; H^1_0(U)),\ v \in L^2(I; H^{-1}(U)), \tag{A-11}$$

we have $(f, w^*) \in Z(I \times U)$ and hence Proposition A.1 yields the existence of a unique solution $u \in f + H^1_{\mathrm{par},\sqcup}(I \times U)$ of (A-1) which satisfies the estimate

$$\|u - f\|_{H^1_{\mathrm{par}}(I\times U)} \le C(\|f\|_{L^2(I;H^1(U))} + \|w^* - \partial_t f\|_{L^2(I;H^{-1}(U))}). \tag{A-12}$$

In other words, we have identified a mapping

$$\partial_t f + v \mapsto f + P(f, v), \tag{A-13}$$

where $f \in L^2(I; H^1_0(U))$, $v \in L^2(I; H^{-1}(U))$, and P is a bounded linear operator from $L^2(I; H^1_0(U)) \times L^2(I; H^{-1}(U))$ to $H^1_{\mathrm{par},\sqcup}(I \times U)$. If we moreover restrict our attention, say, to the set of functions f which vanish in a neighborhood of $\{0\} \times U$, then this mapping provides us with a notion of a solution of (A-1) with null Dirichlet boundary condition on the parabolic boundary of $I \times U$. This additional regularity assumption on the behavior of f near the initial time can of course be weakened as desired.

Note that every w^* of the form (A-11) belongs to $\widehat{H}^{-1}_{\mathrm{par}}(I \times U)$, but the latter space is strictly larger than the set of such w^*. This may at first glance appear at odds with Lemma 3.11; however that lemma required that u^* belong to $L^2(I \times U)$. This hypothesis rules out certain singular distributions which belong to $\widehat{H}^{-1}_{\mathrm{par}}(I \times U)$ but cannot be written in the form (A-11).

Remark A.3. One may wonder if, in analogy with the elliptic setting, one can identify a reflexive subspace E of the space of distributions such that the standard heat operator $(\partial_t - \Delta)$ maps E to its dual E^* surjectively. This is however not possible, as we now explain briefly. Observe first that by Proposition A.1, the heat operator is a bijective mapping from $H^1_{\mathrm{par},\sqcup}(I \times U)$ to $L^2(I; H^{-1}(U))$, and that $L^2(I; H^{-1}(U))$ is strictly smaller than the dual of $H^1_{\mathrm{par},\sqcup}(I \times U)$. Indeed, the dual of $H^1_{\mathrm{par},\sqcup}(I \times U)$ contains all elements of the form $\partial_t v$ for $v \in L^2(I; H^1_0(U))$. Hence, the space E should be strictly between the spaces $H^1_{\mathrm{par},\sqcup}(I \times U)$ and $L^2(I; H^1_0(U))$. Using the decomposition of the solution operator in (A-13), one can then verify that such a space E does not exist.

Before turning to the proof of Proposition A.1, we first recall the following continuity result for elements of a space intermediate between $H^1_{\mathrm{par},\sqcup}(I \times U)$ and $H^1_{\mathrm{par}}(I \times U)$ where the null boundary condition is only imposed in the space direction. We refer to [Temam 1979, Section III.1.4] for a proof.

Lemma A.4. *Let $u \in L^2(I; H^1_0(U))$ be such that $\partial_t u \in L^2(I; H^{-1}(U))$. There exists $\tilde{u} \in C(\bar{I}; L^2(U))$ such that, for almost every $t \in I$, we have $u(t, \cdot) = \tilde{u}(t, \cdot)$.*

From now on, whenever a function u satisfies the conditions of Lemma A.4, we identify it with its continuous representative.

Proof of Proposition A.1. We decompose the proof into four steps.

Step 1. We show that the mapping in (A-9) is uniformly convex. We will in fact prove the stronger statement that the mapping

$$(u, \boldsymbol{g}) \mapsto \int_{I \times U} (A(\nabla u, \boldsymbol{g}, \cdot) - \nabla u \cdot \boldsymbol{g}), \tag{A-14}$$

defined over all pairs $(u, \boldsymbol{g})$ in the set

$$\{(u, \boldsymbol{g}) \in (w + H^1_{\mathrm{par},\sqcup}(I \times U)) \times L^2(I \times U; \mathbb{R}^d) \text{ and } -\nabla \cdot \boldsymbol{g} = w^* - \partial_t u\}, \tag{A-15}$$

is uniformly convex. We first show that the mapping

$$(u, \boldsymbol{g}) \mapsto -\int_{I \times U} \nabla u \cdot \boldsymbol{g}$$

is convex over the set defined in (A-15). By (A-8) (with u^* replaced by w^*), we have

$$\begin{aligned} -\int_{I \times U} \nabla u \cdot \boldsymbol{g} &= -\int_{I \times U} \nabla w \cdot \boldsymbol{g} + \int_{I \times U} (w - u)(w^* - \partial_t u) \\ &= -\int_{I \times U} \nabla w \cdot \boldsymbol{g} + \int_{I \times U} (w - u)(w^* - \partial_t w) + \tfrac{1}{2} \|(u - w)(T, \cdot)\|^2_{L^2(U)}. \end{aligned} \tag{A-16}$$

This expression is clearly convex in the pair $(u, \boldsymbol{g})$. We now complete this step by showing that the mapping

$$(u, \boldsymbol{g}) \mapsto \int_{I \times U} A(\nabla u, \boldsymbol{g}, \cdot)$$

is uniformly convex over the set defined in (A-15). By (A-3), for every $(u, \boldsymbol{g})$ in the set defined in (A-15) and

$$(v, \boldsymbol{h}) \in H^1_{\mathrm{par},\sqcup}(I \times U) \times L^2(I \times U; \mathbb{R}^d) \quad \text{such that } \nabla \cdot \boldsymbol{h} = \partial_t v, \tag{A-17}$$

we have

$$\tfrac{1}{2} A(\nabla(u+v), \boldsymbol{g}+\boldsymbol{h}, \cdot) + \tfrac{1}{2} A(\nabla(u-v), \boldsymbol{g}-\boldsymbol{h}, \cdot) - A(\nabla u, \boldsymbol{g}, \cdot) \geq \frac{1}{2\Lambda}(|\nabla v|^2 + |\boldsymbol{h}|^2).$$

Moreover, by (A-17),

$$\begin{aligned} \|\partial_t v\|_{L^2(I;H^{-1}(U))} &= \sup\left\{ \int_{I\times U} \phi\, \partial_t v \,:\, \phi \in L^2(I; H^1_0(U)),\ \|\nabla\phi\|_{L^2(I\times U)} \leq 1 \right\} \\ &= \sup\left\{ \int_{I\times U} \nabla\phi \cdot \boldsymbol{h} \,:\, \phi \in L^2(I; H^1_0(U)),\ \|\nabla\phi\|_{L^2(I\times U)} \leq 1 \right\} \leq \|\boldsymbol{h}\|_{L^2(I\times U)}. \end{aligned}$$

We have thus shown that

$$\begin{aligned} \int_{I\times U} \left(\tfrac{1}{2} A(\nabla(u+v), \boldsymbol{g}+\boldsymbol{h}, \cdot) + \tfrac{1}{2} A(\nabla(u-v), \boldsymbol{g}-\boldsymbol{h}, \cdot) - A(\nabla u, \boldsymbol{g}, \cdot) \right) \\ \geq \frac{1}{4\Lambda}(\|\nabla v\|^2_{L^2(I\times U)} + \|\partial_t v\|^2_{L^2(I;H^{-1}(U))} + \|\boldsymbol{h}\|^2_{L^2(I\times U)}), \end{aligned}$$

so the proof of uniform convexity is complete.

Step 2. By the result of the previous step, there exists a unique pair $(u_0, \boldsymbol{g}_0)$ in the set defined by (A-15) which minimizes the functional in (A-14). In order to complete the proof, it suffices to show that

$$\int_{I\times U} (A(\nabla u_0, \boldsymbol{g}_0, \cdot) - \nabla u_0 \cdot \boldsymbol{g}_0) = 0. \tag{A-18}$$

Indeed, by (A-5), the identity (A-18) implies

$$\boldsymbol{g}_0 = \boldsymbol{a}(\nabla u_0, \cdot) \quad \text{a.e. in } I \times U,$$

and moreover, by (A-15),

$$\nabla \cdot \boldsymbol{g}_0 = w^* - \partial_t u_0,$$

so that u_0 indeed solves

$$\partial_t u_0 - \nabla \cdot (\boldsymbol{a}(\nabla u_0, \cdot)) = w^*$$

in the weak sense. Our goal is therefore to show (A-18). The fact that the left side of (A-18) is nonnegative is immediate from (A-5). There remains to show that this quantity is nonpositive; that is,

$$\inf_{u \in H^1_{\mathrm{par},\sqcup}(I\times U)} \mathcal{J}[w+u, w^*] \leq 0. \tag{A-19}$$

In order to do so, we consider the perturbed convex minimization problem defined for every $u^* \in L^2(I; H^{-1}(U))$ by

$$G(u^*) := \inf_{u \in H^1_{\mathrm{par},\sqcup}(I\times U)} \left(\mathcal{J}[w+u, w^*+u^*] + \int_{I\times U} u\, u^* \right).$$

Note that (A-19) is equivalent to the statement that $G(0) \le 0$. By the computation in (A-16), for every $u^* \in L^2(I; H^{-1}(U))$ and

$$(u, \boldsymbol{g}) \in H^1_{\mathrm{par},\sqcup}(I \times U) \times L^2(I \times U; \mathbb{R}^d) \quad \text{such that } -\nabla \cdot \boldsymbol{g} = w^* + u^* - \partial_t(w+u), \tag{A-20}$$

we have

$$\int_{I\times U} \big(A(\nabla(w+u), \boldsymbol{g}, \cdot) - \nabla(w+u)\cdot \boldsymbol{g}\big) + \int_{I\times U} u\, u^* = \int_{I\times U} \big(A(\nabla(w+u), \boldsymbol{g}, \cdot) - \nabla w \cdot \boldsymbol{g} - u(w^* - \partial_t w)\big) + \tfrac{1}{2}\|u(T,\cdot)\|^2_{L^2(U)}, \tag{A-21}$$

and hence the function G is convex over $L^2(I; H^{-1}(U))$. Moreover, one can check that it is also locally bounded above, which implies that G is lower semicontinuous, by convexity; see, e.g., [Ekeland and Temam 1976, Lemma I.2.1 and Corollary I.2.2]. Denoting by G^* the convex dual of G, defined for every $v \in L^2(I; H^1_0(U))$ by

$$G^*(v) := \sup_{u^* \in L^2(I;H^{-1}(U))} \left(-G(u^*) + \int_{I\times U} v\, u^*\right),$$

and by G^{**} its bidual, we deduce that $G = G^{**}$, see [Ekeland and Temam 1976, Proposition I.4.1], and in particular,

$$G(0) = G^{**}(0) = \sup_{v \in L^2(I;H^1_0(U))} (-G^*(v)).$$

The statement (A-19) is therefore equivalent to

$$\text{for all } v \in L^2(I; H^1_0(U)), \quad G^*(v) \ge 0. \tag{A-22}$$

The proof of this fact occupies the next two steps.

Step 3. For each $v \in L^2(I; H^1_0(U))$, we have $G^*(v) \in \mathbb{R} \cup \{+\infty\}$. In this step, we show that

$$G^*(v) < +\infty \quad \Longrightarrow \quad \partial_t v \in L^2(I; H^{-1}(U)). \tag{A-23}$$

We note that

$$G^*(v) = \sup\left\{ \int_{I\times U} \big((v-u)\, u^* - A(\nabla(w+u), \boldsymbol{g}, \cdot) + \nabla(w+u)\cdot \boldsymbol{g}\big) : u^* \in L^2(I; H^{-1}(U)),\ (u, \boldsymbol{g}) \text{ satisfy (A-20)} \right\}. \tag{A-24}$$

Restricting to $u^* = \partial_t u$ and to a fixed $\boldsymbol{g} \in L^2(I \times U; \mathbb{R}^d)$ satisfying $-\nabla \cdot \boldsymbol{g} = w^* - \partial_t w$ (which can be constructed as the gradient of the solution of a Dirichlet problem) yields the lower bound

$$G^*(v) \ge \sup\left\{ \int_{I\times U} \big(v\, \partial_t u - A(\nabla(w+u), \boldsymbol{g}, \cdot) + \nabla(w+u)\cdot \boldsymbol{g}\big) - \tfrac{1}{2}\|u(T,\cdot)\|^2_{L^2(U)} : u \in H^1_{\mathrm{par},\sqcup}(I\times U)\right\}.$$

The assumption of $G^*(v) < \infty$ thus implies

$$\sup\left\{ \int_{I\times U} v\, \partial_t u \ :\ u \in H^1_{\mathrm{par},\sqcup}(I\times U),\ \|\nabla u\|_{L^2(I\times U)} \le 1,\ \|u(T,\cdot)\|_{L^2(U)} \le 1 \right\} < \infty.$$

Denoting the supremum above by $C < \infty$, we infer that for every smooth test function u with compact support in $I \times U$,

$$\left| \int_{I \times U} u \, \partial_t v \right| \leq C \|\nabla u\|_{L^2(I \times U)}.$$

By density, we deduce that $\partial_t v$ can be identified with an element of the dual of $L^2(I; H_0^1(U))$. Since this dual space is $L^2(I; H^{-1}(U))$, the proof of (A-23) is complete.

Step 4. In this step, we show that

$$v \in L^2(I; H_0^1(U)) \text{ and } \partial_t v \in L^2(I; H^{-1}(U)) \quad \Longrightarrow \quad G^*(v) \geq 0. \tag{A-25}$$

Together with (A-23), this would complete the proof of (A-22) and therefore of the proposition.

The fact that $G^*(v) \geq 0$ would follow immediately from (A-24) if we could choose $u = v$ and then ensure the equality of the last two terms under the integral. The difficulty we face is that the function u is allowed to range in $H^1_{\mathrm{par},\sqcup}(I \times U)$, while the function v does not belong to this space in general, due to the boundary condition at the initial time. We therefore wish to argue that this constraint on u can be relaxed.

Replacing u^* by $u^* + \partial_t u$ in the supremum in (A-24), we can rewrite $G^*(v)$ as

$$G^*(v) = \sup \left\{ \int_{I \times U} \left((v-u)\,(u^* + \partial_t u) - A(\nabla(w+u), \boldsymbol{g}, \cdot) + \nabla(w+u) \cdot \boldsymbol{g} \right) \right\}, \tag{A-26}$$

where the supremum is taken over every $u^* \in L^2(I; H^{-1}(U))$, $u \in H^1_{\mathrm{par},\sqcup}(I \times U)$ and $\boldsymbol{g} \in L^2(I \times U; \mathbb{R}^d)$ satisfying

$$-\nabla \cdot \boldsymbol{g} = w^* + u^* - \partial_t w. \tag{A-27}$$

Integrating by parts, we can rewrite the term involving $\partial_t u$ on the right side of (A-26) as

$$\int_{I \times U} (v-u)\, \partial_t u = -\int_{I \times U} u \, \partial_t v + \int_U u(T, \cdot) v(T, \cdot) - \tfrac{1}{2} \|u(T, \cdot)\|^2_{L^2(U)}.$$

The functional under the supremum in (A-26) can thus be decomposed into the sum of

$$I_1(u, u^*, \boldsymbol{g}) := \int_{I \times U} \left((v-u)u^* - u \, \partial_t v - A(\nabla(w+u), \boldsymbol{g}, \cdot) + \nabla(w+u) \cdot \boldsymbol{g} \right) \tag{A-28}$$

and

$$I_2(u(T, \cdot)) := \int_U u(T, \cdot) v(T, \cdot) - \tfrac{1}{2} \|u(T, \cdot)\|^2_{L^2(U)}. \tag{A-29}$$

Moreover, for each given $u^* \in L^2(I; H^{-1}(U))$ and $\boldsymbol{g} \in L^2(I \times U; \mathbb{R}^d)$, the mapping $u \mapsto I_1(u, u^*, \boldsymbol{g})$ is continuous for the topology of $L^2(I; H^1(U))$. For any given $b \in H_0^1(U)$ and $\tilde{u} \in L^2(I; H_0^1(U))$, one can find elements of the space

$$\{u \in H^1_{\mathrm{par},\sqcup}(I \times U) : u(T, \cdot) = b\}$$

which approximate $\tilde{u}$ with arbitrary precision for the topology of $L^2(I; H^1(U))$. Hence, for each given $u^* \in L^2(I; H^{-1}(U))$ and $\boldsymbol{g} \in L^2(I \times U; \mathbb{R}^d)$, we have

$$\sup\{I_1(u, u^*, \boldsymbol{g}) + I_2(u(T, \cdot)) : u \in H^1_{\mathrm{par},\sqcup}(I \times U)\} \geq \sup\{I_1(u, u^*, \boldsymbol{g}) + I_2(b) : u \in L^2(I; H^1_0(U)) \text{ and } b \in H^1_0(U)\}.$$

Moreover, the mapping $b \mapsto I_2(b)$ is continuous for the topology of $L^2(U)$, and thus we have in fact

$$\sup\{I_1(u, u^*, \boldsymbol{g}) + I_2(u(T, \cdot)) : u \in H^1_{\mathrm{par},\sqcup}(I \times U)\} = \sup\{I_1(u, u^*, \boldsymbol{g}) + I_2(b) : u \in L^2(I; H^1_0(U)) \text{ and } b \in L^2(U)\}.$$

Selecting $u = v$ and $b = v(T, \cdot)$, we have thus shown

$$G^*(v) \geq \sup\left\{\tfrac{1}{2}\|v(T, \cdot)\|^2_{L^2(U)} + \int_{I\times U}\big(-v\,\partial_t v - A(\nabla(w+v), \boldsymbol{g}, \cdot) + \nabla(w+v)\cdot\boldsymbol{g}\big)\right\},$$

where the supremum is taken over every $u^* \in L^2(I; H^{-1}(U))$ and $\boldsymbol{g} \in L^2(I \times U; \mathbb{R}^d)$ satisfying (A-27). Note that

$$\tfrac{1}{2}\|v(T, \cdot)\|^2_{L^2(U)} - \int_{I\times U} v\,\partial_t v = \tfrac{1}{2}\|v(0, \cdot)\|^2_{L^2(U)} \geq 0.$$

Selecting u^* such that

$$-\nabla\cdot(\boldsymbol{a}(\nabla(w+v), \cdot)) = w^* + u^* - \partial_t w,$$

and then

$$\boldsymbol{g} = \boldsymbol{a}(\nabla(w+v), \cdot),$$

ensures that the constraint (A-27) is satisfied, and by (A-6), that

$$\int_{I\times U}\big(A(\nabla(w+v), \boldsymbol{g}, \cdot) - \nabla(w+v)\cdot\boldsymbol{g}\big) = 0.$$

The proof of (A-25) is therefore complete. □

Appendix B. Meyers-type estimates

In this appendix, we present local and global versions of the Meyers improvement of integrability estimate for gradients of solutions of linear, uniformly parabolic equations with measurable coefficients.

The interior Meyers estimate in the parabolic case was first proved in [Giaquinta and Struwe 1982]. We follow their argument to obtain Proposition B.1 below, which is included for completeness and since the same ideas are needed to prove the global version in Proposition B.2. The statement of the latter will certainly not come as a surprise to experts, but we do not believe it has appeared before. Global versions of the Meyers estimate in the parabolic setting have been previously considered in [Parviainen 2009], but the statement of Proposition B.2 is stronger than the results of that paper since we do not require any additional regularity of the boundary condition in time — a modest technical improvement, but it gives a more natural statement and one which is useful for the application in this paper.

In what follows, we use the same notation for parabolic cylinders as in Section 6; see (1-7). That is, for $(t, x) \in \mathbb{R} \times \mathbb{R}^d$, we define

$$\tilde{I}_r := (-r^2, 0], \quad Q_r(t, x) := (t, x) + \tilde{I}_r \times B_r, \quad \text{and} \quad Q_r := Q_r(0, 0).$$

We fix a coefficient field $\boldsymbol{a} = \boldsymbol{a}(t, x)$ satisfying (1-2) for every $(t, x) \in \mathbb{R} \times \mathbb{R}^d$, and consider the linear parabolic equation

$$\partial_t u - \nabla \cdot (\boldsymbol{a}(t, x)\nabla u) = u^*. \tag{B-1}$$

We remark that the argument we present only makes mild use of linearity and can be adapted to give similar estimates for solutions of nonlinear parabolic equations like the ones considered in Appendix A.

We first present the interior Meyers estimate. Recall that the space $W^{1,p}_{\mathrm{par}}$ is defined in (1-13) and (1-14).

Proposition B.1 (interior Meyers estimate [Giaquinta and Struwe 1982, Theorem 2.1]). *Fix $r > 0$, $p \geq 2$ and suppose $u \in H^1_{\mathrm{par}}(Q_{2r})$ and $u^* \in L^p(I_{2r}; W^{-1,p}(B_{2r}))$ satisfy (B-1) in Q_{2r}. There exist an exponent $\delta(d, \Lambda) > 0$ and a constant $C(d, \Lambda) < \infty$ such that $u \in W^{1,p\wedge(2+\delta)}_{\mathrm{par}}(Q_r)$ and we have the estimate*

$$\|\nabla u\|_{\underline{L}^{p\wedge(2+\delta)}(Q_r)} \leq C(\|\nabla u\|_{\underline{L}^2(Q_{2r})} + \|u^*\|_{\underline{L}^{p\wedge(2+\delta)}(I_{2r}; \underline{W}^{-1,p\wedge(2+\delta)}(B_{2r}))}). \tag{B-2}$$

We next give a global statement of the Meyers estimate with respect to a Cauchy–Dirichlet initial-boundary condition.

Proposition B.2 (global Meyers estimate). *Fix $p \geq 2$. Let $U \in \mathbb{R}^d$ be a bounded Lipschitz domain, $I \subseteq \mathbb{R}$ a bounded interval and set $V := I \times U$. Fix $f \in W^{1,p}_{\mathrm{par}}(V)$, $u^* \in L^p(I; W^{-1,p}(V))$ and suppose*

$$u \in f + H^1_{\mathrm{par},\sqcup}(V)$$

is the unique solution of the Cauchy–Dirichlet problem

$$\begin{cases} \partial_t u - \nabla \cdot (\boldsymbol{a}\nabla u) = u^* & \text{in } V, \\ u = f & \text{on } \partial_\sqcup V. \end{cases}$$

There exist $\delta(V, d, \Lambda) > 0$ and a constant $C(V, d, \Lambda) < \infty$ such that $u \in W^{1,p\wedge(2+\delta)}_{\mathrm{par}}(V)$ and we have the estimate

$$\|u\|_{W^{1,p\wedge(2+\delta)}_{\mathrm{par}}(V)} \leq C(\|f\|_{W^{1,p\wedge(2+\delta)}_{\mathrm{par}}(V)} + \|u^*\|_{L^{p\wedge(2+\delta)}(I; W^{-1,p\wedge(2+\delta)}(V))}). \tag{B-3}$$

The Meyers estimates are consequences of the Caccioppoli inequality, the most basic regularity estimate for divergence-form equations.

Lemma B.3 (parabolic Caccioppoli inequality). *Suppose $u \in H^1_{\mathrm{par}}(Q_{2r})$ and $u^* \in L^2(I_{2r}; H^{-1}(B_{2r}))$ satisfy*

$$\partial_t u - \nabla \cdot (\boldsymbol{a}\nabla u) = u^* \quad \text{in } Q_{2r}.$$

Then there exists $C(d, \Lambda) < \infty$ such that

$$\|\nabla u\|_{L^2(Q_r)} \leq Cr^{-1}\|u\|_{L^2(Q_{2r})} + C\|u^*\|_{L^2(I_{2r}; H^{-1}(B_{2r}))} \tag{B-4}$$

and

$$\sup_{s \in I_r} \|u(s, \cdot)\|_{L^2(B_r)} \leq C\|\nabla u\|_{L^2(Q_{2r})} + C\|u^*\|_{L^2(I_{2r}; H^{-1}(B_{2r}))}. \tag{B-5}$$

Proof. We take $\eta_r \in C_c^\infty(Q_{2r})$ to be a test function satisfying

$$0 \le \eta \le 1, \qquad \eta \equiv 1 \quad \text{on } Q_r, \qquad |\partial_t \eta| + |\nabla \eta|^2 \le C r^{-2}.$$

We test the weak formulation

$$\text{for all } \phi \in L^2(I_{2r}; H_0^1(B_{2r})), \quad \int_{Q_{2r}} \phi(u^* - \partial_t u) = \int_{Q_{2r}} \nabla \phi \cdot \boldsymbol{a} \nabla u$$

with the function $\phi := \eta_r^2 u \in L^2(I_{2r}; H_0^1(B_{2r}))$. We estimate the right side from below by

$$\begin{aligned}
\int_{Q_{2r}} \nabla \phi \cdot \boldsymbol{a} \nabla u &\ge \frac{1}{\Lambda} \int_{Q_{2r}} \eta_r^2 |\nabla u|^2 - C \int_{Q_{2r}} \eta_r |\nabla \eta_r| |u| |\nabla u| \\
&\ge \frac{1}{2\Lambda} \int_{Q_{2r}} \eta_r^2 |\nabla u|^2 - C \int_{Q_{2r}} |\nabla \eta_r|^2 |u|^2 \\
&\ge \frac{1}{2\Lambda} \int_{Q_{2r}} \eta_r^2 |\nabla u|^2 - C r^{-2} \int_{Q_{2r}} |u|^2
\end{aligned}$$

and the left side from above by

$$\begin{aligned}
\int_{Q_{2r}} \eta_r^2 u(u^* - \partial_t u) &\le - \int_{Q_{2r}} \partial_t \left(\tfrac{1}{2} \eta_r^2 u^2 \right) + \int_{Q_{2r}} \eta_r |\partial_t \eta_r| u^2 + \int_{-4r^2}^0 \| (\eta_r^2 u)(t, \cdot) \|_{H^1(B_{2r})} \| u^*(t, \cdot) \|_{H^{-1}(B_{2r})} \, dt \\
&\le -\frac{1}{2} \int_{B_{2r}} \eta_r^2(0, x) u^2(0, x) \, dx + C r^{-2} \int_{Q_{2r}} u^2 + C \| \eta_r^2 u \|_{L^2(I_{2r}; H^1(B_{2r}))} \| u^* \|_{L^2(I_{2r}; H^{-1}(B_{2r}))}.
\end{aligned}$$

Using that

$$\| \eta_r^2 u \|_{L^2(I_{2r}; H^1(B_{2r}))} \le C r^{-1} \| u \|_{L^2(I_{2r} \times B_{2r})} + C \| \eta_r \nabla u \|_{L^2(I_{2r} \times B_{2r})},$$

we get

$$\begin{aligned}
&C \| \eta_r^2 u \|_{L^2(I_{2r}; H^1(B_{2r}))} \| u^* \|_{L^2(I_{2r}; H^{-1}(B_{2r}))} \\
&\qquad \le r^{-2} \| u \|_{L^2(I_{2r} \times B_{2r})}^2 + \frac{1}{4\Lambda} \| \eta_r \nabla u \|_{L^2(I_{2r} \times B_{2r})}^2 + C \| u^* \|_{L^2(I_{2r}; H^{-1}(B_{2r}))}^2.
\end{aligned}$$

Combining the above, we get

$$\frac{1}{2} \int_{B_{2r}} \eta_r^2(0, x) \, u^2(0, x) \, dx + \frac{1}{4\Lambda} \int_{Q_{2r}} \eta_r^2 |\nabla u|^2 \le C r^{-2} \int_{Q_{2r}} |u|^2 + C \| u^* \|_{L^2(I_{2r}; H^{-1}(B_{2r}))}^2.$$

This yields (B-4).

By repeating the above computation, using instead the test function $\phi := \eta_r^2 u \mathbb{1}_{\{t < s\}}$ for fixed $s \in I_{2r}$, and estimating the right side of the weak formulation from below differently, namely

$$\begin{aligned}
\int_{Q_{2r}} \nabla \phi \cdot \boldsymbol{a} \nabla u &\ge -C \| \eta_r \nabla u \|_{L^2(Q_{2r})}^2 - C \| \nabla \eta_r \nabla u \|_{L^2(Q_{2r})} \| u \eta_r \|_{L^2(Q_{2r})} \\
&\ge -C \| \nabla u \|_{L^2(Q_{2r})}^2 - \tfrac{1}{16} r^{-2} \int_{Q_{2r}} \eta_r^2 u^2 \\
&\ge -C \| \nabla u \|_{L^2(Q_{2r})}^2 - \tfrac{1}{4} \sup_{t \in I_{2r}} \int_{B_{2r}} \eta_r^2(t, x) \, u^2(t, x) \, dx,
\end{aligned}$$

we get the bound

$$\frac{1}{2}\int_{B_{2r}} \eta_r^2(s,x)\,u^2(s,x)\,dx \le C\|\nabla u\|^2_{L^2(Q_{2r})} + \frac{1}{4}\sup_{t\in I_{2r}}\int_{B_{2r}} \eta_r^2(t,x)\,u^2(t,x)\,dx + C\|u^*\|^2_{L^2(I_{2r};H^{-1}(B_{2r}))}.$$

Taking the supremum over $s \in I_{2r}$ and rearranging, we get (B-5). □

In the following statement, what is important is that $q < 2$. It is convenient to use the Sobolev exponent $q := 2_*$, although the choice $q = 1$ in $d = 2$ causes technical problems so in that case we just take $q \in \left(\frac{5}{4}, \frac{7}{4}\right)$.

Lemma B.4 (reverse Hölder inequality). *Suppose $u \in H^1_{\mathrm{par}}(Q_{4r})$ and $u^* \in L^2(I_{4r}; H^{-1}(B_{4r}))$ satisfy*

$$\partial_t u - \nabla\cdot(\boldsymbol{a}(x)\nabla u) = u^* \quad \text{in } Q_{4r}.$$

Set $q := 2_* = 2d/(2+d)$ if $d > 2$ or let q be any element of $\left(\frac{5}{4}, \frac{7}{4}\right)$ if $d = 2$. Then there exists $C(d,\Lambda) < \infty$ such that, for every $\alpha > 0$,

$$\|\nabla u\|^2_{\underline{L}^2(Q_r)} \le \frac{C}{\alpha}\|\nabla u\|^2_{\underline{L}^q(Q_{4r})} + \alpha\|\nabla u\|^2_{\underline{L}^2(Q_{4r})} + C\|u^*\|^2_{\underline{L}^2(I_{4r};\underline{H}^{-1}(B_{4r}))}. \tag{B-6}$$

Proof. By subtracting a constant, we may suppose $(u)_{Q_{2r}} = 0$. Let $\xi \in C_c^\infty(B_r)$ with $\int_{B_r}\xi = 1$ and $|\nabla\xi| \le Cr^{-1}$. Define

$$v(t,x) := u(t,x) - w(t), \qquad w(t) := \int_{B_r}\xi(y)u(t,y)\,dy.$$

Then v satisfies

$$\partial_t v - \nabla\cdot(\boldsymbol{a}\nabla v) = u^* - \partial_t w.$$

Applying (B-5) to v, we find that

$$\begin{aligned}\int_{Q_{2r}}|v|^2 &\le \left(\sup_{s\in I_{2r}}\int_{B_{2r}}|v(s,x)|^2\,dx\right)^{\frac12}\int_{I_{2r}}\left(\int_{B_{2r}}|v(t,x)|^2\,dx\right)^{\frac12}dt\\ &\le C\left(\|\nabla u\|_{L^2(Q_{4r})} + \|u^* - \partial_t w\|_{L^2(I_{4r};H^{-1}(B_{4r}))}\right)\int_{I_{2r}}\left(\int_{B_{2r}}|v(t,x)|^2\,dx\right)^{\frac12}dt.\end{aligned}$$

Denote by q' the Hölder conjugate exponent to q and notice that $q' = 2^*$ in $d > 2$ and $q' < \infty$ in $d = 2$. Using the Hölder and Sobolev inequalities, we find that

$$\begin{aligned}\int_{I_{2r}}\left(\int_{B_{2r}}|v(t,x)|^2\,dx\right)^{\frac12}dt &\le \int_{I_{2r}}\left(\int_{B_{2r}}|v(t,x)|^q\,dx\right)^{\frac{1}{2q}}\left(\int_{B_{2r}}|v(t,x)|^{q'}\,dx\right)^{\frac{1}{2q'}}dt\\ &\le Cr^{1+d\left(\frac14-\frac{1}{2q}\right)}\int_{I_{2r}}\left(\int_{B_{2r}}|\nabla v(t,x)|^q\,dx\right)^{\frac{1}{2q}}\left(\int_{B_{2r}}|\nabla v(t,x)|^2\,dx\right)^{\frac14}dt\\ &\le Cr^{1+d\left(\frac14-\frac{1}{2q}\right)}\|\nabla v\|^{\frac12}_{L^q(Q_{2r})}\left(\int_{I_{2r}}\left(\int_{B_{2r}}|\nabla v(t,x)|^2\,dx\right)^{\frac{(2q)'}{4}}dt\right)^{\frac{1}{(2q)'}}.\end{aligned}$$

As $\frac{1}{4}(2q)' \le \frac{1}{2} < 1$, we can use Hölder's inequality in time and then (B-4) and Lemma 3.1 to get

$$\begin{aligned}\left(\int_{I_{2r}}\left(\int_{B_{2r}}|\nabla v(t,x)|^2\,dx\right)^{\frac{(2q)'}{4}}dt\right)^{\frac{2}{(2q)'}} &\le Cr^{\frac{2}{(2q)'}-\frac{1}{2}}\left(\int_{I_{2r}}\int_{B_{2r}}|\nabla v(t,x)|^2\,dx\,dt\right)^{\frac{1}{2}}\\ &\le Cr^{\frac{2}{(2q)'}-\frac{1}{2}}(\|\nabla u\|_{L^2(Q_{4r})}+\|u^*-\partial_t w\|_{L^2(I_{4r};H^{-1}(B_{4r}))}).\end{aligned}$$

Let $\kappa := d\left(\frac{1}{4}-\frac{1}{2q}\right)+\frac{1}{(2q)'}+\frac{3}{4}$. Combining the above, we get

$$\|v\|^2_{L^2(Q_{2r})} \le Cr^{\kappa}\|\nabla v\|^{\frac{1}{2}}_{\underline{L}^q(Q_{2r})}(\|\nabla u\|_{L^2(Q_{4r})}+\|u^*-\partial_t w\|_{L^2(I_{4r};H^{-1}(B_{4r}))})^{\frac{3}{2}}.$$

Combining (B-4) and the previous inequality, we obtain

$$\|\nabla v\|^2_{L^2(Q_r)} \le Cr^{\kappa-2}\|\nabla v\|^{\frac{1}{2}}_{\underline{L}^q(Q_{2r})}(\|\nabla v\|_{L^2(Q_{4r})}+\|u^*-\partial_t w\|_{L^2(I_{4r};H^{-1}(B_{4r}))})^{\frac{3}{2}}+C\|u^*-\partial_t w\|^2_{L^2(I_{4r};H^{-1}(B_{4r}))}.$$

Normalizing the norms, we find that this is the same as

$$\|\nabla v\|^2_{\underline{L}^2(Q_r)} \le \|\nabla v\|^{\frac{1}{2}}_{\underline{L}^q(Q_{2r})}(\|\nabla v\|_{\underline{L}^2(Q_{4r})}+\|u^*-\partial_t w\|_{\underline{L}^2(I_{4r};\underline{H}^{-1}(B_{4r}))})^{\frac{3}{2}}+C\|u^*-\partial_t w\|^2_{\underline{L}^2(I_{4r};\underline{H}^{-1}(B_{4r}))}.$$

Applying Young's inequality, we obtain, for every $\alpha > 0$,

$$\|\nabla v\|^2_{\underline{L}^2(Q_r)} \le \frac{C}{\alpha}\|\nabla v\|^2_{\underline{L}^q(Q_{4r})}+\alpha\|\nabla v\|^2_{\underline{L}^2(Q_{4r})}+C\|u^*-\partial_t w\|^2_{\underline{L}^2(I_{4r};\underline{H}^{-1}(B_{4r}))}. \tag{B-7}$$

It is not difficult to show, by using the equation and the definition of w, that

$$\|\partial_t w\|_{\underline{L}^2(I_{4r};\underline{H}^{-1}(B_{4r}))} \le C(\|\nabla u\|_{\underline{L}^1(Q_{4r})}+\|u^*\|_{\underline{L}^2(I_{4r};\underline{H}^{-1}(B_{4r}))}).$$

Combining the previous two displays yields

$$\|\nabla v\|^2_{\underline{L}^2(Q_r)} \le \frac{C}{\alpha}\|\nabla v\|^2_{\underline{L}^q(Q_{4r})}+\alpha\|\nabla v\|^2_{\underline{L}^2(Q_{4r})}+C\|u^*\|^2_{\underline{L}^2(I_{4r};\underline{H}^{-1}(B_{4r}))}.$$

Since $\nabla v = \nabla u$, this completes the argument. □

To complete the proof of the interior Meyers estimate, we need the following version of Gehring's lemma for parabolic cylinders which states that a reverse Hölder inequality implies an improvement of integrability. This result is standard and so we do not give the proof here. See for instance [Giaquinta and Modica 1979, Proposition 5.1], where the statement is given in cubes rather than parabolic cylinders (which makes no difference in its proof).

Lemma B.5 (Gehring-type lemma). *Assume that $R > 0$, $q > 1$, $F \in L^1(Q_{4R})$, $G \in L^q(Q_{4R})$, $m \in (0,1)$ and $A \in [1,\infty)$. Suppose that, for every $(t,x) \in Q_R$ and $r \in \left(0, \frac{1}{2}R\right]$,*

$$\|F\|_{L^1(Q_r(t,x))} \le A(\|F^m\|^{\frac{1}{m}}_{L^1(Q_{4r}(t,x))}+\|G\|_{L^1(Q_{4r}(t,x))})+\varepsilon\|F\|_{L^1(Q_{4r}(t,x))}.$$

Then there exists $\varepsilon_0(d,m) \in \left(0,\frac{1}{2}\right]$ such that $\varepsilon \le \varepsilon_0$ implies the existence of an exponent $\delta(\varepsilon,A,m,q,d) \in \left(0,\frac{1}{2}\right]$ and $C(\varepsilon,A,m,d) < \infty$ such that $F \in L^{1+\delta}(Q_R)$ and

$$\|F\|_{L^{1+\delta}(Q_R)} \le C(\|F\|_{L^1(Q_{4R})}+\|G\|_{L^{1+\delta}(Q_{4R})}).$$

The statement of Proposition B.1 can now be obtained as a consequence of Lemmas B.4 and B.5 and a routine covering argument. Indeed, any element of $W^{-1,p}(U)$ can be represented as the divergence of an element of $L^p(U;\mathbb{R}^d)$ by the Riesz representation theorem; see [Adams and Fournier 2003, Theorem 3.9]. This allows us to obtain the estimate (B-2). The statement that $\partial_t u$ belongs to $L^{p\wedge(2+\delta)}(I_r; W^{-1,p\wedge(2+\delta)}(B_r))$, with an appropriate estimate, follows from (B-2) and (B-1).

We next give a sketch of the proof of Proposition B.2, which requires us to first revisit the proof of the Caccioppoli inequality to obtain a global version.

Lemma B.6 (global Caccioppoli inequality). *Let $U \subseteq \mathbb{R}^d$ be a bounded Lipschitz domain and define $V := I_2 \times U$. Suppose $f \in H^1_{\mathrm{par}}(V)$, $u \in f + H^1_{\mathrm{par},\sqcup}(V)$ and $u^* \in L^2(I_2; H^{-1}(U))$ satisfy*

$$\partial_t u - \nabla\cdot(\boldsymbol{a}(x)\nabla u) = u^* \quad \text{in } V.$$

Then there exists $C(V,d,\Lambda) < \infty$ such that, for every $r \in (0,1)$ and $(t,x) \in I_1 \times U$,

$$\begin{aligned}\|\nabla(u-f)\|_{L^2(Q_r(t,x)\cap V)} &\le Cr^{-1}\|u-f\|_{L^2(Q_{2r}(t,x)\cap V)} + C\|\nabla f\|_{L^2(Q_{2r}(t,x)\cap V)} + C\|u^*\|_{L^2((t+I_{2r})\cap I_2; H^{-1}(B_{2r}(x)\cap U))} \qquad \text{(B-8)}\end{aligned}$$

and

$$\begin{aligned}\sup_{s\in(t+I_{2r})\cap I_2}\|(u-f)(s,\cdot)\|_{L^2(B_r(x)\cap U)} &\le C\|\nabla(u-f)\|_{L^2(Q_{2r}(t,x)\cap V)} + C\|\nabla f\|_{L^2(Q_{2r}(t,x)\cap V)} + C\|u^*\|_{L^2((t+I_{2r})\cap I_2; H^{-1}(B_{2r}(x)\cap U))}. \qquad \text{(B-9)}\end{aligned}$$

Proof. By replacing u by $\tilde{u} := u - f$ and u^* by $\tilde{u}^* := u^* - (\partial_t - \nabla\cdot\boldsymbol{a}\nabla) f$, we may assume without loss of generality that $f = 0$. The lemma is then obtained by repeating the argument of Lemma B.3 and making obvious adjustments to the notation. □

Following the proof of Lemma B.4, we obtain a global version of the reverse Hölder inequality.

Lemma B.7 (reverse Hölder inequality). *Let $U \subseteq \mathbb{R}^d$ be a bounded Lipschitz domain and define $V := I_2 \times U$. Suppose $f \in H^1_{\mathrm{par}}(V)$, $u \in f + H^1_{\mathrm{par},\sqcup}(V)$ and $u^* \in L^2(I_2; H^{-1}(U))$ satisfy*

$$\partial_t u - \nabla\cdot(\boldsymbol{a}(x)\nabla u) = u^* \quad \text{in } V.$$

Set $q := 2_* = 2d/(2+d)$ if $d > 2$ or let q be any element of $\left(\frac{5}{4}, \frac{7}{4}\right)$ if $d = 2$. Then there exists $C(V,d,\Lambda) < \infty$ such that, for every $r \in (0,1)$, $(t,x) \in I_1 \times U$ and $\alpha > 0$,

$$\begin{aligned}\|\nabla(u-f)\|^2_{\underline{L}^2(Q_r(t,x)\cap V)} \le \frac{C}{\alpha}\|\nabla(u-f)\|^2_{\underline{L}^q(Q_{4r}(t,x)\cap V)} + \alpha\|\nabla(u-f)\|^2_{\underline{L}^2(Q_{4r}(t,x)\cap V)} \\ + \alpha\|\nabla f\|^2_{\underline{L}^2(Q_{4r}(t,x)\cap V)} + C\|u^*\|^2_{\underline{L}^2(t+I_{4r}\cap I_2; \underline{H}^{-1}(B_{4r}(x)\cap U))}.\end{aligned}$$

Proof. The argument is omitted, since it is an easy adaptation of the proof of Lemma B.4. □

Proposition B.2 is now a straightforward consequence of Lemmas B.5 and B.7.

Acknowledgments

Armstrong was partially supported by the NSF Grant DMS-1700329. Mourrat was partially supported by the ANR grant LSD (ANR-15-CE40-0020-03).

References

[Adams and Fournier 2003] R. A. Adams and J. J. F. Fournier, *Sobolev spaces*, 2nd ed., Pure and Applied Mathematics **140**, Elsevier/Academic, Amsterdam, 2003. MR Zbl

[Armstrong and Mourrat 2016] S. N. Armstrong and J.-C. Mourrat, "Lipschitz regularity for elliptic equations with random coefficients", *Arch. Ration. Mech. Anal.* **219**:1 (2016), 255–348. MR Zbl

[Armstrong and Smart 2016] S. N. Armstrong and C. K. Smart, "Quantitative stochastic homogenization of convex integral functionals", *Ann. Sci. Éc. Norm. Supér.* (4) **49**:2 (2016), 423–481. MR Zbl

[Armstrong et al. 2016] S. Armstrong, T. Kuusi, and J.-C. Mourrat, "Mesoscopic higher regularity and subadditivity in elliptic homogenization", *Comm. Math. Phys.* **347**:2 (2016), 315–361. MR Zbl

[Armstrong et al. 2017a] S. Armstrong, T. Kuusi, and J.-C. Mourrat, "The additive structure of elliptic homogenization", *Invent. Math.* **208**:3 (2017), 999–1154. MR Zbl

[Armstrong et al. 2017b] S. Armstrong, T. Kuusi, and J.-C. Mourrat, "Quantitative stochastic homogenization and large-scale regularity", book preprint, 2017, available at http://www.math.ens.fr/∼mourrat/lecturenotes.pdf.

[Bella et al. 2018] P. Bella, A. Chiarini, and B. Fehrman, "A Liouville theorem for stationary and ergodic ensembles of parabolic systems", *Probab. Theory Related Fields* (online publication March 2018). arXiv 1706.03440

[Bensoussan et al. 1978] A. Bensoussan, J.-L. Lions, and G. Papanicolaou, *Asymptotic analysis for periodic structures*, Studies in Mathematics and its Applications **5**, North-Holland, Amsterdam, 1978. MR Zbl

[Brezis and Ekeland 1976a] H. Brezis and I. Ekeland, "Un principe variationnel associé à certaines équations paraboliques: le cas indépendant du temps", *C. R. Acad. Sci. Paris Sér. A-B* **282**:17 (1976), A971–A974. MR Zbl

[Brezis and Ekeland 1976b] H. Brezis and I. Ekeland, "Un principe variationnel associé à certaines équations paraboliques: le cas dépendant du temps", *C. R. Acad. Sci. Paris Sér. A-B* **282**:20 (1976), A1197–A1198. MR Zbl

[Ekeland and Temam 1976] I. Ekeland and R. Temam, *Convex analysis and variational problems*, Studies in Mathematics and its Applications **1**, North-Holland, Amsterdam, 1976. MR Zbl

[Evans 2010] L. C. Evans, *Partial differential equations*, 2nd ed., Graduate Studies in Mathematics **19**, American Mathematical Society, Providence, RI, 2010. MR Zbl

[Fitzpatrick 1988] S. Fitzpatrick, "Representing monotone operators by convex functions", pp. 59–65 in *Workshop/Miniconference on Functional Analysis and Optimization* (Canberra, 1988), edited by S. P. Fitzpatrick and J. R. Giles, Proc. Centre Math. Anal. Austral. Nat. Univ. **20**, Austral. Nat. Univ., Canberra, 1988. MR Zbl

[Ghoussoub 2009] N. Ghoussoub, *Self-dual partial differential systems and their variational principles*, Springer, 2009. MR Zbl

[Ghoussoub and Tzou 2004] N. Ghoussoub and L. Tzou, "A variational principle for gradient flows", *Math. Ann.* **330**:3 (2004), 519–549. MR Zbl

[Giaquinta and Modica 1979] M. Giaquinta and G. Modica, "Regularity results for some classes of higher order nonlinear elliptic systems", *J. Reine Angew. Math.* **311/312** (1979), 145–169. MR Zbl

[Giaquinta and Struwe 1982] M. Giaquinta and M. Struwe, "On the partial regularity of weak solutions of nonlinear parabolic systems", *Math. Z.* **179**:4 (1982), 437–451. MR Zbl

[Gloria and Otto 2011] A. Gloria and F. Otto, "An optimal variance estimate in stochastic homogenization of discrete elliptic equations", *Ann. Probab.* **39**:3 (2011), 779–856. MR Zbl

[Gloria and Otto 2012] A. Gloria and F. Otto, "An optimal error estimate in stochastic homogenization of discrete elliptic equations", *Ann. Appl. Probab.* **22**:1 (2012), 1–28. MR Zbl

[Gloria and Otto 2015] A. Gloria and F. Otto, "The corrector in stochastic homogenization: near-optimal rates with optimal stochastic integrability", preprint, 2015. arXiv 1510.08290v2

[Gloria and Otto 2016] A. Gloria and F. Otto, "The corrector in stochastic homogenization: optimal rates, stochastic integrability, and fluctuations", preprint, 2016. arXiv 1510.08290v3

[Gloria et al. 2015] A. Gloria, S. Neukamm, and F. Otto, "Quantification of ergodicity in stochastic homogenization: optimal bounds via spectral gap on Glauber dynamics", *Invent. Math.* **199**:2 (2015), 455–515. MR Zbl

[Jikov et al. 1994] V. V. Jikov, S. M. Kozlov, and O. A. Oleĭnik, *Homogenization of differential operators and integral functionals*, Springer, 1994. MR Zbl

[Krylov 1982] N. V. Krylov, "Some properties of monotone mappings", *Litovsk. Mat. Sb.* **22**:2 (1982), 80–87. In Russian; translated in *Lith. Math. J.* **22**:2 (1982), 140–145. MR Zbl

[Nualart 2006] D. Nualart, *The Malliavin calculus and related topics*, 2nd ed., Springer, 2006. MR Zbl

[Papanicolaou and Varadhan 1981] G. C. Papanicolaou and S. R. S. Varadhan, "Boundary value problems with rapidly oscillating random coefficients", pp. 835–873 in *Random fields* (Esztergom, 1979), edited by J. Fritz et al., Colloq. Math. Soc. János Bolyai **27**, North-Holland, Amsterdam, 1981. MR Zbl

[Parviainen 2009] M. Parviainen, "Global gradient estimates for degenerate parabolic equations in nonsmooth domains", *Ann. Mat. Pura Appl.* (4) **188**:2 (2009), 333–358. MR Zbl

[Temam 1979] R. Temam, *Navier–Stokes equations: theory and numerical analysis*, Studies in Mathematics and its Applications **2**, North-Holland, Amsterdam, 1979. MR Zbl

[Widder 1961] D. V. Widder, "Series expansions of solutions of the heat equation in n dimensions", *Ann. Mat. Pura Appl.* (4) **55** (1961), 389–409. MR Zbl

Received 22 May 2017. Revised 15 Feb 2018. Accepted 9 Apr 2018.

SCOTT ARMSTRONG: scotta@cims.nyu.edu
Courant Institute of Mathematical Sciences, New York University, New York, NY, United States

ALEXANDRE BORDAS: alexandre.bordas@ens-lyon.fr
Ecole normale supérieure de Lyon, Lyon, France

JEAN-CHRISTOPHE MOURRAT: jean-christophe.mourrat@ens-lyon.fr
Ecole normale supérieure de Lyon, CNRS, Lyon, France

ANALYSIS AND PDE
Vol. 11, No. 8, 2018
dx.doi.org/10.2140/apde.2018.11.2015

HOPF POTENTIALS FOR THE SCHRÖDINGER OPERATOR

LUIGI ORSINA AND AUGUSTO C. PONCE

We establish the Hopf boundary point lemma for the Schrödinger operator $-\Delta+V$ involving potentials V that merely belong to the space $L^1_{\mathrm{loc}}(\Omega)$. More precisely, we prove that among all nonnegative supersolutions u of $-\Delta+V$ which vanish on the boundary $\partial\Omega$ and are such that $Vu\in L^1(\Omega)$, if there exists *one* supersolution that satisfies $\partial u/\partial n<0$ almost everywhere on $\partial\Omega$ with respect to the outward unit vector n, then such a property holds for *every* nontrivial supersolution in the same class. We rely on the existence of nontrivial solutions of the nonhomogeneous Dirichlet problem with boundary datum in $L^\infty(\partial\Omega)$.

1. Introduction and main results

Let $\Omega\subset\mathbb{R}^N$ be a smooth bounded connected open set. The Hopf boundary point lemma for elliptic PDEs asserts that if $u\in C^2(\Omega)\cap C^1(\overline{\Omega})$ satisfies the Dirichlet problem

$$\begin{cases} -\Delta u=\mu & \text{in } \Omega,\\ \quad u=0 & \text{on } \partial\Omega,\end{cases} \tag{1-1}$$

and if $\mu\geq 0$ in Ω, then the normal derivative of u with respect to the outward unit vector n satisfies

$$\frac{\partial u}{\partial n}<0 \quad \text{on } \partial\Omega;$$

see [Evans 2010, Section 6.4.2; Gilbarg and Trudinger 1998, Lemma 3.4; Dupaigne 2011, Proposition A.4.1]. The classical weak maximum principle states that $u\geq 0$ in Ω, so the information that $\partial u/\partial n\leq 0$ merely follows from the minimality of u on $\partial\Omega$. The main issue involving the Hopf lemma is that if $\partial u(a)/\partial n=0$ for some $a\in\partial\Omega$, then $u\equiv 0$ in Ω. A drawback of this formulation lies in the C^1 regularity of u that is required near the boundary.

The Hopf lemma can be also stated quantitatively, based on the Morel–Oswald maximum principle as follows:

$$u(x)\geq c\left(\int_\Omega \mu\, d_{\partial\Omega}\right) d_{\partial\Omega}(x), \tag{1-2}$$

where $c>0$ and $d_{\partial\Omega}:\overline{\Omega}\to\mathbb{R}$ denotes the distance to the boundary; see [Brezis and Cabré 1998, Lemma 3.2; Dupaigne 2011, Proposition A.4.2]. This inequality is equivalent to the pointwise estimate for the Green's function [Zhao 1986]:

$$G(x,y)\geq c'\, d_{\partial\Omega}(x)\, d_{\partial\Omega}(y).$$

MSC2010: primary 35B05, 35B50; secondary 31B15, 31B35.
Keywords: Hopf lemma, boundary point lemma, Schrödinger operator, weak normal derivative.

By smoothness of the solution ζ of the Dirichlet problem (1-1) with constant density $\mu \equiv 1$, we have $d_{\partial\Omega} \geq a\zeta$ for some constant $a > 0$. Inserting this estimate in (1-2) and integrating by parts, we can thus rewrite inequality (1-2) as

$$u(x) \geq c'' \left(\int_\Omega u \right) d_{\partial\Omega}(x). \tag{1-3}$$

Since (1-2) and (1-3) do not explicitly involve the normal derivative, by an approximation argument both inequalities also apply to the solutions of the Dirichlet problem involving rougher data μ, for instance in the class of L^1 functions or finite measures.

Another effective approach to avoid smoothness of the solution near the boundary consists in associating a notion of distributional normal derivative when μ is merely a finite measure in Ω, and the equation is satisfied in the sense of distributions:

$$-\int_\Omega u \, \Delta\varphi = \int_\Omega \varphi \, \mathrm{d}\mu \tag{1-4}$$

for every test function $\varphi \in C_c^\infty(\Omega)$ with compact support in Ω. The zero boundary datum can be encoded by assuming that $u \in W_0^{1,1}(\Omega)$. An equivalent strategy to give a meaning to the Dirichlet problem consists in using the integral formulation (1-4) with the larger class $C_0^\infty(\overline{\Omega})$ of test functions which are smooth in $\overline{\Omega}$ and vanish on $\partial\Omega$; see [Littman et al. 1963; Ponce 2016, Proposition 6.3].

It has been observed by Brezis and the second author in [Brezis and Ponce 2008, Theorem 1.2] that if $u \in W_0^{1,1}(\Omega)$ satisfies (1-4) for some finite measure μ in Ω, then there exists a unique function $F \in L^1(\partial\Omega)$ such that

$$\int_\Omega \nabla u \cdot \nabla\psi = \int_\Omega \psi \, \mathrm{d}\mu + \int_{\partial\Omega} \psi F \, \mathrm{d}\sigma \tag{1-5}$$

for every test function $\psi \in C^\infty(\overline{\Omega})$, where $\sigma = \mathcal{H}^{N-1}\lfloor_{\partial\Omega}$ denotes the surface measure of $\partial\Omega$. The distributional normal derivative of u is then defined as

$$\frac{\partial u}{\partial n} := F.$$

When u is smooth on $\overline{\Omega}$, the notions of classical and distributional normal derivatives of u coincide. The definition of a distributional pairing $\langle \partial u/\partial n, \psi \rangle$ under various assumptions on u and ∇u has been investigated by several authors; see, e.g., [Lions and Magenes 1972; Kohn and Temam 1983; Anzellotti 1983]. The main point here is its realization as a legitimate function in $L^1(\partial\Omega)$, like in [Anzellotti 1984]. We refer the reader to [Ancona 2009; Brezis and Ponce 2008] for additional properties of the distributional normal derivative, including the setting of nonhomogeneous Dirichlet problems.

Identity (1-5) implies in particular that the function u, extended by zero to $\mathbb{R}^N$, is such that Δu is a finite measure in $\mathbb{R}^N$. The distributional normal derivative satisfies a comparison estimate (see Proposition 2.1 below) which combined with (1-3) provides one with the uniform bound

$$\frac{\partial u}{\partial n}(x) \leq -c'' \int_\Omega u. \tag{1-6}$$

One of the motivations of our work comes from the seminal paper [Kato 1972] on the Schrödinger operator $-\Delta + V$ involving potentials V which are merely $L^1_{\mathrm{loc}}(\Omega)$. The Hopf lemma above has an affirmative counterpart for potentials $V \in L^\infty(\Omega)$, but we are interested in situations where V need not be summable near the boundary.

Many classical properties that hold for the Laplacian need no longer be true for $-\Delta + V$ due to some possible singular behavior of V. In this regard, two instructive examples are provided by the smooth functions $u_i : \bar{B}_1 \to \mathbb{R}$ defined by

$$u_1(x) = (1 - |x|^2)^2 \quad \text{and} \quad u_2(x) = (1 - |x|^2)|x - a|,$$

where $B_1 := B_1(0)$ denotes the unit ball in $\mathbb{R}^N$ centered at the origin and $a \in \partial B_1$ is any given point on the boundary. In the first case, we have $\partial u_1/\partial n \equiv 0$ on ∂B_1 and the Schrödinger equation

$$-\Delta u_1 + V u_1 = 0 \quad \text{in } B_1$$

is satisfied in terms of a potential V that behaves like $1/d^2_{\partial B_1}$ near the boundary; in particular, $V \notin L^1(B_1; d_{\partial B_1}\, \mathrm{d}x)$. In the second case, we have $\partial u_2/\partial n < 0$ except at a and the Schrödinger equation is now satisfied for another potential V such that $V \in L^1(B_1; d_{\partial B_1}\, \mathrm{d}x)$ in dimension $N \geq 2$.

We thus have the appearance of an exceptional set where the Hopf lemma fails, and it is our goal in this paper to understand how big such an exceptional set can be. A more refined example which we develop in Section 7 below shows that *every* compact subset $K \subset \partial\Omega$ with zero surface measure is an exceptional set for some suitable potential $V \in L^1(\Omega; d_{\partial\Omega}\, \mathrm{d}x)$. It follows from our Theorem 1 below that there can be essentially no other exceptional sets in this case.

The class of functions we consider consists of supersolutions $u \in W^{1,1}_0(\Omega)$ of the Schrödinger operator $-\Delta + V$ in the sense of distributions. More precisely, we assume that $Vu \in L^1_{\mathrm{loc}}(\Omega)$ and

$$\int_\Omega u\,(-\Delta\varphi + V\varphi) \geq 0$$

for every nonnegative function $\varphi \in C^\infty_c(\Omega)$. By a classical property in the theory of distributions, we have in this case that $-\Delta u + Vu$ is a locally finite measure in Ω. However, the measure Δu need not be finite in Ω and so the distributional normal derivative may not be well-defined in $L^1(\partial\Omega)$ in the sense of [Brezis and Ponce 2008]. For this reason, we define in this paper the normal derivative for functions u that are merely in $W^{1,1}_0(\Omega)$: by $\partial u/\partial n$ we mean the essential infimum of the set

$$\left\{ \frac{\partial w}{\partial n} \in L^1(\partial\Omega) : w \in \mathcal{G}_u \right\},$$

where $\mathcal{G}_u$ denotes the class of functions $w \in W^{1,1}_0(\Omega)$ such that Δw is a finite measure in Ω and $w \leq u$ almost everywhere in Ω; see Section 2 below. In particular, if u is nonnegative, then the normal derivative $\partial u/\partial n$ is a Borel function with values in $[-\infty, 0]$, and if we happen to know that Δu is a finite measure in Ω, then $u \in \mathcal{G}_u$ and $\partial u/\partial n$ coincides with the distributional normal derivative.

To understand the mechanism that is hidden behind the examples above concerning the failure of the Hopf lemma, we introduce the concept of Hopf potential as follows:

Definition 1.1. We say that $V \in L^1_{\mathrm{loc}}(\Omega)$ is a *Hopf potential* whenever there exists a nonnegative function $\zeta_0 \in W^{1,1}_0(\Omega)$ such that

(H_1) $V\zeta_0 \in L^1(\Omega)$,

(H_2) $\partial\zeta_0/\partial n < 0$ almost everywhere on $\partial\Omega$.

As a trivial consequence of this definition, for every Hopf potential V and every $\alpha \in \mathbb{R}$, the function αV is also a Hopf potential. We show in Section 2 that the class of Hopf potentials is actually a vector subspace of $L^1_{\mathrm{loc}}(\Omega)$. Since the solution ζ of the Dirichlet problem (1-1) with constant density $\mu \equiv 1$ behaves as $d_{\partial\Omega}$ near the boundary by the classical Hopf lemma, we have $V\zeta \in L^1(\Omega)$ if and only if $V \in L^1(\Omega; d_{\partial\Omega}\,\mathrm{d}x)$. Therefore, *every* $V \in L^1(\Omega; d_{\partial\Omega}\,\mathrm{d}x)$ *is a Hopf potential*.

We establish the following qualitative counterpart of estimate (1-6) for $-\Delta + V$ when V is a Hopf potential:

Theorem 1. *Let $V \in L^1_{\mathrm{loc}}(\Omega)$ be a Hopf potential and let $u \in W^{1,1}_0(\Omega)$ be a nonnegative supersolution of the Schrödinger operator $-\Delta + V$. If $Vu \in L^1(\Omega)$ and $\int_\Omega u > 0$, then*

$$\frac{\partial u}{\partial n} < 0 \quad \textit{almost everywhere on } \partial\Omega.$$

Theorem 1 above contains as a particular case a Hopf lemma by Bertsch, Smarrazzo and Tesei [Bertsch et al. 2015, Proposition 3.4] which implies the main result in their paper (Theorem 2.1) concerning a characterization of the strong maximum principle in dimension $N = 1$; see also [Bertsch and Rostamian 1985, Lemma 3.6]. To tackle the Hopf lemma in any dimension $N \geq 1$, we rely on a different strategy based on a careful combination of fine properties from measure theory and elliptic PDEs.

One may also consider a localized counterpart of the concept of Hopf potentials, where property (H_2) need not be satisfied by ζ_0 on the entire boundary, but only on a subset of it. In fact, we deduce Theorem 1 from a more general result which is valid for potentials V that merely belong to $L^1_{\mathrm{loc}}(\Omega)$:

Theorem 2. *Let $V \in L^1_{\mathrm{loc}}(\Omega)$ and let $u_i \in W^{1,1}_0(\Omega)$, with $i \in \{1, 2\}$, be two nonnegative supersolutions of the Schrödinger operator $-\Delta + V$. If $Vu_i \in L^1(\Omega)$ and $\int_\Omega u_i > 0$, then for almost every $x \in \partial\Omega$ we have*

$$\frac{\partial u_1}{\partial n}(x) < 0 \quad \textit{if and only if} \quad \frac{\partial u_2}{\partial n}(x) < 0.$$

This theorem yields the remarkable property that once there exists *one* supersolution for $-\Delta + V$ satisfying the conclusion of the classical Hopf lemma on a subset $A \subset \partial\Omega$, then *every* supersolution also satisfies the Hopf lemma on A except for a negligible subset of $\partial\Omega$. Such a conclusion bears some striking analogy with the (straightforward) generalized weak maximum principle for linear elliptic operators of second order [Protter and Weinberger 1984, Chapter 2, Theorem 10]: for the Schrödinger operator $-\Delta + V$ with a possibly signed potential V, the existence of one positive supersolution implies that every nonzero supersolution which vanishes on the boundary must be positive.

The existence of a positive supersolution is also equivalent to the positivity of the energy functional

$$\varphi \in C^\infty_c(\Omega) \longmapsto \int_\Omega (|\nabla\varphi|^2 + V\varphi^2),$$

which is at the heart of the Agmon–Allegretto–Piepenbrink positivity principle; see, e.g., [Dupaigne 2011, Theorem A.6.1] and also [Pinchover 2007; Devyver et al. 2014] for more detailed information and further perspectives. Although one typically assumes that $V \in L^p_{\mathrm{loc}}(\Omega)$ with $p > N/2$, the validity of the strong maximum principle when $V \in L^p_{\mathrm{loc}}(\Omega)$ with $p \geq 1$, see [Ancona 1979; Orsina and Ponce 2016], supports an extension of the Agmon–Allegretto–Piepenbrink principle for potentials V that merely belong to $L^1_{\mathrm{loc}}(\Omega)$ based on the tools we develop to prove Theorems 1 and 2 above; see [Pinchover and Tintarev 2005; Bandle et al. 2008] for Hardy potentials and [Pinchover and Psaradakis 2016] for potentials in Morrey spaces.

A key ingredient of our analysis relies on Proposition 5.1 below, which establishes the equivalence between the validity of the Hopf lemma for the Schrödinger operator $-\Delta + V$ and the existence of nontrivial solutions of the nonhomogeneous Dirichlet problem

$$\begin{cases} -\Delta w + V w = 0 & \text{in } \Omega, \\ w = g & \text{on } \partial\Omega, \end{cases} \tag{1-7}$$

with nonnegative potentials $V \in L^1_{\mathrm{loc}}(\Omega)$, for any datum $g \in L^\infty(\partial\Omega)$. The meaning of a solution of (1-7) is a delicate issue due to the possible singular behavior of V near the boundary. Our approach is based on the use of nonsmooth test functions that satisfy a Dirichlet problem involving interior measure data in the spirit of Stampacchia's definition of weak solutions [1965] via duality. That problem (1-7) has a solution in this sense for every $g \in L^\infty(\partial\Omega)$ can be handled using an approximation procedure starting from variational solutions; see Section 3.

Our strategy to tackle (1-7) differs from the recent work of Véron and Yarur [2012] that investigates problem (1-7) with finite boundary measure data and nonnegative potentials $V \in L^\infty_{\mathrm{loc}}(\Omega)$. They rely on the definition of a solution using test functions like $C^\infty_0(\overline{\Omega})$, which do not take into account the singular behavior of the potential V, and on the Poisson representation of the solution in terms of the Poisson kernel associated to $-\Delta + V$.

Due to the singular behavior of V, it may happen in our case that $w \equiv 0$ is the (unique) solution of (1-7) even if $g \neq 0$. An example of such a counterintuitive phenomenon is given by any potential $V \sim 1/d^2_{\partial\Omega}$, for which the Hopf lemma fails completely. In this case, the equation

$$-\Delta w + V w = 0 \quad \text{in } \Omega$$

can have nontrivial solutions but they satisfy some normalized boundary trace that has been investigated by Marcus and Nguyen [2017].

Another strategy that has been pursued by Ancona [1987] is based on the existence of the Martin kernel K^V_a for $a \in \partial\Omega$ under the assumption that V is a potential in $L^\infty_{\mathrm{loc}}(\Omega)$ that satisfies

$$0 \leq V \lesssim \frac{1}{d^2_{\partial\Omega}}. \tag{1-8}$$

For instance, in the setting of positive solutions of the semilinear equation

$$-\Delta u + u^q = 0 \quad \text{in } \Omega$$

with exponent $q > 1$, the potential $V = u^{q-1}$ satisfies (1-8) by the Keller–Osserman estimate; see [Marcus and Véron 2014, Chapter 4]. In general, the study of fine regular points of the Schrödinger operator $-\Delta + V$ through Martin kernels gives another approach to the existence of solutions of (1-7). In this regard, Ancona [2012], see also [Véron and Yarur 2012], proved that $a \in \partial\Omega$ is a fine regular point for $-\Delta + V$ if and only if

$$\int_\Omega \frac{d_{\partial\Omega}^2(x)}{|x-a|^N} V(x)\, \mathrm{d}x < +\infty. \tag{1-9}$$

When, in addition to (1-8), V belongs to $L^1(\Omega; d_{\partial\Omega}\, \mathrm{d}x)$, integration of the left-hand side of (1-9) over $\partial\Omega$ with respect to a and Fubini's theorem imply that almost every $a \in \partial\Omega$ satisfies (1-9). This agrees with our conclusion concerning the existence of nontrivial solutions for (1-7) since we know in this case that V is a Hopf potential. It is unclear however how one can avoid assumption (1-8) in this setting: Ancona's argument strongly relies on the Harnack principle, which is not true when one merely has $V \in L^1_{\mathrm{loc}}(\Omega)$.

Observe that from the physical point of view the infinite-potential well $1/d_{\partial\Omega}^2$ is so strong that it confines particles inside Ω, which mathematically means that supersolutions must have a vanishing normal derivative on $\partial\Omega$; see [Díaz 2015; 2017; Díaz et al. 2018] and also Example 8.2 below. Although such a conclusion can be successfully deduced from Theorem 2 by looking explicitly for one supersolution such that $\partial u/\partial n \equiv 0$ on $\partial\Omega$, we give a direct proof of this fact by a simple measure-theoretic argument that does not rely on the PDE; see Proposition 2.7.

The paper is organized as follows. In Section 2 we extend the concept of normal derivative to any function in $W_0^{1,1}(\Omega)$, even if Δu is not a finite measure in Ω. In Section 3, we prove the existence of solutions of the nonhomogeneous Dirichlet problem with L^∞ data; the meaning of solution is given by means of duality with solutions of the Dirichlet problem with measure data. In Section 4, we prove the existence of nonnegative solutions of the nonhomogeneous problem when the boundary datum is nonnegative but the inner datum is nonpositive. We then explain how this property implies Theorem 1 in the case of smooth supersolutions. In Section 5 we explain the connection between the Hopf lemma and the existence of nontrivial solutions of (1-7). Theorems 1 and 2 are then proved in Section 6. We then show in Section 7 that every negligible compact subset of $\partial\Omega$ is the zero-set $\{\partial u/\partial n = 0\}$ for some smooth positive solution of the Schrödinger equation $-\Delta u + Vu = 0$ such that $V \in L^1(\Omega; d_{\partial\Omega}\, \mathrm{d}x)$. In Section 8 we explain why Theorems 1 and 2 cannot be true for potentials $V : \Omega \to [0, +\infty]$ that are merely Borel functions.

2. Normal derivative as a Borel function

The notion of distributional normal derivative from [Brezis and Ponce 2008] applies to any function $u \in W_0^{1,1}(\Omega)$ such that Δu is a finite measure in Ω. In this case, the normal derivative $\partial u/\partial n$ is an element in $L^1(\partial\Omega)$ such that

$$\int_\Omega \nabla u \cdot \nabla \psi = -\int_\Omega \psi\, \Delta u + \int_{\partial\Omega} \psi\, \frac{\partial u}{\partial n}\, \mathrm{d}\sigma \quad \text{for every } \psi \in C^\infty(\overline{\Omega}).$$

In this work, we deal with functions $u \in W_0^{1,1}(\Omega)$ such that the distribution Δu need not be a finite measure. The strategy we adopt to define a Borel normal derivative is motivated by the following comparison principle, which can be deduced from Kato's inequality; see [Ponce 2016, Lemma 12.15]:

Proposition 2.1. *Let $v \in W_0^{1,1}(\Omega)$ be such that Δv is a finite measure in Ω. If $v \geq 0$ almost everywhere in Ω, then $\partial v/\partial n \leq 0$ almost everywhere on $\partial\Omega$ with respect to the surface measure.*

Now, to our definition of normal derivative as a Borel function, we begin with any $u \in W_0^{1,1}(\Omega)$. By the essential infimum of the set

$$\mathcal{N}_u := \left\{ \frac{\partial w}{\partial n} \in L^1(\partial\Omega) : w \in \mathcal{G}_u \right\},$$

where

$$\mathcal{G}_u := \{w \in W_0^{1,1}(\Omega) : \Delta w \text{ is a finite measure and } w \leq u \text{ a.e. in } \Omega\},$$

we mean a Borel function $F : \partial\Omega \to [-\infty, \infty]$ such that

(i) $F \leq \partial w/\partial n$ almost everywhere on $\partial\Omega$, for every $w \in \mathcal{G}_u$,

(ii) if $\widetilde{F} : \partial\Omega \to [-\infty, \infty]$ is another Borel function that satisfies (i), then $\widetilde{F} \leq F$ almost everywhere on $\partial\Omega$.

We then define the normal derivative of u as

$$\frac{\partial u}{\partial n} := F.$$

Proposition 2.2. *Such a normal derivative $\partial u/\partial n$ exists for every $u \in W_0^{1,1}(\Omega)$.*

Proof. By the separability of $L^1(\partial\Omega)$, we can extract a countable subset A of $\mathcal{G}_u$ such that $\{\partial v/\partial n : v \in A\}$ is dense in $\mathcal{N}_u$. We claim that the Borel measurable function $F : \partial\Omega \to [-\infty, \infty]$ defined by

$$F(x) := \inf_{v \in A} \frac{\partial v}{\partial n}(x)$$

satisfies properties (i) and (ii) above. Indeed, given $w \in \mathcal{G}_u$, take a sequence $(v_k)_{k\in\mathbb{N}}$ in A such that $(\partial v_k/\partial n)_{k\in\mathbb{N}}$ converges to $\partial w/\partial n$ in $L^1(\partial\Omega)$. Passing to a subsequence if necessary, we may assume that the convergence holds almost everywhere on $\partial\Omega$. Since $\partial v_k/\partial n \geq F$ on $\partial\Omega$, we deduce that $\partial w/\partial n \geq F$ almost everywhere on $\partial\Omega$. Hence, F satisfies property (i). We now let $\widetilde{F}$ be another function that satisfies property (i), and for each $v \in A$ denote by $E_v \subset \partial\Omega$ a set of surface measure zero such that $\widetilde{F}(x) \leq \partial v(x)/\partial n$ for every $x \in \partial\Omega \setminus E_v$. Since A is countable, the set $E = \bigcup_{v\in A} E_v$ also has surface measure zero and

$$\widetilde{F}(x) \leq \frac{\partial v}{\partial n}(x) \quad \text{for every } x \in \partial\Omega \setminus E,$$

for every $v \in A$. Taking the infimum of the right-hand side over v we deduce that $\widetilde{F} \leq F$ on $\partial\Omega \setminus E$, which gives property (ii). □

In the pointwise approximation of a Borel normal derivative, one can restrict the attention to the study of monotone sequences in $\mathcal{G}_u$ and $\mathcal{N}_u$:

Proposition 2.3. *For every $u \in W_0^{1,1}(\Omega)$, there exists a nondecreasing sequence $(w_k)_{k\in\mathbb{N}}$ in $\mathcal{G}_u$ such that $(\partial w_k/\partial n)_{k\in\mathbb{N}}$ is a nonincreasing sequence in $\mathcal{N}_u$ that converges almost everywhere to $\partial u/\partial n$ on $\partial\Omega$.*

In order to prove Proposition 2.3 we rely on Kato's inequality up to the boundary, which implies that if $\zeta \in W_0^{1,1}(\Omega)$ and $\Delta\zeta$ is a finite measure in Ω, then $\Delta[(\zeta - a)^+]$ is also a finite measure in Ω for every $a \in \mathbb{R}$ and

$$\|\Delta[(\zeta - a)^+]\|_{\mathcal{M}(\Omega)} \le 2\|\Delta\zeta\|_{\mathcal{M}(\Omega)}; \tag{2-1}$$

see [Brezis and Ponce 2008, Theorem 1.1; Ponce 2016, Proposition 7.7]. Here, $\mathcal{M}(\Omega)$ denotes the vector space of finite Borel measures ν in Ω equipped with the norm

$$\|\nu\|_{\mathcal{M}(\Omega)} = |\nu|(\Omega),$$

which makes $\mathcal{M}(\Omega)$ a Banach space. The normal derivative $\partial(\zeta - a)^+/\partial n$ is then well-defined in the distributional sense as an element in $L^1(\partial\Omega)$. Ancona [2009, Remark 6.2] subsequently proved using tools from potential theory that

$$\frac{\partial(\zeta - a)^+}{\partial n} = \begin{cases} \partial\zeta/\partial n & \text{if } a < 0, \\ \min\{\partial\zeta/\partial n, 0\} & \text{if } a = 0, \\ 0 & \text{if } a > 0. \end{cases} \tag{2-2}$$

These properties can be illustrated by the following lemma:

Lemma 2.4. *If $\zeta_i \in W_0^{1,1}(\Omega)$, with $i \in \{1, 2\}$, are such that $\Delta\zeta_i$ are finite measures in Ω, then the function $\zeta = \max\{\zeta_1, \zeta_2\}$ belongs to $W_0^{1,1}(\Omega)$, is such that $\Delta\zeta$ is a finite measure in Ω, and*

$$\frac{\partial\zeta}{\partial n} = \min\left\{\frac{\partial\zeta_1}{\partial n}, \frac{\partial\zeta_2}{\partial n}\right\} \quad \textit{almost everywhere on } \partial\Omega.$$

Proof of Lemma 2.4. Observe that

$$\zeta = \zeta_2 + (\zeta_1 - \zeta_2)^+.$$

Thus, $\zeta \in W_0^{1,1}(\Omega)$. By Kato's inequality up to the boundary (2-1) applied to the function $\zeta_1 - \zeta_2$ and $a = 0$, we deduce that the measure $\Delta[(\zeta_1 - \zeta_2)^+]$ is finite in Ω, whence so is the measure $\Delta\zeta$. By (2-2) we have

$$\frac{\partial}{\partial n}(\zeta_1 - \zeta_2)^+ = \min\left\{\frac{\partial}{\partial n}(\zeta_1 - \zeta_2), 0\right\} = -\frac{\partial\zeta_2}{\partial n} + \min\left\{\frac{\partial\zeta_1}{\partial n}, \frac{\partial\zeta_2}{\partial n}\right\},$$

and the conclusion follows. □

Proof of Proposition 2.3. Let $(v_k)_{k\in\mathbb{N}}$ be a sequence in $\mathcal{G}_u$ such that $(\partial v_k/\partial n)_{k\in\mathbb{N}}$ is dense in $\mathcal{N}_u$. As in the proof of Proposition 2.2, we have

$$\frac{\partial u}{\partial n} = \inf_{j\in\mathbb{N}} \frac{\partial v_j}{\partial n} \quad \text{almost everywhere on } \partial\Omega.$$

Define by induction the nondecreasing sequence $(w_k)_{k\in\mathbb{N}}$ as $w_0 := v_0$ and, for $k \in \mathbb{N}_*$,

$$w_k := \max\{w_{k-1}, v_k\}.$$

By Lemma 2.4 we have $w_k \in \mathcal{G}_u$ for every $k \in \mathbb{N}$. In particular, $\partial u/\partial n \le \partial w_k/\partial n$ almost everywhere on $\partial\Omega$. By comparison of normal derivatives, the sequence $(\partial w_k/\partial n)_{k\in\mathbb{N}}$ is monotone and nonincreasing; hence

$$\frac{\partial u}{\partial n} \le \lim_{k\to\infty} \frac{\partial w_k}{\partial n}$$

and also

$$\lim_{k\to\infty} \frac{\partial w_k}{\partial n} \le \frac{\partial w_j}{\partial n} \le \frac{\partial v_j}{\partial n}$$

almost everywhere on $\partial\Omega$ for every $j \in \mathbb{N}$. Taking the infimum of the right-hand side over j, we deduce that

$$\lim_{k\to\infty} \frac{\partial w_k}{\partial n} = \frac{\partial u}{\partial n} \quad \text{almost everywhere on } \partial\Omega. \qquad \square$$

As a consequence of Proposition 2.3, we observe that for the sake of investigating the set where the normal derivative of a function u is negative, one does not need to rely on the entire family $\mathcal{G}_u$ nor even on a countable subset of it, but on a single suitably chosen element:

Proposition 2.5. *For every nonnegative function $u \in W_0^{1,1}(\Omega)$, there exists a nonnegative function $v \in \mathcal{G}_u$ such that*

$$\frac{\partial v}{\partial n} < 0 \quad \textit{almost everywhere on } \left\{\frac{\partial u}{\partial n} < 0\right\}.$$

Proof. Let $(w_k)_{k\in\mathbb{N}}$ be a nondecreasing sequence in $\mathcal{G}_u$ satisfying the conclusion of Proposition 2.3. Replacing each w_k by its positive part if necessary, we may assume by Lemma 2.4 that each function w_k is nonnegative in Ω and in particular $\partial w_0/\partial n$ is nonpositive on $\partial\Omega$. Hence, in the sum

$$\frac{\partial w_0}{\partial n} + \sum_{j=1}^{\infty} \left(\frac{\partial w_j}{\partial n} - \frac{\partial w_{j-1}}{\partial n}\right) = \frac{\partial u}{\partial n},$$

we have

$$\frac{\partial w_0}{\partial n} \le 0 \quad \text{and} \quad \frac{\partial w_j}{\partial n} - \frac{\partial w_{j-1}}{\partial n} \le 0$$

almost everywhere on $\partial\Omega$ for every $j \in \mathbb{N}_*$. In addition, for almost every $x \in \{\partial u/\partial n < 0\}$, one of these terms, possibly depending on x, must be negative. The conclusion is thus satisfied with

$$v := w_0 + \sum_{j=1}^{\infty} \epsilon_j (w_j - w_{j-1}),$$

where $(\epsilon_j)_{j\in\mathbb{N}}$ is a sequence in $(0, 1]$ such that

$$\sum_{j=1}^{\infty} \epsilon_j \big(\|\nabla(w_j - w_{j-1})\|_{L^1(\Omega)} + \|\Delta(w_j - w_{j-1})\|_{\mathcal{M}(\Omega)}\big) < +\infty.$$

Indeed, we have $v \le u$ in Ω and such a choice of sequence $(\epsilon_j)_{j\in\mathbb{N}}$ ensures that v belongs to $W_0^{1,1}(\Omega)$ and Δv is a finite measure in Ω by completeness of $W_0^{1,1}(\Omega)$ and $\mathcal{M}(\Omega)$. $\square$

Corollary 2.6. *The class of Hopf potentials is a vector subspace of $L^1_{\mathrm{loc}}(\Omega)$.*

Proof. Let $V_i \in L^1_{\mathrm{loc}}(\Omega)$, with $i \in \{1,2\}$, be two Hopf potentials, and denote by $\zeta_i \in W^{1,1}_0(\Omega)$ a nonnegative function that satisfies properties (H$_1$) and (H$_2$) with respect to V_i. We now verify that $\alpha_1 V_1 + \alpha_2 V_2$ is a Hopf potential for every $\alpha_1, \alpha_2 \in \mathbb{R}$ by using the function $\zeta := \min\{\zeta_1, \zeta_2\}$. Observe that

$$\big|(\alpha_1 V_1 + \alpha_2 V_2)\zeta\big| \le |\alpha_1 V_1|\zeta_1 + |\alpha_2 V_2|\zeta_2 \in L^1(\Omega).$$

By Proposition 2.5 there exists a nonnegative function $v_i \in \mathcal{G}_{\zeta_i}$ such that $\partial v_i/\partial n < 0$ almost everywhere on $\partial\Omega$. Since $\zeta \ge \min\{v_1, v_2\}$, the counterpart of Lemma 2.4 for the minimum of two functions gives in this case $\min\{v_1, v_2\} \in \mathcal{G}_\zeta$ and

$$\frac{\partial \zeta}{\partial n} \le \frac{\partial}{\partial n} \min\{v_1, v_2\} = \max\left\{\frac{\partial v_1}{\partial n}, \frac{\partial v_2}{\partial n}\right\}.$$

Therefore, $\partial\zeta/\partial n < 0$ almost everywhere on $\partial\Omega$. □

The growth of the potential V like $1/d^2_{\partial\Omega}$ near the boundary is critical for the validity of the Hopf lemma:

Proposition 2.7. *If $u \in W^{1,1}_0(\Omega)$ is a nonnegative function such that*

$$\int_\Omega \frac{u}{d^2_{\partial\Omega}} < \infty, \tag{2-3}$$

then $\partial u/\partial n = 0$ almost everywhere on $\partial\Omega$.

Proof. By the comparison principle for normal derivatives (Proposition 2.1), it suffices to verify that for every nonnegative function $v \in \mathcal{G}_u$ we have $\partial v/\partial n = 0$ almost everywhere on $\partial\Omega$. To this end, denote by μ the measure in $\mathbb{R}^N$ such that $\mu = \Delta v$ in Ω and $\mu \equiv 0$ on the Borel subsets of $\mathbb{R}^N \setminus \Omega$; we also extend v by zero to $\mathbb{R}^N \setminus \Omega$. For every $\psi \in C^\infty(\mathbb{R}^N)$, we then have

$$\int_{\partial\Omega} \psi \frac{\partial v}{\partial n}\, \mathrm{d}\sigma = \int_\Omega \psi\, \Delta v + \int_\Omega \nabla v \cdot \nabla\psi = \int_{\mathbb{R}^N} \psi\, \mathrm{d}\mu - \int_{\mathbb{R}^N} v\, \Delta\psi.$$

Given $x \in \partial\Omega$ and $r > 0$, we apply this identity with $\psi(y) = \varphi((y-x)/r)$, where $\varphi \in C^\infty_c(\mathbb{R}^N)$ is such that $\varphi \equiv 1$ in B_1, $\varphi \equiv 0$ in $\mathbb{R}^N \setminus B_2$, and $0 \le \varphi \le 1$ in $\mathbb{R}^N$. By the nonpositivity of $\partial v/\partial n$ we then get

$$\int_{\partial\Omega\cap B_r(x)} \left|\frac{\partial v}{\partial n}\right| \mathrm{d}\sigma \le -\int_{\partial\Omega} \psi \frac{\partial v}{\partial n}\, \mathrm{d}\sigma \le |\mu|(B_{2r}(x)) + C\int_{B_{2r}(x)} \frac{v}{d^2_{\partial\Omega}}. \tag{2-4}$$

The set

$$E_1 := \left\{x \in \mathbb{R}^N : \limsup_{r\to 0} \frac{1}{r^{N-1}} \int_{B_r(x)} \frac{v}{d^2_{\partial\Omega}} > 0\right\}$$

satisfies $\mathcal{H}^{N-1}(E_1) = 0$, where $\mathcal{H}^{N-1}$ is the Hausdorff measure of E_1 of dimension $N-1$; see, e.g., [Evans and Gariepy 2015, Theorem 2.10]. Since $\mu \equiv 0$ on $\partial\Omega$, by outer regularity of μ the set

$$E_2 := \left\{x \in \mathbb{R}^N : \limsup_{r\to 0} \frac{|\mu|(B_r(x))}{r^{N-1}} > 0\right\}$$

also satisfies $\mathcal{H}^{N-1}(E_2) = 0$, with the same proof as for E_1. Dividing both sides of (2-4) by r^{N-1}, it follows that for every $x \in \partial\Omega \setminus (E_1 \cup E_2)$ we have

$$\lim_{r\to 0} \frac{1}{r^{N-1}} \int_{\partial\Omega\cap B_r(x)} \left|\frac{\partial v}{\partial n}\right| \, \mathrm{d}\sigma = 0,$$

and then $\partial v/\partial n = 0$ almost everywhere on $\partial\Omega$ as claimed. □

The choice of the Sobolev space $W_0^{1,1}(\Omega)$ to define the normal derivative is sufficient for our purposes, but one might be interested in a condition that does not require the weak (distributional) derivative to be in $L^1(\Omega;\mathbb{R}^N)$. In fact, the presentation above easily adapts to functions $u \in L^1(\Omega)$ which vanish on the boundary in the sense that

$$\lim_{r\to 0} \frac{1}{r} \int_{\{x\in\Omega : d_{\partial\Omega}(x)<r\}} |u| = 0. \tag{2-5}$$

The reason is that any function $v \in L^1(\Omega)$ such that (2-5) holds and Δv is a finite measure in Ω necessarily belongs to $W_0^{1,1}(\Omega)$; see [Ponce 2016, Propositions 6.3 and 20.2]. Therefore, a family $\widetilde{\mathcal{G}}_u$ defined in terms of (2-5) coincides with our class $\mathcal{G}_u$. An interesting aspect of (2-5) is that such a condition automatically holds for any function $u \in L^1(\Omega)$ that satisfies (2-3).

3. The Dirichlet problem with nonhomogeneous data

Given $f \in L^\infty(\Omega)$ and $g \in L^\infty(\partial\Omega)$, the concept of solution of the nonhomogeneous Dirichlet problem for the Schrödinger operator $-\Delta + V$ with $V \in L^1_{\mathrm{loc}}(\Omega)$,

$$\begin{cases} -\Delta v + Vv = f & \text{in } \Omega, \\ \qquad\quad v = g & \text{on } \partial\Omega, \end{cases} \tag{3-1}$$

can be straightforwardly defined by L^1-L^∞ duality using as test function the solution of the Dirichlet problem

$$\begin{cases} -\Delta\zeta + V\zeta = \mu & \text{in } \Omega, \\ \qquad\quad \zeta = 0 & \text{on } \partial\Omega \end{cases} \tag{3-2}$$

involving $\mu \in L^1(\Omega)$, in the spirit of Stampacchia's definition of weak solutions [1965]. In this case, a solution $v \in L^\infty(\Omega)$ of (3-1) is meant to satisfy the identity

$$\int_\Omega v\mu = \int_\Omega f\zeta - \int_{\partial\Omega} g\frac{\partial\zeta}{\partial n} \, \mathrm{d}\sigma \quad \text{for every } \mu \in L^1(\Omega). \tag{3-3}$$

While this notion is enough to investigate the Hopf lemma involving smooth supersolutions of $-\Delta + V$, to deal with nonsmooth ones we rely on a larger class of test functions. Namely, we allow any solution of (3-2) involving finite measures μ in Ω which are diffuse with respect to the $W^{1,2}$ capacity. The main result of this section ensures the existence of solutions of (3-1) in this stronger sense:

Proposition 3.1. *Let $V \in L^1_{\mathrm{loc}}(\Omega)$ be a nonnegative function. Given $f \in L^\infty(\Omega)$ and $g \in L^\infty(\partial\Omega)$, there exists $v \in W^{1,2}_{\mathrm{loc}}(\Omega) \cap L^\infty(\Omega)$ such that*

$$\int_\Omega \hat{v} \, \mathrm{d}\mu = \int_\Omega f\zeta - \int_{\partial\Omega} g\frac{\partial\zeta}{\partial n} \, \mathrm{d}\sigma \tag{3-4}$$

for every finite measure μ in Ω which is diffuse with respect to the $W^{1,2}$ capacity, where $\hat{v}$ is the precise representative of v and $\zeta \in W_0^{1,1}(\Omega)$ satisfies

$$-\Delta\zeta + V\zeta = \mu \quad \text{in the sense of distributions in } \Omega. \tag{3-5}$$

In particular, there exists a constant $C > 0$ such that

$$\|v\|_{L^\infty(\Omega)} \le C(\|f\|_{L^\infty(\Omega)} + \|g\|_{L^\infty(\partial\Omega)}). \tag{3-6}$$

We recall that the $W^{1,2}$ capacity of a compact subset $K \subset \mathbb{R}^N$ is defined as

$$\operatorname{cap}_{W^{1,2}}(K) = \inf\left\{\|\varphi\|^2_{W^{1,2}(\mathbb{R}^N)} : \varphi \in C_c^\infty(\mathbb{R}^N) \text{ is nonnegative and } \varphi > 1 \text{ on } K\right\}.$$

A Borel measure μ in Ω is diffuse with respect to the $W^{1,2}$ capacity whenever $|\mu|(K) = 0$ for every compact subset K such that $\operatorname{cap}_{W^{1,2}}(K) = 0$. This is the analogue of the notion of absolute continuity between two measures from measure theory.

We say that $x \in \Omega$ is a Lebesgue point of v and $\hat{v}(x) \in \mathbb{R}$ is the value of the precise representative of v at x whenever

$$\lim_{r\to 0} \fint_{B_r(x)} |v - \hat{v}(x)| = 0,$$

where $\fint_{B_r(x)}$ denotes the average integral over $B_r(x)$. For an L^1_{loc} function, the exceptional set (i.e., the complement of the Lebesgue set in Ω) has Lebesgue measure zero. Since in our case v is a $W^{1,2}_{\mathrm{loc}}$ function, the exceptional set of v is typically smaller and has $W^{1,2}$ capacity zero; see [Evans and Gariepy 2015, Theorem 4.19; Ponce 2016, Proposition 8.6]. Thus, the exceptional set of v is irrelevant for diffuse measures.

The existence of solutions of the Dirichlet problem (3-2) for finite diffuse measures μ is proved in [Orsina and Ponce 2008] for potentials $V \in L^1_{\mathrm{loc}}(\Omega)$ and depends upon a contraction estimate, following an idea of Brezis and Strauss [1973]:

Proposition 3.2. *Let $V \in L^1_{\mathrm{loc}}(\Omega)$ be a nonnegative function. For every finite measure μ in Ω which is diffuse with respect to the $W^{1,2}$ capacity, there exists a unique function $\zeta \in W_0^{1,1}(\Omega)$ that satisfies (3-5). In addition, $V\zeta \in L^1(\Omega)$ and*

$$\|V\zeta\|_{L^1(\Omega)} \le \|\mu\|_{\mathcal{M}(\Omega)}. \tag{3-7}$$

The contraction estimate (3-7) holds for any solution of the Dirichlet problem (3-5) provided that V is nonnegative, independently of the fact that μ is diffuse or not; see, e.g., [Brezis et al. 2007, Proposition 4.B.3]. This is a formal consequence of using $\operatorname{sgn}\zeta$ as a test function.

The classical weak maximum principle implies by duality the estimate $\|\zeta\|_{L^1(\Omega)} \le C'\|\Delta\zeta\|_{\mathcal{M}(\Omega)}$. Thus, as a consequence of (3-7) and the identity (3-5) satisfied by $\Delta\zeta$,

$$\frac{1}{C'}\|\zeta\|_{L^1(\Omega)} \le \|\Delta\zeta\|_{\mathcal{M}(\Omega)} \le \|\mu\|_{\mathcal{M}(\Omega)} + \|V\zeta\|_{L^1(\Omega)} \le 2\|\mu\|_{\mathcal{M}(\Omega)}. \tag{3-8}$$

Additionally, the existence of the distributional normal derivative $\partial\zeta/\partial n$ in $L^1(\partial\Omega)$ relies on the estimate $\|\partial\zeta/\partial n\|_{L^1(\partial\Omega)} \le \|\Delta\zeta\|_{\mathcal{M}(\Omega)}$. Proceeding as in (3-8) we get

$$\left\| \frac{\partial \zeta}{\partial n} \right\|_{L^1(\partial\Omega)} \le \|\Delta\zeta\|_{\mathcal{M}(\Omega)} \le 2\|\mu\|_{\mathcal{M}(\Omega)}. \tag{3-9}$$

To understand the role played by diffuse measures and the $W^{1,2}$ capacity in this problem, one should keep in mind a classical result in potential theory which says that for every such a measure μ one can find a sequence $(\mu_k)_{k\in\mathbb{N}}$ of finite measures that converges strongly to μ in the space of finite Borel measures $\mathcal{M}(\Omega)$ and such that the solution of the Dirichlet problem

$$\begin{cases} -\Delta w_k = \mu_k & \text{in } \Omega, \\ w_k = 0 & \text{on } \partial\Omega \end{cases} \tag{3-10}$$

is a bounded function for every $k \in \mathbb{N}$. Those measures can be obtained for example as an application of the Hahn–Banach theorem in the spirit of [Feyel and de la Pradelle 1977; Dal Maso 1983]; see [Ponce and Wilmet 2017, Proposition 2.1] for details. Another strategy relies on the Frostman–Maria boundedness principle by taking $\mu_k := \mu\lfloor_{E_k}$, where E_k is a sublevel set of the solution of the Dirichlet problem (3-10) with datum μ; see Lemma 13.2 in [Ponce 2016] and the remark following that statement. The property that $w_k \in L^\infty(\Omega)$ and $\Delta w_k \in \mathcal{M}(\Omega)$ implies by interpolation that $w_k \in W_0^{1,2}(\Omega)$; hence μ_k acts as an element in the dual space $(W_0^{1,2}(\Omega))'$.

The existence of a variational solution of the Dirichlet problem (3-2) with better datum $\mu \in (W_0^{1,2}(\Omega))'$ relies on a standard variational approach based on the minimization of the functional

$$E(z) = \frac{1}{2}\int_\Omega (|\nabla z|^2 + Vz^2) - \mu[z] \tag{3-11}$$

in the class $W_0^{1,2}(\Omega) \cap L^2(\Omega; V\,\mathrm{d}x)$. The unique minimizer ζ satisfies the Euler–Lagrange equation

$$\int_\Omega (\nabla\zeta\cdot\nabla z + V\zeta z) = \mu[z] \tag{3-12}$$

for every $z \in W_0^{1,2}(\Omega)\cap L^2(\Omega; V\,\mathrm{d}x)$. Since $V \in L^1_{\mathrm{loc}}(\Omega)$, the set $C_c^\infty(\Omega)$ is contained in the minimization class $W_0^{1,2}(\Omega)\cap L^2(\Omega; V\,\mathrm{d}x)$. The Euler–Lagrange equation implies in this case that

$$-\Delta\zeta + V\zeta = \mu \quad \text{in the sense of distributions in } \Omega. \tag{3-13}$$

When μ is in addition a finite measure in Ω, one deduces that $V\zeta \in L^1(\Omega)$ using the test function $z = T_\epsilon(\zeta)/\epsilon$, where $T_\epsilon : \mathbb{R}\to\mathbb{R}$ is the truncation function at $\pm\epsilon$ defined for $t \in \mathbb{R}$ by

$$T_\epsilon(t) = \begin{cases} -\epsilon & \text{if } t < -\epsilon, \\ t & \text{if } -\epsilon \le t \le \epsilon, \\ \epsilon & \text{if } t > \epsilon. \end{cases}$$

Indeed, z satisfies $\nabla\zeta\cdot\nabla z \ge 0$ and $|z| \le 1$. Applying the Euler–Lagrange equation (3-12) with z as above, and letting ϵ tend to zero, the contraction estimate (3-7) follows from Fatou's lemma. Equation (3-13)

thus implies that $\Delta\zeta$ is also a finite measure in Ω. One may prove that $\Delta\zeta$ is a finite measure even for nonnegative Borel functions V, although (3-13) need not be satisfied in this case; see Section 8 below.

We illustrate these tools with a sketch of the proof of Proposition 3.2:

Proof of Proposition 3.2. By linearity, it suffices to consider the case where μ is nonnegative. Take a sequence of measures $(\mu_k)_{k\in\mathbb{N}}$ converging strongly to μ in $\mathcal{M}(\Omega)$ such that the solution w_k of the Dirichlet problem (3-10) with density μ_k is bounded. Hence, μ_k belongs to $(W_0^{1,2}(\Omega))'$ and then the Dirichlet problem (3-2) with datum μ_k has a unique solution $\zeta_k \in W_0^{1,2}(\Omega)$. By the linearity of (3-13), we have the contraction estimate

$$\|V\zeta_k - V\zeta_l\|_{L^1(\Omega)} \le \|\mu_k - \mu_l\|_{\mathcal{M}(\Omega)}$$

for every $k, l \in \mathbb{N}$. Hence, the sequence $(V\zeta_k)_{k\in\mathbb{N}}$ is Cauchy in $L^1(\Omega)$. Therefore, $(\Delta\zeta_k)_{k\in\mathbb{N}}$ converges strongly in $\mathcal{M}(\Omega)$, and thus by the elliptic estimates of Littman, Stampacchia and Weinberger [Littman et al. 1963], see also [Ponce 2016, Proposition 5.1], the sequence $(\zeta_k)_{k\in\mathbb{N}}$ converges strongly in $W_0^{1,p}(\Omega)$ for every $1 \le p < N/(N-1)$. In particular, its limit ζ belongs to $W_0^{1,1}(\Omega)$ and satisfies (3-13). □

Before proving Proposition 3.1, we first address a question of existence of solutions that includes the easier L^1-L^∞ duality setting and we also develop an approximation scheme of solutions that will be used in the next section:

Lemma 3.3. *Let $(g_k)_{k\in\mathbb{N}}$ be a uniformly bounded sequence in $C^\infty(\partial\Omega)$ that converges almost everywhere to $g \in L^\infty(\partial\Omega)$. Then, for every $f \in L^\infty(\Omega)$ and $k \in \mathbb{N}$, there exist $f_k \in L^1_{\mathrm{loc}}(\Omega)$ and $v_k \in W^{1,2}(\Omega)\cap L^\infty(\Omega)$ such that, for every minimizer $\zeta \in W_0^{1,2}(\Omega) \cap L^2(\Omega;\, V\,\mathrm{d}x)$ of the energy functional (3-11) with datum $\mu \in (W_0^{1,2}(\Omega))' \cap \mathcal{M}(\Omega)$, we have*

(i) *$f_k \in L^1(\Omega;\, |\zeta|\,\mathrm{d}x)$, $v_k = g_k$ in the sense of traces on $\partial\Omega$ and*

$$\int_\Omega (\nabla v_k \cdot \nabla z + V v_k z) = \int_\Omega f_k z \quad \text{for every } z \in W_0^{1,2}(\Omega) \cap L^1(\Omega;\, V\,\mathrm{d}x),$$

(ii) *$(f_k)_{k\in\mathbb{N}}$ converges to f in $L^1(\Omega;\, |\zeta|\,\mathrm{d}x)$,*

(iii) *$(v_k)_{k\in\mathbb{N}}$ is uniformly bounded and converges in $L^1(\Omega)$ to the L^1-L^∞ duality solution of (3-1).*

Proof of Lemma 3.3. We construct the function v_k of the form $v_k = u_k + \psi_k$, where $\psi_k \in C^\infty(\overline{\Omega})$ is the harmonic extension of g_k to $\overline{\Omega}$ and $u_k \in W_0^{1,2}(\Omega)$ satisfies

$$-\Delta u_k + V u_k = f - T_k(V)\psi_k \quad \text{in the sense of distributions in } \Omega.$$

Our motivation is that v_k formally satisfies

$$\begin{cases} -\Delta v_k + V v_k = f + (V - T_k(V))\psi_k & \text{in } \Omega, \\ v_k = g_k & \text{on } \partial\Omega, \end{cases}$$

with a warning concerning the fact that $V\psi_k$ need not belong to $L^1(\Omega)$.

Let

$$f_k := f + (V - T_k(V))\psi_k \in L^1_{\mathrm{loc}}(\Omega). \tag{3-14}$$

The assumption $\mu \in \mathcal{M}(\Omega)$ implies $V\zeta \in L^1(\Omega)$; see (3-8). Since both f and ψ_k are bounded, we then have $f_k \in L^1(\Omega; |\zeta|\,dx)$ and the sequence $(f_k)_{k\in\mathbb{N}}$ converges to f in this space by the dominated convergence theorem. We now show that

$$\int_\Omega (\nabla v_k \cdot \nabla z + V v_k z) = \int_\Omega f_k z \quad \text{for every } z \in W_0^{1,2}(\Omega) \cap L^1(\Omega; V\,dx), \tag{3-15}$$

and in particular with $z = \zeta$.

On one hand, we observe that u_k can be obtained by minimization of the energy functional (3-11) with datum $f - T_k(V)\psi_k$. The Euler–Lagrange equation satisfied by u_k with test function z gives in this case

$$\int_\Omega (\nabla u_k \cdot \nabla z + V u_k z) = \int_\Omega (f - T_k(V)\psi_k) z. \tag{3-16}$$

On the other hand, since $z \in W_0^{1,2}(\Omega)$ and ψ_k is the harmonic extension of g_k,

$$\int_\Omega \nabla\psi_k \cdot \nabla z = -\int_\Omega \Delta\psi_k z = 0. \tag{3-17}$$

For $z \in L^1(\Omega; V\,dx)$, the integral $\int_\Omega V\psi_k z$ is finite since ψ_k is bounded. Thus adding this integral on both sides of (3-16) and using (3-17) we get (3-15).

We now prove that $(v_k)_{k\in\mathbb{N}}$ is uniformly bounded. To this end, it suffices to establish that $(u_k)_{k\in\mathbb{N}}$ has such a property, and this follows from the pointwise estimate

$$|u_k| \le \|\psi_k\|_{L^\infty(\Omega)} + \|f\|_{L^\infty(\Omega)}\zeta \quad \text{almost everywhere in } \Omega, \tag{3-18}$$

where $\zeta \in C_0^\infty(\overline{\Omega})$ is the solution of the Dirichlet problem (1-1) with constant density $\mu \equiv 1$. Indeed, the function $Z_k := u_k - \|\psi_k\|_{L^\infty(\Omega)} - \|f\|_{L^\infty(\Omega)}\zeta$ satisfies

$$\Delta Z_k = V u_k + T_k(V)\psi_k - f + \|f\|_{L^\infty(\Omega)} \quad \text{in the sense of distributions in } \Omega.$$

Thus, by the classical Kato's inequality [1972], see also [Ponce 2016, Proposition 6.6], and the nonnegativity of V,

$$\Delta Z_k^+ \ge \chi_{\{Z_k \ge 0\}}(V u_k + T_k(V)\psi_k - f + \|f\|_{L^\infty(\Omega)}) \ge 0$$

in the sense of distributions in Ω. Since $Z_k^+ \in W_0^{1,2}(\Omega)$, the weak maximum principle implies $Z_k^+ \le 0$ almost everywhere in Ω; see, e.g., [Ponce 2016, Propositions 6.1 and 6.5]. Therefore, Z_k^+ vanishes in Ω and we get

$$u_k \le \|\psi_k\|_{L^\infty(\Omega)} + \|f\|_{L^\infty(\Omega)}\zeta \quad \text{almost everywhere in } \Omega.$$

A similar estimate holds for $-u_k$ and one deduces (3-18). As $\|\psi_k\|_{L^\infty(\Omega)} = \|g_k\|_{L^\infty(\partial\Omega)}$ and the sequence $(g_k)_{k\in\mathbb{N}}$ is uniformly bounded, we deduce that $(v_k)_{k\in\mathbb{N}}$ is uniformly bounded. This type of property where the potential of the Schrödinger operator forces the equation to have bounded solutions from data that are merely L^1 has been further investigated by Arcoya and Boccardo [2015], based on suitable choices of test functions.

We are left with the convergence of the sequence $(v_k)_{k\in\mathbb{N}}$. We have just proved that $(v_k)_{k\in\mathbb{N}}$ is uniformly bounded. Since $V \in L^1_{\mathrm{loc}}(\Omega)$ and

$$-\Delta v_k + V v_k = f_k \quad \text{in the sense of distributions in } \Omega,$$

the sequence $(\Delta v_k)_{k\in\mathbb{N}}$ is bounded in $L^1(\omega)$ for every $\omega \Subset \Omega$. It then follows by interpolation that $(v_k)_{k\in\mathbb{N}}$ is bounded in $W^{1,2}(\omega)$ for every $\omega \Subset \Omega$. Thus, there exists a subsequence $(v_{k_j})_{j\in\mathbb{N}}$ that converges to some function $v \in W^{1,2}_{\mathrm{loc}}(\Omega)$ weakly in $W^{1,2}(\omega)$ for every $\omega \Subset \Omega$ and strongly in $L^1(\Omega)$; the latter holds in the entire domain Ω by uniform boundedness of $(v_k)_{k\in\mathbb{N}}$.

To identify the limit v, we return to (3-16) and (3-17) to prove that under the additional assumption that $\mu \in L^\infty(\Omega)$, one has

$$\int_\Omega v_k \mu = \int_\Omega f_k \zeta - \int_{\partial\Omega} g_k \frac{\partial \zeta}{\partial n}\, \mathrm{d}\sigma. \tag{3-19}$$

For such a μ, the quantity $\mu[u_k]$ can be computed through integration of $u_k\mu$. From the Euler–Lagrange equation satisfied by ζ with test function u_k and (3-16), we get

$$\int_\Omega u_k \mu = \mu[u_k] = \int_\Omega (f - T_k(V)\psi_k)\zeta.$$

Since $v_k = g$ on $\partial\Omega$ and ζ has a distributional normal derivative, (3-17) with $z = \zeta$ implies

$$-\int_\Omega \psi_k\, \Delta\zeta + \int_{\partial\Omega} g_k \frac{\partial\zeta}{\partial n}\, \mathrm{d}\sigma = 0.$$

Thus adding $\int_\Omega V\psi_k\zeta$ on both sides and using the fact that $\mu = -\Delta\zeta + V\zeta$ in the sense of measures in Ω, we get

$$\int_\Omega \psi_k \mu = \int_\Omega V\psi_k\zeta - \int_{\partial\Omega} g_k \frac{\partial\zeta}{\partial n}\, \mathrm{d}\sigma.$$

A combination of the first and third identities implies (3-19). As $k = k_j$ tends to infinity, we have v satisfies

$$\int_\Omega v\mu = \int_\Omega f\zeta - \int_{\partial\Omega} g \frac{\partial\zeta}{\partial n}\, \mathrm{d}\sigma \quad \text{for every } \mu \in L^\infty(\Omega).$$

This already gives the uniqueness of the limit and in particular the entire sequence $(v_k)_{k\in\mathbb{N}}$ converges to v in $L^1(\Omega)$. That this identity holds for every $\mu \in L^1(\Omega)$ follows from approximation of μ by bounded functions $(\mu_k)_{k\in\mathbb{N}}$. Indeed, the solutions $(\zeta_k)_{k\in\mathbb{N}}$ associated to that sequence converge to ζ in $L^1(\Omega)$ and $(\partial\zeta_k/\partial n)_{k\in\mathbb{N}}$ converges to $\partial\zeta/\partial n$ in $L^1(\partial\Omega)$ by estimates (3-8) and (3-9). It thus suffices to use μ_k as test function and let k tend to infinity. □

A finite measure ν in Ω that belongs to the dual space $(W^{1,2}_0(\Omega))'$ satisfies

$$\left|\int_\Omega \varphi\, \mathrm{d}\nu\right| \le C\|\varphi\|_{W^{1,2}(\Omega)} \quad \text{for every } \varphi \in C^\infty_c(\Omega). \tag{3-20}$$

By the density of $C^\infty_c(\Omega)$ in $W^{1,2}_0(\Omega)$, the linear functional

$$\varphi \in C^\infty_c(\Omega) \longmapsto \int_\Omega \varphi\, \mathrm{d}\nu$$

has a unique continuous extension to $W_0^{1,2}(\Omega)$. Denoting such an extension by $\nu[u]$ for every $u \in W_0^{1,2}(\Omega)$, one can represent $\nu[u]$ as integration of u with respect to ν. Indeed, estimate (3-20) implies that ν, as a measure, is diffuse with respect to the $W^{1,2}$ capacity, the precise representative $\hat{u}$ has an exceptional set with $W^{1,2}$ capacity zero, and $\hat{u} \in L^1(\Omega; \nu)$; see, e.g., [Grun-Rehomme 1977; Ponce 2016, Proposition 16.5]. Moreover,

$$\nu[u] = \int_\Omega \hat{u} \, \mathrm{d}\nu \quad \text{for every } u \in W_0^{1,2}(\Omega). \tag{3-21}$$

Proof of Proposition 3.1. Estimate (3-6) is a straightforward consequence of the L^1-L^∞ duality. Indeed, for any $\mu \in L^1(\Omega)$, by estimates (3-8) and (3-9) we have

$$\begin{aligned}\left|\int_\Omega v\mu\right| = |\mu[v]| &\le \|f\|_{L^\infty(\Omega)}\|\zeta\|_{L^1(\Omega)} + \|g\|_{L^\infty(\partial\Omega)}\left\|\frac{\partial\zeta}{\partial n}\right\|_{L^1(\partial\Omega)} \\ &\le 2(C'\|f\|_{L^\infty(\Omega)} + \|g\|_{L^\infty(\partial\Omega)})\|\mu\|_{L^1(\Omega)}.\end{aligned}$$

By L^1-L^∞ duality, we deduce that $v \in L^\infty(\Omega)$ and

$$\|v\|_{L^\infty(\Omega)} \le 2(C'\|f\|_{L^\infty(\Omega)} + \|g\|_{L^\infty(\partial\Omega)}).$$

We thus have the conclusion with $C := 2\max\{C', 1\}$.

The proof of Lemma 3.3 may be seen as a first step in establishing Proposition 3.1. We follow the notation there: We recall that $(g_k)_{k\in\mathbb{N}}$ is a uniformly bounded sequence in $C^\infty(\partial\Omega)$ that converges almost everywhere to g and $(f_k)_{k\in\mathbb{N}}$ is defined by (3-14) and converges to f in $L^1(\Omega; |\zeta|\,\mathrm{d}x)$, where ζ is the solution of (3-5) with $\mu \in (W_0^{1,2}(\Omega))' \cap \mathcal{M}(\Omega)$. The sequence $(v_k)_{k\in\mathbb{N}}$ defined by $v_k = u_k + \psi_k$ is uniformly bounded and also bounded in $W^{1,2}(\omega)$ for every open subset $\omega \Subset \Omega$.

We now prove that if $\mu \in (W_0^{1,2}(\Omega))' \cap \mathcal{M}(\Omega)$ has *compact support in* Ω, then

$$\mu[v_k\varphi] = \int_\Omega f_k\zeta - \int_{\partial\Omega} g_k\frac{\partial\zeta}{\partial n}\,\mathrm{d}\sigma, \tag{3-22}$$

where $\varphi \in C_c^\infty(\Omega)$ is any function such that $\varphi = 1$ on $\operatorname{supp}\mu$. Proceeding as in the case where μ was assumed to belong to $L^\infty(\Omega)$, we have

$$\mu[u_k] = \int_\Omega (f - T_k(V)\psi_k)\zeta \tag{3-23}$$

and

$$\int_\Omega \psi_k \,\mathrm{d}\mu = \int_\Omega V\psi_k\zeta - \int_{\partial\Omega} g_k\frac{\partial\zeta}{\partial n}\,\mathrm{d}\sigma. \tag{3-24}$$

For $\varphi \in C_c^\infty(\Omega)$ such that $\varphi = 1$ on $\operatorname{supp}\mu$, we also have

$$\mu[u_k] = \mu[u_k\varphi] \quad \text{and} \quad \int_\Omega \psi_k \,\mathrm{d}\mu = \int_\Omega \psi_k\,\varphi \,\mathrm{d}\mu = \mu[\psi_k\varphi]. \tag{3-25}$$

A combination of (3-23), (3-24) and (3-25) then implies (3-22).

Take a subsequence $(v_{k_j})_{j\in\mathbb{N}}$ that converges to v weakly in $W^{1,2}(\omega)$ for every $\omega \Subset \Omega$ and in $L^1(\Omega)$. By (3-21), we have

$$\lim_{j\to\infty} \mu[v_{k_j}\varphi] = \mu[v\varphi] = \int_\Omega \widehat{v\varphi}\, \mathrm{d}\mu = \int_\Omega \hat{v}\, \mathrm{d}\mu.$$

In view of the convergences of $(f_k)_{k\in\mathbb{N}}$ and $(g_k)_{k\in\mathbb{N}}$, as $k = k_j$ tends to infinity in (3-22) we conclude that

$$\int_\Omega \hat{v}\, \mathrm{d}\mu = \int_\Omega f\zeta - \int_{\partial\Omega} g\frac{\partial\zeta}{\partial n}\, \mathrm{d}\sigma, \tag{3-26}$$

when $\mu \in \mathcal{M}(\Omega)\cap(W_0^{1,2}(\Omega))'$ has compact support in Ω.

We finally prove that v satisfies identity (3-4) for every test function ζ as in the statement. To this end, we may assume that μ is nonnegative. As in the proof of Proposition 3.2, take a sequence $(\mu_k)_{k\in\mathbb{N}}$ of finite measures that converges strongly to μ in $\mathcal{M}(\Omega)$ and such that, for each measure μ_k, the solution w_k of the Dirichlet problem (3-10) with density μ_k is bounded. We may also assume that each μ_k has compact support in Ω. By interpolation, $w_k \in W_0^{1,2}(\Omega)$ and then $\mu_k \in (W_0^{1,2}(\Omega))'$. Denoting by $\zeta_k \in W_0^{1,2}(\Omega)\cap L^2(\Omega; V\,\mathrm{d}x)$ the solution of (3-5) with μ_k, it follows from (3-26) that

$$\int_\Omega \hat{v}\, \mathrm{d}\mu_k = \int_\Omega f\zeta_k - \int_{\partial\Omega} g\frac{\partial\zeta_k}{\partial n}\, \mathrm{d}\sigma.$$

On one hand, since the function $\hat{v}$ is bounded, by strong convergence of the sequence $(\mu_k)_{k\in\mathbb{N}}$ we have

$$\lim_{k\to\infty}\int_\Omega \hat{v}\, \mathrm{d}\mu_k = \int_\Omega \hat{v}\, \mathrm{d}\mu.$$

On the other hand, by estimates (3-8) and (3-9) the sequence $(\zeta_k)_{k\in\mathbb{N}}$ converges to ζ in $L^1(\Omega)$ and $(\partial\zeta_k/\partial n)_{k\in\mathbb{N}}$ converges to $\partial\zeta/\partial n$ in $L^1(\partial\Omega)$. By the boundedness of f and g we get

$$\int_\Omega \hat{v}\, \mathrm{d}\mu = \lim_{k\to\infty}\left(\int_\Omega f\zeta_k - \int_{\partial\Omega} g\frac{\partial\zeta_k}{\partial n}\, \mathrm{d}\sigma\right) = \int_\Omega f\zeta - \int_{\partial\Omega} g\frac{\partial\zeta}{\partial n}\, \mathrm{d}\sigma. \qquad \square$$

4. Construction of positive test functions

Given any nontrivial nonnegative boundary datum $g \in L^\infty(\partial\Omega)$, the main result of this section gives a recipe to construct $f \in L^\infty(\Omega)$ such that $f < 0$ almost everywhere in Ω while the Dirichlet problem (3-1) with mixed sign datum (f, g) has a *nonnegative* solution.

Proposition 4.1. *There exists a bounded continuous function $H : \mathbb{R}_+ \to \mathbb{R}_+$, with $H(t) > 0$ for $t > 0$, such that, for any nonnegative functions $V \in L^1_{\mathrm{loc}}(\Omega)$ and $g \in L^\infty(\partial\Omega)$, if w satisfies the Dirichlet problem*

$$\begin{cases} -\Delta w + Vw = 0 & \text{in } \Omega,\\ w = g & \text{on } \partial\Omega, \end{cases}$$

and if v satisfies

$$\begin{cases} -\Delta v + Vv = H(w) & \text{in } \Omega,\\ v = 0 & \text{on } \partial\Omega, \end{cases}$$

then we have $w \geq v$ almost everywhere in Ω.

We illustrate this proposition with a proof of Theorem 1 for smooth supersolutions involving potentials in $L^1(\Omega; d_{\partial\Omega}\,\mathrm{d}x)$. An important ingredient is the following strong maximum principle for L^1 potentials that was proved independently by Ancona [1979] and Trudinger [1978]; see also [Brezis and Ponce 2003]:

Proposition 4.2. *Let $V \in L^1_{\mathrm{loc}}(\Omega)$. If $u \in L^1(\Omega)$ is a nonnegative supersolution of the Schrödinger operator $-\Delta + V$ and if $\int_\Omega u > 0$, then $u > 0$ almost everywhere in Ω.*

Proof of Theorem 1 when $V \in L^1(\Omega; d_{\partial\Omega}\,\mathrm{d}x)$ and $u \in C_0^\infty(\overline{\Omega})$. Since u is nonnegative, we may assume from the beginning that V is also nonnegative. Assume by contradiction that the set $A := \{\partial u/\partial n = 0\}$ is not negligible with respect to the surface measure on $\partial\Omega$. We solve the Dirichlet problems of Proposition 4.1 starting with the boundary condition $g = \chi_A$.

Using the notation in that statement, the function $w - v$ is nonnegative. Since $w - v$ satisfies the Dirichlet problem with datum $(-H(w), \chi_A)$, using u as test function, i.e., taking $u = \zeta$ in (3-3), we have

$$\int_\Omega (w - v)(-\Delta u + Vu) = \int_\Omega (-H(w))u - \int_{\partial\Omega} \chi_A \frac{\partial u}{\partial n}\,\mathrm{d}\sigma.$$

Observe that u is an admissible test function because $V \in L^1(\Omega; d_{\partial\Omega}\,\mathrm{d}x)$, which implies $-\Delta u + Vu \in L^1(\Omega)$. By the choice of A, the last integral vanishes. Thus,

$$\int_\Omega H(w)u = -\int_\Omega (w - v)(-\Delta u + Vu) \le 0,$$

by nonnegativity of the integrand on the right-hand side. This implies $H(w)u = 0$ almost everywhere in Ω. Since A has positive surface measure, w and v are not identically zero; this follows from the fact that one can use as a test function in both problems the solution of the Dirichlet problem (1-1) with datum $\mu \equiv 1$ (cf. Proposition 5.1 below). By the strong maximum principle above for the Schrödinger operator $-\Delta + V$, we have $w > 0$ almost everywhere in Ω; hence $H(w)$ satisfies the same property. Therefore, $u \equiv 0$ and the conclusion follows for smooth supersolutions. □

To prove Proposition 4.1, we need a version of Kato's inequality adapted to the nonhomogeneous Dirichlet problem involving potentials $V \in L^1_{\mathrm{loc}}(\Omega)$.

Lemma 4.3. *Let $V \in L^1_{\mathrm{loc}}(\Omega)$ be a nonnegative function. Given $f \in L^\infty(\Omega)$ and $g \in L^\infty(\partial\Omega)$, if $v \in L^\infty(\Omega)$ satisfies the Dirichlet problem (3-1), then*

$$\int_\Omega v^+ \le \int_{\{v>0\}} f\xi - \int_{\partial\Omega} g^+ \frac{\partial\xi}{\partial n}\,\mathrm{d}\sigma,$$

where $\xi \in W_0^{1,2}(\Omega) \cap L^\infty(\Omega)$ is the (nonnegative) solution of the Dirichlet problem (3-2) with datum $\mu \equiv 1$.

Proof of Lemma 4.3. The proof is based on an approximation of v by functions $v_k \in W^{1,2}(\Omega) \cap L^\infty(\Omega)$, following the notation in Lemma 3.3. By the contraction estimate (3-7), $\Delta\xi \in L^1(\Omega)$ and ξ has a distributional normal derivative $\partial\xi/\partial n \in L^1(\partial\Omega)$. Thus,

$$\int_\Omega (\nabla\xi \cdot \nabla\psi + V\xi\psi) = \int_\Omega \psi + \int_{\partial\Omega} \psi \frac{\partial\xi}{\partial n}\,\mathrm{d}\sigma \quad \text{for every } \psi \in C^\infty(\overline{\Omega}).$$

Since $\xi \in W_0^{1,2}(\Omega)$ and $V\xi \in L^1(\Omega)$, this identity holds by approximation of ψ for every $\psi \in W^{1,2}(\Omega) \cap L^\infty(\Omega)$. In particular, we can take $\psi = J(v_k)$, where $J : \mathbb{R} \to \mathbb{R}$ is a smooth function to be chosen later on. Since $J(v_k) = J(g_k)$ in the sense of traces on $\partial\Omega$, where $g_k \in C^\infty(\partial\Omega)$ is an approximation of g, we get

$$\int_\Omega \big(J'(v_k)\nabla\xi \cdot \nabla v_k + V\xi J(v_k)\big) = \int_\Omega J(v_k) + \int_{\partial\Omega} J(g_k)\frac{\partial\xi}{\partial n}\,\mathrm{d}\sigma. \tag{4-1}$$

On the other hand, applying the Euler–Lagrange equation in the statement of Lemma 3.3 with test function $z = J'(v_k)\xi$ (which satisfies $|z| \le C|\xi|$ and thus $z \in L^1(\Omega; V\,\mathrm{d}x)$), we have

$$\int_\Omega \big(J''(v_k)|\nabla v_k|^2\xi + J'(v_k)\nabla\xi \cdot \nabla v_k + V v_k J'(v_k)\xi\big) = \int_\Omega f_k J'(v_k)\xi.$$

Assuming that $J'' \ge 0$, by the nonnegativity of ξ we get

$$\int_\Omega \big(J'(v_k)\nabla v_k \cdot \nabla\xi + V v_k J'(v_k)\xi\big) \le \int_\Omega f_k J'(v_k)\xi. \tag{4-2}$$

Subtracting (4-2) from (4-1), we get

$$\int_\Omega V\xi[J(v_k) - v_k J'(v_k)] \ge \int_\Omega J(v_k) - \int_\Omega f_k J'(v_k)\xi + \int_{\partial\Omega} J(g_k)\frac{\partial\xi}{\partial n}\,\mathrm{d}\sigma.$$

We now take J convex such that $J(t) = 0$ for $t \le 0$ and $0 \le J(t) \le t$ for $t \ge 0$. In particular, for every $t \in \mathbb{R}$ we have $J(t) \le J'(t)t$. Since V and ξ are nonnegative, the integrand on the left-hand side is nonpositive and we deduce that

$$\int_\Omega J(v_k) \le \int_\Omega f_k J'(v_k)\xi - \int_{\partial\Omega} J(g_k)\frac{\partial\xi}{\partial n}\,\mathrm{d}\sigma.$$

The sequence $(f_k)_{k\in\mathbb{N}}$ provided by Lemma 3.3 converges to f in $L^1(\Omega; \xi\,\mathrm{d}x)$. As k tends to infinity, by pointwise convergence and the boundedness of $(v_k)_{k\in\mathbb{N}}$ and $(g_k)_{k\in\mathbb{N}}$ and by the dominated convergence theorem we thus get

$$\int_\Omega J(v) \le \int_\Omega f J'(v)\xi - \int_{\partial\Omega} J(g)\frac{\partial\xi}{\partial n}\,\mathrm{d}\sigma.$$

To conclude, we apply this inequality to a sequence $(J_i)_{i\in\mathbb{N}}$ of convex functions as above that converges pointwise to $t \mapsto t^+$ and such that $(J_i')_{i\in\mathbb{N}}$ converges pointwise to $\chi_{(0,\infty)}$. As i tends to infinity, we have the conclusion. □

Proof of Proposition 4.1. For every $\epsilon > 0$, we claim that if z_ϵ satisfies the Dirichlet problem

$$\begin{cases} -\Delta z_\epsilon + V z_\epsilon = \chi_{\{w>\epsilon C\}} & \text{in } \Omega, \\ z_\epsilon = 0 & \text{on } \partial\Omega \end{cases}$$

for $C > 0$ sufficiently large, then we have $w \ge \epsilon z_\epsilon$. Indeed, by Kato's inequality above applied to $v = \epsilon z_\epsilon - w$, we have

$$\int_\Omega (\epsilon z_\epsilon - w)^+ \le \int_{\{\epsilon z_\epsilon > w\}} \epsilon\,\chi_{\{w>\epsilon C\}}\xi - \int_{\partial\Omega} (-g)^+\frac{\partial\xi}{\partial n}\,\mathrm{d}\sigma = \epsilon \int_{\{\epsilon z_\epsilon > w > \epsilon C\}} \xi,$$

since g is assumed to be nonnegative. Observe that

$$z_\epsilon \le \xi \le \|\xi\|_{L^\infty(\Omega)} \quad \text{for every } \epsilon > 0.$$

Taking $C := \|\xi\|_{L^\infty(\Omega)}$, the set $\{\epsilon z_\epsilon > w > \epsilon C\}$ is then negligible and we deduce that

$$\int_\Omega (\epsilon z_\epsilon - w)^+ \le 0.$$

Hence, $w \ge \epsilon z_\epsilon$ almost everywhere in Ω. Applying this conclusion with $\epsilon = 1/2^k$ for every $k \in \mathbb{N}_*$, we get

$$w = \sum_{k=1}^{\infty} \frac{1}{2^k} w \ge \sum_{k=1}^{\infty} \frac{1}{2^k} \cdot \frac{1}{2^k} z_{1/2^k} =: \tilde{v},$$

where $\tilde{v}$ satisfies the Dirichlet problem

$$\begin{cases} -\Delta \tilde{v} + V\tilde{v} = \tilde{f} & \text{in } \Omega, \\ \tilde{v} = 0 & \text{on } \partial\Omega, \end{cases}$$

with

$$\tilde{f}(x) := \sum_{k=1}^{\infty} \frac{1}{2^k} \cdot \frac{1}{2^k} \chi_{\{w > C/2^k\}}(x) = \sum_{k=1}^{\infty} \frac{1}{2^k} \cdot \frac{1}{2^k} \chi_{(C/2^k,\infty)}(w(x)).$$

Observe that for every $s \ge 0$,

$$\begin{aligned} \sum_{k=1}^{\infty} \frac{1}{2^k} \cdot \frac{1}{2^k} \chi_{(C/2^k,\infty)}(s) &\ge \sum_{k=1}^{\infty} \int_{1/2^k}^{1/2^{k-1}} \frac{t}{2} \chi_{(Ct,\infty)}(s)\, \mathrm{d}t \\ &= \int_0^1 \frac{t}{2} \chi_{(0,s/C)}(t)\, \mathrm{d}t = \int_0^{\min\{1,s/C\}} \frac{t}{2}\, \mathrm{d}t = \tfrac{1}{4}(\min\{1, s/C\})^2 =: H(s). \end{aligned}$$

By this computation, we thus have $\tilde{f} \ge H(w)$ in Ω. Hence by comparison it follows that the solution v of the Dirichlet problem with datum $H(w)$ satisfies $\tilde{v} \ge v$ in Ω. Therefore, $w \ge v$ as we wanted to prove. □

5. Nontrivial solutions for the nonhomogeneous problem

The existence of nontrivial solutions of the boundary value problem

$$\begin{cases} -\Delta w + Vw = 0 & \text{in } \Omega, \\ w = g & \text{on } \partial\Omega \end{cases} \tag{5-1}$$

in the sense of Proposition 3.1 is related to the existence of supersolutions of the Schrödinger operator $-\Delta + V$ with negative normal derivative through the following:

Proposition 5.1. *Let $V \in L^1_{\mathrm{loc}}(\Omega)$ and $g \in L^\infty(\partial\Omega)$ be nonnegative functions. Then, the (nonnegative) solution $w \in W^{1,2}_{\mathrm{loc}}(\Omega) \cap L^\infty(\Omega)$ of the Dirichlet problem (5-1) with datum g satisfies*

$$\int_\Omega w > 0$$

if and only if there exists a (nonnegative) supersolution $\zeta \in W_0^{1,1}(\Omega)$ *of the Schrödinger operator* $-\Delta + V$ *such that the measure* $-\Delta\zeta + V\zeta$ *is finite and diffuse in* Ω *and*

$$\int_{\partial\Omega} g \frac{\partial \zeta}{\partial n} \, \mathrm{d}\sigma < 0. \tag{5-2}$$

Proposition 5.1 is used in the proofs of Theorems 1 and 2. We deduce a posteriori from Theorem 2 that once condition (5-2) holds for *one* supersolution, then it holds for *every* nontrivial supersolution $\zeta \in W_0^{1,1}(\Omega)$ such that $V\zeta \in L^1(\Omega)$.

The nonnegativity of w and ζ follows from Proposition 3.1 and Lemma 4.3. Indeed, by Lemma 4.3 applied to $v := -w$ we have

$$\int_\Omega (-w)^+ \le -\int_{\partial\Omega} (-g)^+ \frac{\partial \xi}{\partial n} \, \mathrm{d}\sigma = 0.$$

Thus, $-w \le 0$ almost everywhere in Ω. The same argument applies to solutions of (3-1) such that $f \ge 0$ in Ω and $g \ge 0$ on $\partial\Omega$. In particular, by Proposition 3.1 and the nonnegativity of the solutions of (3-1) in this case we have

$$\int_\Omega f\zeta - \int_{\partial\Omega} g \frac{\partial \zeta}{\partial n} \, \mathrm{d}\sigma = \int_\Omega \hat{v} \, \mathrm{d}\mu \ge 0,$$

where $\mu = -\Delta\zeta + V\zeta$. Since this is true for every nonnegative f and g, we deduce that $\zeta \ge 0$ in Ω and $\partial\zeta/\partial n \le 0$ on $\partial\Omega$.

Proof of Proposition 5.1. "$\Longleftarrow$" Since $\partial\zeta/\partial n \le 0$ almost everywhere on $\partial\Omega$, by Proposition 3.1 we have

$$\int_\Omega \hat{w} \, \mathrm{d}\mu = -\int_{\partial\Omega} g \frac{\partial \zeta}{\partial n} \, \mathrm{d}\sigma > 0.$$

In particular, w is a nonzero solution of (5-1). Since g is nonnegative on $\partial\Omega$, w is nonnegative in Ω and the implication follows.

"$\Longrightarrow$" Since w is a nontrivial nonnegative solution, there exists a compact subset $K \subset \Omega$ with positive $W^{1,2}$ capacity which is contained in the Lebesgue set of w and is such that $\hat{w} \ge \epsilon$ on K for some constant $\epsilon > 0$. Take a finite nonnegative diffuse measure μ supported on K such that $\mu(K) > 0$. The existence of such a measure follows from the Hahn–Banach theorem; see, e.g., [Ponce 2016, Proposition A.17]. We take as supersolution the function $\zeta \in W_0^{1,1}(\Omega)$ such that

$$-\Delta\zeta + V\zeta = \mu \quad \text{in the sense of distributions in } \Omega;$$

see Proposition 3.2. Applying ζ as a test function in the Dirichlet problem satisfied by w, we have

$$\epsilon\mu(K) \le \int_\Omega \hat{w} \, \mathrm{d}\mu = -\int_{\partial\Omega} g \frac{\partial \zeta}{\partial n} \, \mathrm{d}\sigma. \qquad \square$$

The previous proposition raises the question of how to construct supersolutions of the Schrödinger operator $-\Delta + V$ with pointwise control on its distributional normal derivative to ensure that (5-2) is satisfied.

Proposition 5.2. *Let $V \in L^1_{\mathrm{loc}}(\Omega)$ be a nonnegative function. For every $w \in W_0^{1,1}(\Omega)$ such that Δw is a finite measure and $Vw \in L^1(\Omega)$, there exists a nonnegative function $\tilde{w} \in W_0^{1,2}(\Omega) \cap L^\infty(\Omega)$ such that*

(O_1) *$\Delta\tilde{w}$ is a finite measure in Ω and $V\tilde{w} \in L^1(\Omega)$,*

(O_2) *$\partial\tilde{w}/\partial n \le \partial w/\partial n$ almost everywhere on $\partial\Omega$,*

(O_3) *$-\Delta\tilde{w} + V\tilde{w} \ge 0$ in the sense of distributions in Ω.*

For example, when V is a nonnegative Hopf potential, taking $w := v$ as the function given by Proposition 2.5 with $u := \zeta_0$ one finds, as an application of Proposition 5.2, a supersolution of $-\Delta + V$ having a distributional normal derivative $\partial\tilde{w}/\partial n < 0$ almost everywhere on $\partial\Omega$.

Proof of Proposition 5.2. Since $w \in W_0^{1,1}(\Omega)$ and the measure Δw is finite in Ω, by interpolation we have $T_1(w) \in W_0^{1,2}(\Omega)$. Since $Vw \in L^1(\Omega)$, we also have $VT_1(w) \in L^1(\Omega)$. By Kato's inequality up to the boundary (2-1), $\Delta T_1(w)$ is a finite measure in Ω. By (2-2), we also have

$$\frac{\partial T_1(w)}{\partial n} = \frac{\partial w}{\partial n}. \tag{5-3}$$

Thus, replacing w by $T_1(w)$ if necessary, we may henceforth assume that $w \in W_0^{1,2}(\Omega) \cap L^\infty(\Omega)$. The measure Δw in this case is diffuse with respect to the $W^{1,2}$ capacity; hence the measure

$$\nu := -\Delta w + Vw$$

is also finite and diffuse in Ω, and so is its positive part ν^+.

By Proposition 3.2, the Dirichlet problem with nonnegative potential V,

$$\begin{cases} -\Delta z + Vz = \nu^+ & \text{in } \Omega, \\ z = 0 & \text{on } \partial\Omega, \end{cases}$$

has a solution. It is not clear for example why z is bounded, for this reason we now prove that $\tilde{w} := T_1(z)$ satisfies the required properties. The contraction estimate (3-7) implies $Vz \in L^1(\Omega)$, and hence the measure Δz is finite. Proceeding as in the first part of the proof, we have $\tilde{w} \in W_0^{1,2}(\Omega)$, $\Delta\tilde{w}$ is a finite measure in Ω, and

$$\frac{\partial\tilde{w}}{\partial n} = \frac{\partial z}{\partial n}.$$

By the comparison principle between solutions of the Dirichlet problem we have $z \ge w$ in Ω. Then, by comparison between normal derivatives,

$$\frac{\partial\tilde{w}}{\partial n} = \frac{\partial z}{\partial n} \le \frac{\partial w}{\partial n},$$

which is (O_2). Since z is nonnegative and $\Delta z \le Vz$ in the sense of distributions in Ω, a straightforward variant of Kato's inequality yields

$$\Delta\tilde{w} = \Delta(\min\{z, 1\}) \le \chi_{\{z<1\}} Vz \tag{5-4}$$

in the sense of distributions in Ω; see [Ponce 2016, Proposition 6.9]. By the nonnegativity of V, the right-hand side is smaller than $V\tilde{w}$, and we deduce that $\tilde{w}$ satisfies (O_3). □

Ancona's argument leading to (2-2) is based on tools from potential theory. There is another strategy which allows one to prove a smooth counterpart of this formula based on a PDE approach, which is enough to prove Proposition 5.2. More precisely, given a smooth function $\Phi : \mathbb{R} \to \mathbb{R}$ such that Φ'' has compact support, it has been proved in [Dal Maso et al. 1999] using the notion of renormalized solution that, for every $u \in W_0^{1,1}(\Omega)$ such that Δu is a finite measure in Ω, one has that $\Delta\Phi(u)$ is also a finite measure in Ω and the following holds:

$$\Delta\Phi(u) = \Phi'(u)(\Delta u)_{\mathrm{d}} + \Phi''(u)|\nabla u|^2. \tag{5-5}$$

Here, $(\Delta u)_{\mathrm{d}}$ denotes the part of the measure Δu which is diffuse with respect to the $W^{1,2}$ capacity that arises from the Lebesgue decomposition of measures.

The approximation scheme from [Brezis and Ponce 2008] to prove that the distributional normal derivative belongs to $L^1(\partial\Omega)$ is based on a strong approximation of the measure Δu by measures with compact support. In this case, one deduces using the identity (5-5) that the solutions u_k of the approximating Dirichlet problems are such that $(\Delta\Phi(u_k))_{k\in\mathbb{N}}$ converges strongly to $\Delta\Phi(u)$ in Ω, and one then deduces that

$$\frac{\partial\Phi(u)}{\partial n} = \Phi'(0)\frac{\partial u}{\partial n}.$$

In particular, if Φ is an approximation of the truncation function such that $\Phi'(0) = 1$, one gets an equality between the normal derivatives as in (2-2) and (5-3).

6. Proofs of the main results

Proof of Theorem 2. Since u_1 and u_2 are nonnegative, they are also supersolutions for the Schrödinger operator $-\Delta + V^+$ with nonnegative potential. We may thus assume from the beginning that V is nonnegative. We split the proof into three steps:

Step 1: We prove the theorem under the additional assumption that the measures Δu_1 and Δu_2 are finite and diffuse.

By Proposition 2.1, both $\partial u_1/\partial n$ and $\partial u_2/\partial n$ are nonpositive on $\partial\Omega$. Let us prove that

$$\frac{\partial u_1}{\partial n} < 0 \quad \text{almost everywhere on } \left\{\frac{\partial u_2}{\partial n} < 0\right\} \tag{6-1}$$

with respect to the surface measure on $\partial\Omega$. Assume by contradiction that there exists a Borel set $A \subset \{\partial u_1/\partial n = 0\}$ such that

$$\int_A \frac{\partial u_2}{\partial n}\, \mathrm{d}\sigma < 0.$$

By Proposition 5.1, the solution w of the Dirichlet problem (5-1) with datum $g = \chi_A$ is nontrivial. Since Ω is connected, by the strong maximum principle for L^1 potentials (Proposition 4.2), we then have $w > 0$ almost everywhere in Ω. Denoting by v the solution of the Dirichlet problem in Proposition 4.1 with datum $(H(w), 0)$, the function $w - v$ is nonnegative. By Proposition 3.1 applied to the solution $w - v$

and test function u_1, we get

$$0 \le \int_\Omega \widehat{w-v}\ \mathrm{d}(-\Delta u_1 + V u_1) = \int_\Omega (-H(w))u_1 - \int_{\partial\Omega} \chi_A \frac{\partial u_1}{\partial n}\ \mathrm{d}\sigma = -\int_\Omega H(w)u_1,$$

where the last equality follows from the fact that $\partial u_1/\partial n = 0$ on A. Thus, the integral on the right-hand side is nonpositive, while the integrand is nonnegative; hence $H(w)u_1 = 0$ almost everywhere in Ω. Since $H(w) > 0$ almost everywhere in Ω, we then have $u_1 = 0$ almost everywhere in Ω. This contradicts the nontriviality of u_1. Thus, (6-1) holds.

Step 2: We prove the theorem under the additional assumption that the measures Δu_1 and Δu_2 are diffuse but not necessarily finite in Ω.

Take a nonzero finite measure μ in Ω such that $0 \le \mu \le -\Delta u_1 + V u_1$. In particular, μ is diffuse and so by Proposition 3.2 the Dirichlet problem (3-2) has a solution $\tilde{u}_1$ and $\Delta\tilde{u}_1$ is a finite measure in Ω. Since V is nonnegative, by comparison we have $\tilde{u}_1 \le u_1$. By the definition of $\partial u_1/\partial n$ as an essential infimum of normal derivatives over $\mathcal{G}_{u_1}$,

$$\frac{\partial u_1}{\partial n} \le \frac{\partial \tilde{u}_1}{\partial n} \quad \text{almost everywhere on } \partial\Omega. \tag{6-2}$$

We next take any $w_2 \in \mathcal{G}_{u_2}$ and apply Proposition 5.2 to this function to get a supersolution $\tilde{w}_2 \in W_0^{1,2}(\Omega)$ of the Schrödinger operator $-\Delta + V$ such that

$$\frac{\partial \tilde{w}_2}{\partial n} \le \frac{\partial w_2}{\partial n} \quad \text{almost everywhere on } \partial\Omega. \tag{6-3}$$

Observe that both $\tilde{u}_1$ and $\tilde{w}_2$ satisfy the assumptions of the previous step. Thus,

$$\frac{\partial \tilde{u}_1}{\partial n} < 0 \quad \text{almost everywhere on } \left\{\frac{\partial \tilde{w}_2}{\partial n} < 0\right\}.$$

Combining (6-2) and (6-3), we thus get

$$\frac{\partial u_1}{\partial n} < 0 \quad \text{almost everywhere on } \left\{\frac{\partial w_2}{\partial n} < 0\right\}.$$

Since this property holds for every $w_2 \in \mathcal{G}_{u_2}$, we obtain (6-1).

Step 3: Proof of the theorem completed.

In the general case, it suffices to apply the previous argument to $T_1(u_i)$. Indeed, by Kato's inequality, the $\Delta T_1(u_i)$ are locally finite measures in Ω. Since $T_1(u_i) \in W_0^{1,2}(\Omega)$, the measures $\Delta T_1(u_i)$ are diffuse with respect to the $W^{1,2}$ capacity. A straightforward variant of Kato's inequality (cf. (5-4) above) implies that the $T_1(u_i)$ are supersolutions of $-\Delta + V$. By Step 2, assertion (6-1) above thus applies to $T_1(u_i)$. By (2-2), we have

$$\frac{\partial T_1(u_i)}{\partial n} = \frac{\partial u_i}{\partial n} \quad \text{almost everywhere on } \partial\Omega,$$

and (6-1) for u_i follows. It now suffices to switch the roles of u_1 and u_2 to conclude. □

Proof of Theorem 1. Since the supersolution u is nonnegative, we may replace V by V^+, and assume that V is nonnegative. Let ζ_0 be given by Definition 1.1. Since $\partial\zeta_0/\partial n < 0$ almost everywhere on $\partial\Omega$, it suffices to prove that

$$\frac{\partial u}{\partial n} < 0 \quad \text{almost everywhere on } \left\{\frac{\partial \zeta_0}{\partial n} < 0\right\}. \tag{6-4}$$

To this end, we apply Proposition 5.2 to any function $w \in \mathcal{G}_{\zeta_0}$ to get a nonnegative supersolution $\tilde{w} \in W_0^{1,2}(\Omega)$ of the Schrödinger operator $-\Delta + V$ such that $V\tilde{w} \in L^1(\Omega)$ and

$$\frac{\partial \tilde{w}}{\partial n} \le \frac{\partial w}{\partial n} \quad \text{almost everywhere on } \partial\Omega.$$

By Theorem 2, for almost every $x \in \partial\Omega$ we have $\partial u(x)/\partial n < 0$ if and only if $\partial\tilde{w}(x)/\partial n < 0$. In particular,

$$\frac{\partial u}{\partial n} < 0 \quad \text{almost everywhere on } \left\{\frac{\partial w}{\partial n} < 0\right\}.$$

Since this property holds for every $w \in \mathcal{G}_{\zeta_0}$, we have (6-4) and the conclusion follows. □

7. Exceptional sets for the Hopf lemma

For any negligible compact subset $K \subset \partial\Omega$, we prove that K is the level set $\{\partial u/\partial n = 0\}$ of the normal derivative of a positive smooth solution of

$$-\Delta u + Vu = 0 \quad \text{in } \Omega \tag{7-1}$$

for some $V \in L^1(\Omega; d_{\partial\Omega}\, dx)$. More generally, given any positive function $\zeta \in C_0^\infty(\overline{\Omega})$, with a normal derivative $\partial\zeta/\partial n$ that possibly vanishes on part of $\partial\Omega$, we find a solution of (7-1), for some $V \in L^1_{\mathrm{loc}}(\Omega)$, whose normal derivative vanishes on a larger subset of $\partial\Omega$ that includes K. This is the content of our next result:

Proposition 7.1. *Let $\zeta \in C_0^\infty(\overline{\Omega})$ be such that $\zeta > 0$ in Ω. For every compact set $K \subset \partial\Omega$ such that $\mathcal{H}^{N-1}(K) = 0$, there exists $u \in C^\infty(\overline{\Omega}\setminus K) \cap C_0(\overline{\Omega})$ with $0 < u \le \zeta$ in Ω such that*

(i) *$D^2u \in L^1(\Omega)$ and $D^2u/u \in L^1(\Omega; \zeta\ dx)$,*

(ii) *$\partial u/\partial n$ is well-defined in the classical sense and is continuous on $\partial\Omega$,*

(iii) *$\partial u(x)/\partial n = 0$ if and only if $x \in K$ or $\partial\zeta(x)/\partial n = 0$.*

Thus, the function $V := \Delta u/u$ belongs to $L^1(\Omega; \zeta\ dx)$ by property (i) *above; in particular, $Vu \in L^1(\Omega)$ and*

$$-\Delta u + Vu = 0 \quad \text{in } \Omega.$$

It is unclear from our construction whether the upper bound in (1-8) is satisfied by the nonnegative potential V^+. We need the following variant of a second-order inequality by Bourdaud [1991, Théorème 3]:

Lemma 7.2. *Let $H : \mathbb{R} \to \mathbb{R}$ be a convex smooth function such that H' is bounded. If $\zeta \in C_0^1(\overline{\Omega})$ is nonnegative, then for every $\varphi \in C^\infty(\overline{\Omega})$ we have*

$$\|D^2[H(\varphi)]\|_{L^1(\Omega;\zeta\,dx)} \le C(\|D^2\varphi\|_{L^1(\Omega;\zeta\,dx)} + \|\nabla\varphi\|_{L^1(\Omega)}).$$

Proof of Lemma 7.2. In view of the composition formula

$$D^2H(\varphi) = H'(\varphi)D^2\varphi + \nabla[H'(\varphi)] \otimes \nabla\varphi,$$

we only need to estimate the second term on the right-hand side. For every $e \in \mathbb{R}^N$ such that $|e| = 1$, by the convexity of H the quantity

$$(\nabla[H'(\varphi)] \otimes \nabla\varphi)[e, e] = \partial_e[H'(\varphi)]\partial_e\varphi = H''(\varphi)(\partial_e\varphi)^2$$

is nonnegative. Since $\zeta = 0$ on $\partial\Omega$, by integration by parts we get

$$\begin{aligned}\int_\Omega (\nabla[H'(\varphi)] \otimes \nabla\varphi)[e, e]\,\zeta &= -\int_\Omega H'(\varphi)(\partial^2_{e,e}\varphi\,\zeta + \partial_e\varphi\,\partial_e\zeta)\\ &\le \|H'\|_{L^\infty(\mathbb{R})}(\|D^2\varphi\|_{L^1(\Omega;\zeta\,dx)} + \|\nabla\varphi\|_{L^1(\Omega)}\|\nabla\zeta\|_{L^\infty(\Omega)}).\end{aligned}$$

This implies the conclusion. □

We also need the following property of the Hausdorff measure $\mathcal{H}^{N-1}$:

Lemma 7.3. *Let $K \subset \mathbb{R}^N$ be a compact set. For every $\epsilon > 0$ and every open set $\omega \supset K$, there exists a nonnegative function $\varphi \in C_c^\infty(\omega)$ such that $\varphi > 1$ on K and*

$$\|D^2\varphi\|_{L^1(\mathbb{R}^N;d_K\,dx)} + \|\nabla\varphi\|_{L^1(\mathbb{R}^N)} \le C\mathcal{H}^{N-1}(K) + \epsilon,$$

where $d_K : \mathbb{R}^N \to \mathbb{R}$ denotes the distance to K.

Proof of Lemma 7.3. Let $0 < \delta \le d(K, \partial\omega)/4$, and take finitely many balls $(B_{r_i}(x_i))_{i\in\{1,\dots,\ell\}}$ that intersect K such that $K \subset \bigcup_{i=1}^\ell B_{r_i}(x_i)$,

$$\sum_{i=1}^\ell r_i^{N-1} \le C'\mathcal{H}^{N-1}(K) + \epsilon,$$

and $r_i \le \delta$ for every $i \in \{1, \dots, \ell\}$. Given a nonnegative function $\theta \in C_c^\infty(B_2)$ such that $\theta > 1$ on B_1, we have the conclusion with

$$\varphi(x) = \sum_{i=1}^\ell \theta\left(\frac{x - x_i}{r_i}\right).$$

Note that for $x \in B_{2r_i}(x_i)$ we have $d_K(x) \le 3r_i$. Thus, for every $x \in \mathbb{R}^N$,

$$|D^2\varphi(x)|d_K(x) \le \sum_{i=1}^\ell \frac{3}{r_i}\left|D^2\theta\left(\frac{x - x_i}{r_i}\right)\right|.$$

A similar pointwise estimate is satisfied by $|\nabla\varphi(x)|$ and we conclude by integration over $\mathbb{R}^N$ and a change of variables in the integral. □

Proof of Proposition 7.1. Let $(\epsilon_k)_{k\in\mathbb{N}}$ be a summable sequence of positive numbers. We construct by induction a decreasing sequence of open sets $(\omega_k)_{k\in\mathbb{N}}$ that contain K and a sequence of nonnegative functions $(\varphi_k)_{k\in\mathbb{N}}$ in $C_c^\infty(\omega_k)$ such that $\varphi_k > 1$ on ω_{k+1} as follows. Take a bounded open subset $\omega_0 \subset \mathbb{R}^N$

that contains K and such that $|\omega_0| \le \epsilon_0$. Given ω_k, let $\varphi_k \in C_c^\infty(\omega_k)$ be the function given by Lemma 7.3 with open set ω_k and parameter ϵ_k. We then take an open subset ω_{k+1} such that

$$K \subset \omega_{k+1} \subset \{\varphi_k > 1\} \quad \text{and} \quad |\omega_{k+1}| \le \epsilon_{k+1}.$$

We now take a convex smooth function $H : \mathbb{R} \to \mathbb{R}$ such that $H(0) = 1$, $H(t) = 0$ for $t \ge 1$ and H' is bounded. For each $k \in \mathbb{N}$, let

$$\psi_k = H(\varphi_k)\zeta.$$

We have in particular $\psi_k = 0$ on ω_{k+1} and $\psi_k = \zeta$ on $\overline{\Omega} \setminus \omega_k$. By the triangle inequality and Lemma 7.2, we have

$$\begin{aligned}\|D^2\psi_k - D^2\zeta\|_{L^1(\Omega)} &\le \|D^2[H(\varphi_k)]\|_{L^1(\Omega;\zeta\,dx)} + C_1\|\nabla[H(\varphi_k)]\|_{L^1(\Omega)} + C_2\|H(\varphi_k) - 1\|_{L^1(\Omega)} \\ &\le C_3(\|D^2\varphi_k\|_{L^1(\Omega;\zeta\,dx)} + \|\nabla\varphi_k\|_{L^1(\Omega)} + |\omega_k|).\end{aligned}$$

Since $\zeta \in C_0^\infty(\overline{\Omega})$ and $K \subset \partial\Omega$, we have $\zeta \le C_4 d_K$ in $\overline{\Omega}$. By the choice of ω_k and φ_k and the assumption $\mathcal{H}^{N-1}(K) = 0$, we deduce that

$$\|D^2\psi_k - D^2\zeta\|_{L^1(\Omega)} \le C_5\epsilon_k. \tag{7-2}$$

In particular, the sequence $(D^2\psi_k)_{k\in\mathbb{N}}$ is bounded in $L^1(\Omega)$.

Take

$$u = \sum_{j=0}^{\infty} \frac{1}{2^{j+1}}\psi_j.$$

By construction, we have $u \in C_0(\overline{\Omega})$, u is smooth in $\overline{\Omega} \setminus K$, and $0 < u \le \zeta$ in Ω. Moreover, u has a normal derivative given pointwise by

$$\frac{\partial u}{\partial n} = \left(\sum_{j=0}^{\infty} \frac{1}{2^{j+1}} H(\varphi_j)\right)\frac{\partial \zeta}{\partial n}.$$

In particular, $\partial u/\partial n$ is continuous on $\partial\Omega$ and

$$\left\{\frac{\partial u}{\partial n} = 0\right\} = K \cup \left\{\frac{\partial \zeta}{\partial n} = 0\right\}.$$

By the L^1 estimate of $D^2\psi_k$, we also have $D^2u \in L^1(\Omega)$.

We conclude with the proof that $D^2u/u \in L^1(\Omega;\, \zeta\, dx)$. This is based on the pointwise estimate

$$\frac{|D^2u|}{u}\zeta \le \sum_{j=k}^{\infty} \frac{1}{2^{j-k}}|D^2\psi_j| \quad \text{on } \omega_k \setminus \omega_{k+1}, \tag{7-3}$$

which is a consequence of the following facts:

(a) $\psi_j = 0$ in ω_k for every $j < k$.

(b) $u \ge \zeta/2^{k+1}$ on $\Omega \setminus \omega_{k+1}$, since $\psi_j = \zeta$ on this set for every $j \ge k+1$.

By (7-3) and the triangle inequality,

$$\frac{|D^2u|}{u}\zeta \le \sum_{j=k}^{\infty} \frac{1}{2^{j-k}}|D^2\psi_j - D^2\zeta| + 2|D^2\zeta| \quad \text{on } \omega_k \setminus \omega_{k+1}.$$

By this estimate and the fact that $u = \zeta$ in $\Omega \setminus \omega_0$, we have

$$\begin{aligned}\int_\Omega \frac{|D^2u|}{u}\zeta &= \sum_{k=0}^{\infty}\int_{\omega_k\setminus\omega_{k+1}} \frac{|D^2u|}{u}\zeta + \int_{\Omega\setminus\omega_0}\frac{|D^2u|}{u}\zeta \\ &\le \sum_{k=0}^{\infty}\sum_{j=k}^{\infty}\frac{1}{2^{j-k}}\int_{\omega_k\setminus\omega_{k+1}}|D^2\psi_j - D^2\zeta| + 2\int_\Omega |D^2\zeta|.\end{aligned}$$

Interchanging the order of summation and using (7-2), we get

$$\begin{aligned}\sum_{k=0}^{\infty}\sum_{j=k}^{\infty}\frac{1}{2^{j-k}}\int_{\omega_k\setminus\omega_{k+1}}|D^2\psi_j - D^2\zeta| &= \sum_{j=0}^{\infty}\sum_{k=0}^{j}\frac{1}{2^{j-k}}\int_{\omega_k\setminus\omega_{k+1}}|D^2\psi_j - D^2\zeta| \\ &\le \sum_{j=0}^{\infty}\int_{\omega_0\setminus\omega_{j+1}}|D^2\psi_j - D^2\zeta| \le \sum_{j=0}^{\infty} C_5\epsilon_j < \infty. \qquad \square\end{aligned}$$

8. Potentials that are merely Borel functions

Instead of dealing with potentials V in $L^1_{\mathrm{loc}}(\Omega)$, one could wish to work with general Borel functions $V : \Omega \to [0, +\infty]$, but as we explain in this section, the counterparts of Theorems 1 and 2 need not be true. The minimization approach that yields a variational solution of the Dirichlet problem

$$\begin{cases} -\Delta\zeta + V\zeta = \mu & \text{in } \Omega, \\ \zeta = 0 & \text{on } \partial\Omega \end{cases}$$

with datum $\mu \in (W_0^{1,2}(\Omega))'$ can be implemented as in Section 3 above; see [Dal Maso and Mosco 1986; 1987]. However, since test functions in $C_c^\infty(\Omega)$ need not belong to the minimization class $W_0^{1,2}(\Omega) \cap L^2(\Omega; V\,\mathrm{d}x)$, the equation may not be satisfied in the sense of distributions. In this case, the following holds:

Proposition 8.1. *Let $V : \Omega \to [0, +\infty]$ be a Borel function and $\mu \in (W_0^{1,2}(\Omega))'$. If μ is a finite measure in Ω, then the variational solution $\zeta \in W_0^{1,2}(\Omega) \cap L^2(\Omega; V\,\mathrm{d}x)$ is such that $V\zeta \in L^1(\Omega)$, $\Delta\zeta$ is a finite measure in Ω and*

$$\|V\zeta\|_{L^1(\Omega)} + \|\Delta\zeta\|_{\mathcal{M}(\Omega)} \le 3\|\mu\|_{L^1(\Omega)}.$$

If in addition we have $\mu \ge 0$ in Ω, then

$$-\Delta\zeta + V\zeta \le \mu \quad \textit{in the sense of distributions in } \Omega.$$

Proof. For every $k \in \mathbb{N}$, denote by ζ_k the minimizer of the functional E_k associated to the bounded potential $V_k := T_k(V)$. In this case, the equation

$$-\Delta\zeta_k + V_k\zeta_k = \mu$$

is satisfied in the sense of distributions and

$$\|V_k\zeta_k\|_{L^1(\Omega)} \le \|\mu\|_{L^1(\Omega)}. \tag{8-1}$$

Thus, $\Delta\zeta_k$ is a finite measure in Ω and

$$\|\Delta\zeta_k\|_{\mathcal{M}(\Omega)} \le \|\mu\|_{\mathcal{M}(\Omega)} + \|V_k\zeta_k\|_{L^1(\Omega)} \le 2\|\mu\|_{\mathcal{M}(\Omega)}. \tag{8-2}$$

Since ζ_k is a minimizer of E_k and since $T_k(V) \le V$, for every $k \in \mathbb{N}$ we also have

$$E_k(\zeta_k) \le E_k(\zeta) \le E(\zeta).$$

We deduce that the sequence $(\zeta_k)_{k\in\mathbb{N}}$ is bounded in $W_0^{1,2}(\Omega)$, whence by this inequality it must converge to the minimizer ζ. By Fatou's lemma, as k tends to infinity in the contraction estimate (8-1) we deduce that $V\zeta \in L^1(\Omega)$. By lower semicontinuity of the norm and estimate (8-2), we also have that $\Delta\zeta$ is a finite measure in Ω.

Observe that if $\mu \ge 0$, then $\zeta_k \ge 0$. By the equation satisfied by ζ_k, as k tends to infinity we deduce from Fatou's lemma that

$$\int_\Omega \zeta\,(-\Delta\varphi + V\varphi) \le \mu[\varphi] = \int_\Omega \varphi\ \mathrm{d}\mu$$

for every nonnegative function $\varphi \in C_c^\infty(\Omega)$. □

Applying Proposition 8.1 to the positive and negative parts of μ, it follows that there exists a finite measure λ in Ω such that

$$-\Delta\zeta + V\zeta = \mu + \lambda \quad \text{in the sense of distributions in } \Omega. \tag{8-3}$$

This measure λ possibly depends on μ and arises due to the singular character of V, but it can vanish even for very singular potentials.

Example 8.2. Take the potential $V_\alpha : B_1 \to [0, +\infty]$ defined by

$$V_\alpha(x) = \frac{1}{|x_1|^\alpha}$$

with $\alpha \ge 1$, so that $V_\alpha \notin L^1_{\mathrm{loc}}(B_1)$. We have proved in [Orsina and Ponce 2008, Proposition 9.2] that for every exponent $\alpha \ge 1$ the Dirichlet problem uncouples in the sense that the variational solution satisfies two independent (homogeneous) Dirichlet problems in B_1^+ and B_1^-, where

$$B_1^+ = \{x \in B_1 : x_1 > 0\} \quad \text{and} \quad B_1^- = \{x \in B_1 : x_1 < 0\}.$$

Solving separately the Dirichlet problems on B_1^+ and B_1^- with $\mu \in L^2(B_1)$ and denoting by ζ_+ and ζ_- these solutions, the function

$$\zeta := \begin{cases} \zeta_+ & \text{in } B_1^+, \\ \zeta_- & \text{in } B_1^- \end{cases}$$

belongs to $W_0^{1,2}(B_1) \cap L^2(B_1; V_\alpha \, dx)$ and satisfies

$$\int_{B_1} \zeta \, (-\Delta \varphi + V_\alpha \varphi) = \int_{B_1} \varphi \mu$$

for every $\varphi \in C_c^\infty(B_1 \setminus (\partial B_1^+ \cap \partial B_1^-))$. When $\alpha \geq 2$, by Proposition 2.7 this identity actually holds for every $\varphi \in C_c^\infty(B_1)$. Hence,

$$-\Delta \zeta + V_\alpha \zeta = \mu \quad \text{in the sense of distributions in } \Omega,$$

and thus the measure λ that satisfies (8-3) is identically zero.

The singularity of V_α in the previous example is so strong that Theorems 1 and 2 are simply false. The reason is that the operator $-\Delta + V_\alpha$ with $\alpha \geq 2$ behaves as if the domain B_1 were disconnected, with two connected components B_1^+ and B_1^-. Indeed, the function ζ defined above with constant datum $\mu \equiv 1$ satisfies $\partial \zeta / \partial n < 0$ on $\partial B_1 \setminus \{x_1 = 0\}$ by a local application of the classical Hopf lemma. However, the function

$$\tilde{\zeta} = \begin{cases} \zeta_+ & \text{in } B_1^+, \\ 0 & \text{in } B_1^- \end{cases}$$

is also a supersolution for $-\Delta + V_\alpha$, but the normal derivative $\partial \tilde{\zeta} / \partial n$ is negative only on half of the boundary ∂B_1.

Acknowledgements

A. C. Ponce was supported by the Fonds de la Recherche scientifique (FNRS) under research grants J.0026.15 and J.0020.18. He warmly thanks the Dipartimento di Matematica of the "Sapienza" Università di Roma for the invitation. He also acknowledges the hospitality of the Academia Belgica in Rome.

References

[Ancona 1979] A. Ancona, "Une propriété d'invariance des ensembles absorbants par perturbation d'un opérateur elliptique", *Comm. Partial Differential Equations* **4**:4 (1979), 321–337. MR Zbl

[Ancona 1987] A. Ancona, "Negatively curved manifolds, elliptic operators, and the Martin boundary", *Ann. of Math.* (2) **125**:3 (1987), 495–536. MR Zbl

[Ancona 2009] A. Ancona, "Elliptic operators, conormal derivatives and positive parts of functions", *J. Funct. Anal.* **257**:7 (2009), 2124–2158. MR Zbl

[Ancona 2012] A. Ancona, "Positive solutions of Schrödinger equations and fine regularity of boundary points", *Math. Z.* **272**:1-2 (2012), 405–427. Correction in **272**:1-2 (2012), 429. MR Zbl

[Anzellotti 1983] G. Anzellotti, "Pairings between measures and bounded functions and compensated compactness", *Ann. Mat. Pura Appl.* (4) **135** (1983), 293–318. MR Zbl

[Anzellotti 1984] G. Anzellotti, "On the extremal stress and displacement in Hencky plasticity", *Duke Math. J.* **51**:1 (1984), 133–147. MR Zbl

[Arcoya and Boccardo 2015] D. Arcoya and L. Boccardo, "Regularizing effect of the interplay between coefficients in some elliptic equations", *J. Funct. Anal.* **268**:5 (2015), 1153–1166. MR Zbl

[Bandle et al. 2008] C. Bandle, V. Moroz, and W. Reichel, "'Boundary blowup' type sub-solutions to semilinear elliptic equations with Hardy potential", *J. Lond. Math. Soc.* (2) **77**:2 (2008), 503–523. MR Zbl

[Bertsch and Rostamian 1985] M. Bertsch and R. Rostamian, "The principle of linearized stability for a class of degenerate diffusion equations", *J. Differential Equations* **57**:3 (1985), 373–405. MR Zbl

[Bertsch et al. 2015] M. Bertsch, F. Smarrazzo, and A. Tesei, "A note on the strong maximum principle", *J. Differential Equations* **259**:8 (2015), 4356–4375. MR Zbl

[Bourdaud 1991] G. Bourdaud, "Le calcul fonctionnel dans les espaces de Sobolev", *Invent. Math.* **104**:2 (1991), 435–446. MR Zbl

[Brezis and Cabré 1998] H. Brezis and X. Cabré, "Some simple nonlinear PDE's without solutions", *Boll. Unione Mat. Ital. Sez. B Artic. Ric. Mat.* (8) **1**:2 (1998), 223–262. MR Zbl

[Brezis and Ponce 2003] H. Brezis and A. C. Ponce, "Remarks on the strong maximum principle", *Differential Integral Equations* **16**:1 (2003), 1–12. MR Zbl

[Brezis and Ponce 2008] H. Brezis and A. C. Ponce, "Kato's inequality up to the boundary", *Commun. Contemp. Math.* **10**:6 (2008), 1217–1241. MR Zbl

[Brezis and Strauss 1973] H. Brezis and W. A. Strauss, "Semi-linear second-order elliptic equations in L^1", *J. Math. Soc. Japan* **25** (1973), 565–590. MR Zbl

[Brezis et al. 2007] H. Brezis, M. Marcus, and A. C. Ponce, "Nonlinear elliptic equations with measures revisited", pp. 55–109 in *Mathematical aspects of nonlinear dispersive equations*, edited by J. Bourgain et al., Ann. of Math. Stud. **163**, Princeton Univ. Press, 2007. MR Zbl

[Dal Maso 1983] G. Dal Maso, "On the integral representation of certain local functionals", *Ricerche Mat.* **32**:1 (1983), 85–113. MR Zbl

[Dal Maso and Mosco 1986] G. Dal Maso and U. Mosco, "Wiener criteria and energy decay for relaxed Dirichlet problems", *Arch. Rational Mech. Anal.* **95**:4 (1986), 345–387. MR Zbl

[Dal Maso and Mosco 1987] G. Dal Maso and U. Mosco, "Wiener's criterion and Γ-convergence", *Appl. Math. Optim.* **15**:1 (1987), 15–63. MR Zbl

[Dal Maso et al. 1999] G. Dal Maso, F. Murat, L. Orsina, and A. Prignet, "Renormalized solutions of elliptic equations with general measure data", *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **28**:4 (1999), 741–808. MR Zbl

[Devyver et al. 2014] B. Devyver, M. Fraas, and Y. Pinchover, "Optimal Hardy weight for second-order elliptic operator: an answer to a problem of Agmon", *J. Funct. Anal.* **266**:7 (2014), 4422–4489. MR Zbl

[Díaz 2015] J. I. Díaz, "On the ambiguous treatment of the Schrödinger equation for the infinite potential well and an alternative via flat solutions: the one-dimensional case", *Interfaces Free Bound.* **17**:3 (2015), 333–351. MR Zbl

[Díaz 2017] J. I. Díaz, "On the ambiguous treatment of the Schrödinger equation for the infinite potential well and an alternative via singular potentials: the multi-dimensional case", *SeMA J.* **74**:3 (2017), 255–278. Zbl

[Díaz et al. 2018] J. I. Díaz, D. Gómez-Castro, J. M. Rakotoson, and R. Temam, "Linear diffusion with singular absorption potential and/or unbounded convective flow: the weighted space approach", *Discrete Contin. Dyn. Syst.* **38**:2 (2018), 509–546. MR Zbl

[Dupaigne 2011] L. Dupaigne, *Stable solutions of elliptic partial differential equations*, Chapman & Hall/CRC Monographs and Surveys in Pure and Applied Mathematics **143**, Chapman & Hall/CRC, Boca Raton, FL, 2011. MR Zbl

[Evans 2010] L. C. Evans, *Partial differential equations*, 2nd ed., Graduate Studies in Mathematics **19**, Amer. Math. Soc., Providence, RI, 2010. MR Zbl

[Evans and Gariepy 2015] L. C. Evans and R. F. Gariepy, *Measure theory and fine properties of functions*, CRC Press, Boca Raton, FL, 2015. MR Zbl

[Feyel and de la Pradelle 1977] D. Feyel and A. de la Pradelle, "Topologies fines et compactifications associées à certains espaces de Dirichlet", *Ann. Inst. Fourier* (*Grenoble*) **27**:4 (1977), 121–146. MR Zbl

[Gilbarg and Trudinger 1998] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, Grundlehren der Mathematischen Wissenschaften **224**, Springer, 1998.

[Grun-Rehomme 1977] M. Grun-Rehomme, "Caractérisation du sous-différentiel d'intégrandes convexes dans les espaces de Sobolev", *J. Math. Pures Appl.* (9) **56**:2 (1977), 149–156. MR Zbl

[Kato 1972] T. Kato, "Schrödinger operators with singular potentials", *Israel J. Math.* **13** (1972), 135–148. MR Zbl

[Kohn and Temam 1983] R. Kohn and R. Temam, "Dual spaces of stresses and strains, with applications to Hencky plasticity", *Appl. Math. Optim.* **10**:1 (1983), 1–35. MR Zbl

[Lions and Magenes 1972] J.-L. Lions and E. Magenes, *Non-homogeneous boundary value problems and applications, I*, Die Grundlehren der Mathematischen Wissenschaften **181**, Springer, 1972. MR Zbl

[Littman et al. 1963] W. Littman, G. Stampacchia, and H. F. Weinberger, "Regular points for elliptic equations with discontinuous coefficients", *Ann. Scuola Norm. Sup. Pisa* (3) **17** (1963), 43–77. MR Zbl

[Marcus and Nguyen 2017] M. Marcus and P.-T. Nguyen, "Moderate solutions of semilinear elliptic equations with Hardy potential", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **34**:1 (2017), 69–88. MR Zbl

[Marcus and Véron 2014] M. Marcus and L. Véron, *Nonlinear second order elliptic equations involving measures*, De Gruyter Series in Nonlinear Analysis and Applications **21**, De Gruyter, Berlin, 2014. MR Zbl

[Orsina and Ponce 2008] L. Orsina and A. C. Ponce, "Semilinear elliptic equations and systems with diffuse measures", *J. Evol. Equ.* **8**:4 (2008), 781–812. MR Zbl

[Orsina and Ponce 2016] L. Orsina and A. C. Ponce, "Strong maximum principle for Schrödinger operators with singular potential", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **33**:2 (2016), 477–493. MR Zbl

[Pinchover 2007] Y. Pinchover, "Topics in the theory of positive solutions of second-order elliptic and parabolic partial differential equations", pp. 329–355 in *Spectral theory and mathematical physics: a Festschrift in honor of Barry Simon's 60th birthday*, edited by F. Gesztesy et al., Proc. Sympos. Pure Math. **76**, Amer. Math. Soc., Providence, RI, 2007. MR Zbl

[Pinchover and Psaradakis 2016] Y. Pinchover and G. Psaradakis, "On positive solutions of the (p, A)-Laplacian with potential in Morrey space", *Anal. PDE* **9**:6 (2016), 1317–1358. MR Zbl

[Pinchover and Tintarev 2005] Y. Pinchover and K. Tintarev, "Existence of minimizers for Schrödinger operators under domain perturbations with application to Hardy's inequality", *Indiana Univ. Math. J.* **54**:4 (2005), 1061–1074. MR Zbl

[Ponce 2016] A. C. Ponce, *Elliptic PDEs, measures and capacities: from the Poisson equations to nonlinear Thomas–Fermi problems*, EMS Tracts in Mathematics **23**, European Mathematical Society, Zürich, 2016. MR Zbl

[Ponce and Wilmet 2017] A. C. Ponce and N. Wilmet, "Schrödinger operators involving singular potentials and measure data", *J. Differential Equations* **263**:6 (2017), 3581–3610. MR Zbl

[Protter and Weinberger 1984] M. H. Protter and H. F. Weinberger, *Maximum principles in differential equations*, Springer, 1984. MR Zbl

[Stampacchia 1965] G. Stampacchia, "Le problème de Dirichlet pour les équations elliptiques du second ordre à coefficients discontinus", *Ann. Inst. Fourier* (*Grenoble*) **15**:1 (1965), 189–258. MR Zbl

[Trudinger 1978] N. S. Trudinger, "On the positivity of weak supersolutions of nonuniformly elliptic equations", *Bull. Austral. Math. Soc.* **19**:3 (1978), 321–324. MR Zbl

[Véron and Yarur 2012] L. Véron and C. Yarur, "Boundary value problems with measures for elliptic equations with singular potentials", *J. Funct. Anal.* **262**:3 (2012), 733–772. MR Zbl

[Zhao 1986] Z. X. Zhao, "Green function for Schrödinger operator and conditioned Feynman–Kac gauge", *J. Math. Anal. Appl.* **116**:2 (1986), 309–334. MR Zbl

Received 3 Jun 2017. Accepted 9 Apr 2018.

LUIGI ORSINA: orsina@mat.uniroma1.it
Dipartimento di Matematica, "Sapienza" Università di Roma, Roma, Italy

AUGUSTO C. PONCE: augusto.ponce@uclouvain.be
Institut de Recherche en Mathématique et Physique, Université catholique de Louvain, Louvain-la-Neuve, Belgium

ANALYSIS AND PDE
Vol. 11, No. 8, 2018
dx.doi.org/10.2140/apde.2018.11.2049

MONOTONICITY OF NONPLURIPOLAR PRODUCTS AND COMPLEX MONGE–AMPÈRE EQUATIONS WITH PRESCRIBED SINGULARITY

TAMÁS DARVAS, ELEONORA DI NEZZA AND CHINH H. LU

We establish the monotonicity property for the mass of nonpluripolar products on compact Kähler manifolds, and we initiate the study of complex Monge–Ampère-type equations with prescribed singularity type. Using the variational method of Berman, Boucksom, Guedj and Zeriahi we prove existence and uniqueness of solutions with small unbounded locus. We give applications to Kähler–Einstein metrics with prescribed singularity, and we show that the log-concavity property holds for nonpluripolar products with small unbounded locus.

1. Introduction and main results

Let X be a compact Kähler manifold of complex dimension n, and let θ be a smooth closed real $(1,1)$-form on X such that $\{\theta\}$ is big. Broadly speaking, the purpose of this article is threefold. First, we develop the potential theory of nonpluripolar products without any restrictions on the singularity type by combining techniques of Witt Nyström [2017] and previous work of the authors [Darvas et al. 2018]. Second, given $\phi \in \mathrm{PSH}(X,\theta)$, we introduce and study the spaces $\mathcal{E}(X,\theta,\phi)$ and $\mathcal{E}^1(X,\theta,\phi)$, generalizing the content of [Boucksom et al. 2010] to the relative framework. These latter spaces contain potentials that are slightly more singular than ϕ, and satisfy a (relative) full mass/finite energy condition. Lastly, with sufficient potential theory developed, we focus on the variational study of the complex Monge–Ampère equation

$$(\theta + i\partial\bar{\partial}u)^n = f\omega^n, \tag{1}$$

where $f \geq 0$, $f \in L^p(\omega^n)$, $p > 1$, and the singularity type of $u \in \mathrm{PSH}(X,\theta)$ is the same as that of ϕ. As it will turn out, this equation is well-posed only for potentials ϕ with a certain type of "model" singularity, which includes the case of analytic singularities, and we provide existence of unique solutions with small unbounded locus. As we will see, on the right-hand side of (1) one may even consider more general (nonpluripolar) Radon measures.

When θ is a Kähler form, $f > 0$ is smooth, and $\phi = 0$, the above equation was solved (with smooth solutions) by Yau [1978], see also [Aubin 1978], resolving the famous Calabi conjecture. Using both a priori estimates and pluripotential theory, this result was later extended in many different directions; see [Kołodziej 1998; 2003; Guedj and Zeriahi 2007; Boucksom et al. 2010; Berman et al. 2013; Berman

MSC2010: primary 32Q15, 32U05, 32W20; secondary 32Q20.
Keywords: Monge–Ampère equation, variational approach, pluripotential theory.

2013; Phong and Sturm 2014]. Our approach seems to unify all existing works (in the compact setting), under the theme of solutions with arbitrary prescribed (model) singularity type.

At the end of the paper, we give applications of our results to singular Kähler–Einstein metrics and establish the log-concavity property for certain nonpluripolar products. Other applications will be treated in a sequel.

Though we will work in the general framework of big cohomology classes throughout the paper, we note that all our results seem to be new in the particular case of Kähler classes as well.

Monotonicity of nonpluripolar products and relative finite energy. Unless otherwise specified, we fix a background Kähler structure (X, ω) for the remainder of the paper.

We say that a potential $u \in L^1(X, \omega^n)$ is θ-plurisubharmonic (θ-psh) if locally u is the difference of a plurisubharmonic and a smooth function, and $\theta_u := \theta + i\partial\bar{\partial}u \geq 0$ in the sense of currents. The set of θ-psh potentials is denoted by $\mathrm{PSH}(X, \theta)$. We say that $\{\theta\}$ is *pseudoeffective* if $\mathrm{PSH}(X, \theta)$ is nonempty. Along these lines, $\{\theta\}$ is *big* if $\mathrm{PSH}(X, \theta - \varepsilon\omega)$ is nonempty for some $\varepsilon > 0$.

If u and v are two θ-psh functions on X, then u is said to be *less singular* than v if $v \leq u + C$ for some $C \in \mathbb{R}$. We say that u has the same singularities as v if u is less singular than v, and v is less singular than u. This defines an equivalence relation on $\mathrm{PSH}(X, \theta)$ whose equivalence classes are the *singularity types* $[u]$, $u \in \mathrm{PSH}(X, \theta)$.

Given closed positive $(1, 1)$-currents $T_1 := \theta^1_{u_1}, \ldots, T_p := \theta^p_{u_p}$, where the θ^j are closed smooth real $(1, 1)$-forms, generalizing the construction of [Bedford and Taylor 1987] in the local setting, it was shown in [Boucksom et al. 2010] that one can define the *nonpluripolar product* of these currents:

$$\theta^1_{u_1} \wedge \cdots \wedge \theta^p_{u_p} := \langle T_1 \wedge \cdots \wedge T_p \rangle.$$

The resulting positive (p, p)-current does not charge pluripolar sets and it is *closed*. For a θ-psh function u, the *nonpluripolar complex Monge–Ampère measure* of u is simply $\theta^n_u := \theta_u \wedge \cdots \wedge \theta_u$.

It was recently proved by Witt Nyström [2017, Theorem 1.2] that the complex Monge–Ampère mass of θ-psh potentials decreases as the singularity type increases. Our main result about monotonicity of nonpluripolar products generalizes this result to the case of different cohomology classes $\{\theta^j\}$, fully proving what was conjectured by Boucksom, Eyssidieux, Guedj and Zeriahi (see the comments after [Boucksom et al. 2010, Theorem 1.16] in which they prove that the result holds for potentials with small unbounded locus):

Theorem 1.1. *Let θ^j, $j \in \{1, \ldots, n\}$, be smooth closed real $(1, 1)$-forms on X. Let $u_j, v_j \in \mathrm{PSH}(X, \theta^j)$ such that u_j is less singular than v_j for all $j \in \{1, \ldots, n\}$. Then*

$$\int_X \theta^1_{u_1} \wedge \cdots \wedge \theta^n_{u_n} \geq \int_X \theta^1_{v_1} \wedge \cdots \wedge \theta^n_{v_n}.$$

To prove the above theorem, we first need to generalize the main convergence theorems of Bedford–Taylor theory [1987]; see also [Xing 1996; 2009]. This is done collectively in the next result, further elaborated in Theorem 2.3 below:

Theorem 1.2. *Let θ^j, $j \in \{1, \dots, n\}$, be smooth closed real $(1,1)$-forms on X. Suppose that we have $u_j, u_j^k \in \mathrm{PSH}(X, \theta^j)$ such that $u_j^k \to u_j$ in capacity as $k \to \infty$, and*

$$\int_X \theta_{u_1}^1 \wedge \cdots \wedge \theta_{u_n}^n \geq \limsup_{k\to\infty} \int_X \theta_{u_1^k}^1 \wedge \cdots \wedge \theta_{u_n^k}^n. \tag{2}$$

Then $\theta_{u_1^k}^1 \wedge \cdots \wedge \theta_{u_n^k}^n \to \theta_{u_1}^1 \wedge \cdots \wedge \theta_{u_n}^n$ in the weak sense of measures.

We recall that a sequence $\{u_k\}_k$ converges in capacity to u if for any $\delta > 0$ we have

$$\lim_{k\to\infty} \mathrm{Cap}_\omega\{|u_k - u| \geq \delta\} = 0,$$

where Cap_ω is the Monge–Ampère capacity associated to ω; see [Guedj and Zeriahi 2017, Definition 4.23].

We note that condition (2) is necessary in this generality, even in the Kähler case. Indeed, if $u \in \mathrm{PSH}(X, \omega)$ is a pluricomplex Green potential, then the cut-offs $u_j := \max(u, -j) \in \mathrm{PSH}(X, \omega)$ satisfy $u_j \searrow u$. However, $\int_X \omega_{u_j}^n = \int_X \omega^n > 0$ for all j, and $\int_X \omega_u^n = 0$; hence $\omega_{u_j}^n$ cannot converge to ω_u^n weakly.

As noted above, Theorem 1.2 generalizes classical theorems of Bedford and Taylor (when u_j^k, u_j are uniformly bounded) and also results from [Boucksom et al. 2010] (when u_j^k, u_j have full mass). In both of these cases, there are severe restrictions on the singularity class of the potentials u_j^k, u_j. On the other hand, the above theorem shows that there is no need for restrictions on singularity type of the potentials involved. Instead, one needs only a semicontinuity condition on the total masses.

To develop the variational approach to (1), with the above general results in hand, we initiate the study of relative full mass/relative finite energy currents. Let $\phi \in \mathrm{PSH}(X, \theta)$. We say that $v \in \mathrm{PSH}(X, \theta)$ has *full mass relative* to ϕ ($v \in \mathcal{E}(X, \theta, \phi)$) if v is more singular than ϕ and $\int_X \theta_v^n = \int_X \theta_\phi^n$. In our investigation of these classes, the following well-known envelope constructions will be of great help:

$$\psi \quad \to \quad P_\theta(\psi, \phi),\ P_\theta[\psi](\phi),\ P_\theta[\psi] \in \mathrm{PSH}(X, \theta) \quad \text{where } \psi \in \mathrm{PSH}(X, \theta).$$

These were introduced by Ross and Witt Nyström [2014] in their construction of geodesic rays, building on ideas of Rashkovskii and Sigurdsson [2005] in the local setting. Due to the frequency of these operators appearing in this work, we choose to follow slightly different notations. The starting point is the "rooftop envelope"

$$P_\theta(\psi, \phi) := \sup\{v \in \mathrm{PSH}(X, \theta) \mid v \leq \min(\psi, \phi)\}.$$

This allows us to introduce

$$P_\theta[\psi](\phi) := \big(\lim_{C\to\infty} P_\theta(\psi + C, \phi)\big)^*,$$

and it is easy to see that $P_\theta[\psi](\phi)$ only depends on the singularity type of ψ. When $\phi = 0$ or $\phi = V_\theta$, we will simply write $P_\theta[\psi] := P_\theta[\psi](0) = P_\theta[\psi](V_\theta)$, and we refer to this potential as the *envelope of the singularity type* $[\psi]$.

Using the techniques of our recent work [Darvas et al. 2018], we can give a generalization of [Darvas 2017, Theorem 3], paralleling [Darvas et al. 2018, Theorem 1.2]. This result characterizes membership in $\mathcal{E}(X, \theta, \phi)$ solely in terms of singularity type:

Theorem 1.3. *Suppose $\phi \in \mathrm{PSH}(X,\theta)$ and $\int_X \theta_\phi^n > 0$. The following are equivalent*:

(i) $u \in \mathcal{E}(X,\theta,\phi)$.

(ii) *ϕ is less singular than u, and $P_\theta[u](\phi) = \phi$.*

(iii) *ϕ is less singular than u, and $P_\theta[u] = P_\theta[\phi]$.*

Without the *nonzero mass* condition $\int_X \theta_\phi^n > 0$ this characterization cannot hold (see Remark 3.3). The equivalence between (i) and (iii) in the above theorem shows that $P_\theta[u]$ is the same potential for any $u \in \mathcal{E}(X,\theta,\phi)$, and is equal to $P_\theta[\phi]$. Given this and the inclusion $\mathcal{E}(X,\theta,\phi) \subset \mathcal{E}(X,\theta,P_\theta[\phi])$, one is tempted to consider only potentials ϕ in the image of the operator $\psi \to P_\theta[\psi]$, when studying the classes of relative full mass $\mathcal{E}(X,\theta,\phi)$. These potentials seemingly play the same role as V_θ, the potential with minimal singularities from [Boucksom et al. 2010]. Implementation of this idea will be further motivated by the results of the next subsection.

In addition to the above result, we also establish analogs of many classical results for $\mathcal{E}(X,\theta,\phi)$, like the comparison, maximum and domination principles. Some of these are routine, while others, like the domination principle, require new techniques and a more involved analysis compared to the existing literature (see Proposition 3.11).

Complex Monge–Ampère equations with prescribed singularity. With the potential theoretic tools developed, we focus on solving (1). A simple minded example shows that this equation is not well-posed for arbitrary $\phi \in \mathrm{PSH}(X,\theta)$ (see the introduction of Section 4). Instead, one needs to consider only potentials ϕ that are fixed points of the operator $\psi \to P_\theta[\psi]$, i.e., $\psi = P_\theta[\psi]$. Such potentials ψ will be called *model potentials*, and their singularity types $[\psi]$ will be called *model-type singularities*. In this direction we have the following result:

Theorem 1.4. *Suppose $\phi \in \mathrm{PSH}(X,\theta)$ has small unbounded locus, and $\phi = P_\theta[\phi]$. Let $f \in L^p(\omega^n)$, $p > 1$ such that $f \geq 0$ and $\int_X f\omega^n = \int_X \theta_\phi^n > 0$. Then the following hold*:

(i) *There exists $u \in \mathrm{PSH}(X,\theta)$, unique up to a constant, such that $[u] = [\phi]$ and*

$$\theta_u^n = f\omega^n. \tag{3}$$

(ii) *For any $\lambda > 0$ there exists a unique $v \in \mathrm{PSH}(X,\theta)$ such that $[v] = [\phi]$ and*

$$\theta_v^n = e^{\lambda v} f\omega^n. \tag{4}$$

That ϕ has *small unbounded locus* means that ϕ is locally bounded outside a closed complete pluripolar set $A \subset X$. It will be interesting to see if this condition is simply technical, or otherwise necessary. This seemingly extra condition on ϕ does have some benefits. Indeed, since in this setting solutions are locally bounded on $X \setminus A$, one can interpret (3) and (4) in the following simple way: u and v satisfy (3) and (4) on $X \setminus A$, in the sense of Bedford and Taylor.

Remark 1.5. As argued in Theorem 4.34, if (3) can be solved for all $f \in L^p(X)$, $p > 1$, (with the constraint $[u] = [\phi]$) then ϕ *must have* model-type singularity. Consequently, our choice of ϕ in the above theorem is not ad hoc, but truly natural!

In our study of the above equations, we will start with a much more general context. In particular, we will show in Theorems 4.28 and 4.23 below that instead of $f\omega^n$, one can consider, on the right-hand side of (3) and (4), nonpluripolar measures, thereby generalizing [Boucksom et al. 2010, Theorems A, D].

Remark 1.6. Naturally, $V_\theta = P_\theta[V_\theta]$, but our reader may wonder if there are other interesting enough potentials with model-type singularity. We believe this to be the case, as evidenced below:

- By Theorem 3.12 below, $P_\theta[\psi] = P_\theta[P_\theta[\psi]]$ for any $\psi \in \mathrm{PSH}(X,\theta)$ with $\int_X \theta_\psi^n > 0$. In particular, $P_\theta[\psi]$ is a model potential, giving an abundance of potentials with model-type singularity.
- By Proposition 4.35 below, if $\psi \in \mathrm{PSH}(X,\theta)$ has small unbounded locus, and $\theta_\psi^n/\omega^n \in L^p(\omega^n)$, $p > 1$, with $\int_X \theta_\psi^n > 0$, then ψ has model-type singularity.
- All *analytic singularity types* (those that can be locally written as $c\log\big(\sum_j |f_j|^2\big) + g$, where the f_j are holomorphic, $c > 0$ and g is smooth) are of model type [Ross and Witt Nyström 2014, Remark 4.6; Rashkovskii and Sigurdsson 2005]; see also Proposition 4.36. In particular, discrete logarithmic singularity types are of model type, making a connection with pluricomplex Green currents [Coman and Guedj 2009; Phong and Sturm 2014; Rashkovskii and Sigurdsson 2005].
- By [Ross and Witt Nyström 2014; Darvas 2017; Darvas et al. 2018], potentials with model-type singularity naturally arise as degenerations along geodesic rays and in particular along test configurations.

Complex Monge–Ampère equations with bounded/minimally singular solutions have been intensely studied in the past; see [Kołodziej 1998; 2003; Guedj and Zeriahi 2007; Boucksom et al. 2010; Berman et al. 2013], to name only a few works in a fast expanding literature. To our knowledge, in the compact case, only [Phong and Sturm 2014] discusses at length solutions that are not "minimally singular", without severe restrictions on the right-hand side of the equation. They treat the case of solutions to (3) with isolated algebraic singularities in the Kähler case, with a view toward constructing pluricomplex Green currents on X. Given the specific setting, [Phong and Sturm 2014, Theorem 3] obtains more precise regularity estimates compared to ours, using blowup techniques. In our general framework better estimates are likely not possible. However, for smooth f, we suspect that away from the singularity locus our solution u should be as regular as ϕ (up to order 2). For a general result on the regularity of certain model potentials we refer to [Ross and Witt Nyström 2017, Theorem 1.1].

Lastly, let us mention that in [Berman 2013, Section 4] solutions to complex Monge–Ampère equations with divisorial singularity type are used in the construction/approximation of geodesic rays corresponding to certain test configurations. In Section 5 of the same work, Berman speculates that solutions with more general singularity type should allow for better understanding of degenerations along test configurations/geodesic rays, and we believe our treatise will lead to more results of this flavor.

In addition to the results in the compact setting mentioned above, finding singular/unbounded solutions to the related Dirichlet problem on domains in $\mathbb{C}^n$, or more generally on compact manifolds with boundary, was studied by a number of authors. We only mention [Lempert 1983; Bedford and Demailly 1988; Guan 1998; Phong and Sturm 2010a; 2010b] to highlight a few works in a fast expanding literature.

Applications. Solutions of complex Monge–Ampère equations are linked to existence of special Kähler metrics. In particular, we can think of the solution to (3) as a potential with prescribed singularity type and prescribed Ricci curvature in the philosophy of the Calabi–Yau theorem. As an immediate application of our solution to (4) we obtain existence of singular *Kähler–Einstein* (KE) metrics with prescribed singularity type on Kähler manifolds of general type. An analogous result also holds on Calabi–Yau manifolds as well, via solutions of (3).

Corollary 1.7. *Let X be a smooth projective variety of general type ($K_X > 0$) and let h be a smooth Hermitian metric on K_X with $\theta := \Theta(h) > 0$. Suppose also that $\phi \in \mathrm{PSH}(X,\theta)$ is a model potential, has small unbounded locus and $\int_X \theta_\phi^n > 0$. Then there exists a unique singular KE metric $he^{-\phi_{\mathrm{KE}}}$ on K_X ($\theta_{\phi_{\mathrm{KE}}}^n = e^{\phi_{\mathrm{KE}}+f_\theta}\theta^n$, where f_θ is the Ricci potential of θ satisfying $\mathrm{Ric}\,\theta = \theta + i\partial\bar\partial f_\theta$), with $\phi_{\mathrm{KE}} \in \mathrm{PSH}(X,\theta)$ having the same singularity type as ϕ.*

As another application we confirm the log-concavity conjecture [Boucksom et al. 2010, Conjecture 1.23] in the case of currents with potentials having small unbounded locus:

Theorem 1.8. *Let $T_1,\dots,T_n$ be positive closed $(1,1)$-currents on a compact Kähler manifold X. Assume that each T_j has a potential with small unbounded locus. Then*

$$\int_X \langle T_1 \wedge \cdots \wedge T_n\rangle \geq \left(\int_X \langle T_1^n\rangle\right)^{1/n} \cdots \left(\int_X \langle T_n^n\rangle\right)^{1/n}.$$

Possible future directions. It is well known that, for $\lambda < 0$, (4) does not always have a solution. More importantly, solvability of this equation is tied together with existence of KE metrics on Fano manifolds. It would be interesting to see if the techniques of [Darvas and Rubinstein 2017] apply to give characterizations for existence of KE metrics with prescribed singularity type in terms of energy properness.

By [Darvas 2017; Darvas et al. 2018] the geometry of geodesic rays and properties of (relative) full mass potentials seems to be intimately related. In a future work we will explore this avenue further, by introducing a metric geometry on the space of singularity types, via the constructions of [Darvas 2017; Darvas et al. 2018]. By understanding the metric properties of this space, we hope to study degenerations of singularity types along complex Monge–Ampère equations.

Organization of the paper. Most of our notation and terminology carries over from [Darvas et al. 2018], and we refer the reader to the introductory sections of this work. In Section 2 we prove Theorems 1.1 and 1.2. In Section 3 we develop the theory of the relative full mass classes $\mathcal{E}(X,\theta,\phi)$ and we exploit properties of envelopes to prove Theorem 1.3. In Section 4 we generalize the variational methods of [Berman et al. 2013] to prove Theorem 1.4. Finally, Theorem 1.8 is proved in Section 5.

2. The monotonicity property and convergence of nonpluripolar products

To begin, from the main result of [Witt Nyström 2017] we deduce the following proposition:

Proposition 2.1. *Let θ^j, $j \in \{1, \dots, n\}$, be smooth closed real $(1,1)$-forms on X whose cohomology classes are pseudoeffective. Let $u_j, v_j \in \mathrm{PSH}(X, \theta^j)$ be such that u_j has the same singularity type as v_j, $j \in \{1, \dots, n\}$. Then*

$$\int_X \theta^1_{u_1} \wedge \cdots \wedge \theta^n_{u_n} = \int_X \theta^1_{v_1} \wedge \cdots \wedge \theta^n_{v_n}.$$

The proof of this result uses the arguments in [Boucksom et al. 2010, Corollary 2.15].

Proof. First we note that we can assume that the classes $\{\theta^j\}$ are in fact big. Indeed, if this is not the case we can just replace each θ^j with $\theta^j + \varepsilon\omega$, and using the multilinearity of the nonpluripolar product [Boucksom et al. 2010, Proposition 1.4] we can let $\varepsilon \to 0$ at the end of our argument to conclude the statement for pseudoeffective classes.

For each $t \in \Delta = \{t = (t_1, \dots, t_n) \in \mathbb{R}^n \mid t_j > 0\}$ consider $u_t := \sum_j t_j u_j$, $v_t := \sum_j t_j v_j$ and $\theta^t := \sum_j t_j \theta^j$. Clearly, $\{\theta^t\}$ is big, and u_t has the same singularities as v_t. Hence it follows from [Witt Nyström 2017, Theorem 1.2] that $\int_X (\theta^t_{u_t})^n = \int_X (\theta^t_{v_t})^n$ for all $t \in \Delta$. On the other hand, using multilinearity of the nonpluripolar product again [Boucksom et al. 2010, Proposition 1.4], we see that both $t \to \int_X (\theta^t_{u_t})^n$ and $t \to \int_X (\theta^t_{v_t})^n$ are homogeneous polynomials of degree n. Our last identity forces all the coefficients of these polynomials to be equal, giving the statement of our result. □

We recall a classical convergence theorem from Bedford–Taylor theory. We refer to [Guedj and Zeriahi 2017, Theorem 4.26] for a proof of this result, which is merely a slight generalization of [Xing 1996, Theorem 1].

Proposition 2.2. *Let $\Omega \subset \mathbb{C}^n$ be an open set. Suppose $\{f_j\}_j$ are uniformly bounded quasicontinuous functions which converge in capacity to another quasicontinuous function f on Ω. Let $\{u_1^j\}_j, \{u_2^j\}_j, \dots, \{u_n^j\}_j$ be uniformly bounded plurisubharmonic functions on Ω, converging in capacity to $u_1, u_2, \dots, u_n$ respectively. Then we have the following weak convergence of measures*:

$$f_j i\partial\bar\partial u_1^j \wedge i\partial\bar\partial u_2^j \wedge \cdots \wedge i\partial\bar\partial u_n^j \to f i\partial\bar\partial u_1 \wedge i\partial\bar\partial u_2 \wedge \cdots \wedge i\partial\bar\partial u_n.$$

The following lower-semicontinuity property of nonpluripolar products will be key in the sequel:

Theorem 2.3. *Let θ^j, $j \in \{1, \dots, n\}$, be smooth closed real $(1,1)$-forms on X whose cohomology classes are big. Suppose that for all $j \in \{1, \dots, n\}$ we have $u_j, u_j^k \in \mathrm{PSH}(X, \theta^j)$ such that $u_j^k \to u_j$ in capacity as $k \to \infty$. Then for all positive bounded quasicontinuous functions χ we have*

$$\liminf_{k\to\infty} \int_X \chi \theta^1_{u_1^k} \wedge \cdots \wedge \theta^n_{u_n^k} \geq \int_X \chi \theta^1_{u_1} \wedge \cdots \wedge \theta^n_{u_n}.$$

If additionally,

$$\int_X \theta^1_{u_1} \wedge \cdots \wedge \theta^n_{u_n} \geq \limsup_{k\to\infty} \int_X \theta^1_{u_1^k} \wedge \cdots \wedge \theta^n_{u_n^k}, \tag{5}$$

then $\theta^1_{u_1^k} \wedge \cdots \wedge \theta^n_{u_n^k} \to \theta^1_{u_1} \wedge \cdots \wedge \theta^n_{u_n}$ in the weak sense of measures on X.

Proof. Set $\Omega := \bigcap_{j=1}^n \mathrm{Amp}(\theta^j)$ and fix an open relatively compact subset U of Ω. Then the functions V_{θ^j} are bounded on U. We now use a classical idea in pluripotential theory. Fix $C > 0$, $\varepsilon > 0$ and consider

$$f_j^{k,C,\varepsilon} := \frac{\max(u_j^k - V_{\theta^j} + C, 0)}{\max(u_j^k - V_{\theta^j} + C, 0) + \varepsilon}, \quad j = 1, \dots, n, \quad k \in \mathbb{N}^*,$$

and

$$u_j^{k,C} := \max(u_j^k, V_{\theta^j} - C).$$

Observe that for C, j fixed, the functions $u_j^{k,C} \geq V_{\theta^j} - C$ are uniformly bounded in U (since V_{θ^j} is bounded in U) and converge in capacity to u_j^C as $k \to \infty$. Moreover, $f_j^{k,C,\varepsilon} = 0$ if $u_j^k \leq V_{\theta^j} - C$. By locality of the nonpluripolar product we can write

$$f^{k,C,\varepsilon} \chi\theta_{u_1^k}^1 \wedge \cdots \wedge \theta_{u_n^k}^n = f^{k,C,\varepsilon} \chi\theta_{u_1^{k,C}}^1 \wedge \cdots \wedge \theta_{u_n^{k,C}}^n,$$

where $f^{k,C,\varepsilon} = f_1^{k,C,\varepsilon} \cdots f_n^{k,C,\varepsilon}$. For each fixed C, ε, the functions $f^{k,C,\varepsilon}$ are quasicontinuous, uniformly bounded (with values in $[0, 1]$) and converge in capacity to $f^{C,\varepsilon} := f_1^{C,\varepsilon} \cdots f_n^{C,\varepsilon}$, where $f_j^{C,\varepsilon}$ is defined by

$$f_j^{C,\varepsilon} := \frac{\max(u_j - V_{\theta^j} + C, 0)}{\max(u_j - V_{\theta^j} + C, 0) + \varepsilon}.$$

With the information above we can apply Proposition 2.2 to get

$$f^{k,C,\varepsilon} \chi\theta_{u_1^{k,C}}^1 \wedge \cdots \wedge \theta_{u_n^{k,C}}^n \to f^{C,\varepsilon} \chi\theta_{u_1^C}^1 \wedge \cdots \wedge \theta_{u_n^C}^n \quad \text{as } k \to \infty,$$

in the weak sense of measures on U. In particular since $0 \leq f^{k,C,\varepsilon} \leq 1$ we have

$$\liminf_{k\to\infty} \int_X \chi\theta_{u_1^k}^1 \wedge \cdots \wedge \theta_{u_n^k}^n \geq \liminf_{k\to\infty} \int_U f^{k,C,\varepsilon} \chi\theta_{u_1^{k,C}}^1 \wedge \cdots \wedge \theta_{u_n^{k,C}}^n \geq \int_U f^{C,\varepsilon} \chi\theta_{u_1^C}^1 \wedge \cdots \wedge \theta_{u_n^C}^n.$$

Now, letting $\varepsilon \to 0$ and then $C \to \infty$, by definition of the nonpluripolar product we obtain

$$\liminf_{k\to\infty} \int_X \chi\theta_{u_1^k}^1 \wedge \cdots \wedge \theta_{u_n^k}^n \geq \int_U \chi\theta_{u_1}^1 \wedge \cdots \wedge \theta_{u_n}^n.$$

Finally, letting U increase to Ω and noting that the complement of Ω is pluripolar we conclude the proof of the first statement of the theorem.

To prove the last statement, we set $\mu_k := \theta_{u_1^k}^1 \wedge \cdots \wedge \theta_{u_n^k}^n$ and $\mu := \theta_{u_1}^1 \wedge \cdots \wedge \theta_{u_n}^n$. Note that the total mass of these measures is bounded by $\int_X \theta^1 \wedge \cdots \wedge \theta^n$ [Boucksom et al. 2010, Definition 1.17]. As a result, by the Banach–Alaoglu theorem, it suffices to show that any cluster point of $\{\mu_k\}_k$ coincides with μ. Let ν be such a cluster point and assume (after extracting a subsequence) that μ_k converges weakly to ν. Condition (5) implies that $\nu(X) \leq \mu(X)$. It suffices to argue that $\nu \geq \mu$, which is a consequence of the first statement, thus finishing the proof. □

Now we move on to the monotonicity of nonpluripolar products:

Theorem 2.4. *Let θ^j, $j \in \{1, \dots, n\}$, be smooth closed real $(1,1)$-forms on X whose cohomology classes are pseudoeffective. Let $u_j, v_j \in \mathrm{PSH}(X, \theta^j)$ be such that u_j is less singular than v_j for all $j \in \{1, \dots, n\}$. Then*

$$\int_X \theta_{u_1}^1 \wedge \cdots \wedge \theta_{u_n}^n \geq \int_X \theta_{v_1}^1 \wedge \cdots \wedge \theta_{v_n}^n.$$

Proof. By the same reason as in Proposition 2.1, we can assume that the classes $\{\theta^j\}$ are in fact big. For each $t > 0$ we set $v_j^t := \max(u_j - t, v_j)$ for $j = 1, \dots, n$. Observe that the v_j^t converge decreasingly to v_j as $t \to \infty$. In particular, by [Guedj and Zeriahi 2005, Proposition 3.7] the convergence holds in capacity. As v_j^t and u_j have the same singularity type, it follows from Proposition 2.1 that

$$\int_X \theta_{u_1}^1 \wedge \cdots \wedge \theta_{u_n}^n = \int_X \theta_{v_1^t}^1 \wedge \cdots \wedge \theta_{v_n^t}^n.$$

Letting $t \to \infty$, the first part of Theorem 2.3 allows us to conclude the argument. □

Remark 2.5. We note that condition (5) in Theorem 2.3 is automatically satisfied if $u_j^k \nearrow u_j$ a.e. as $k \to \infty$. Indeed, in this case $u_j^k \to u_j$ in capacity, see [Guedj and Zeriahi 2017, Proposition 4.25], and by Theorem 2.4 we have $\int_X \theta_{u_1}^1 \wedge \cdots \wedge \theta_{u_n}^n \geq \limsup_k \int_X \theta_{u_1^k}^1 \wedge \cdots \wedge \theta_{u_n^k}^n$.

On the other hand, if $u_j^k, u_j \in \mathcal{E}(X, \theta^j)$, by Corollary 3.2 below, it follows that (5) is again automatically satisfied. Moreover, in the next section we will show that this last property holds for potentials of relative full mass as well (see Corollary 3.15), giving Theorem 2.4 a more broad spectrum of applications.

3. Pluripotential theory with relative full mass

3A. ***Nonpluripolar products of relative full mass.*** Suppose θ^j, $j \in \{1, \dots, n\}$, are smooth closed real $(1,1)$-forms on X with $\{\theta^j\}$ pseudoeffective. Let $\phi_j, \psi_j \in \mathrm{PSH}(X, \theta^j)$ be such that ϕ_j is less singular than ψ_j. We say that $\theta_{\psi_1}^1 \wedge \cdots \wedge \theta_{\psi_n}^n$ has full mass with respect to $\theta_{\phi_1}^1 \wedge \cdots \wedge \theta_{\phi_n}^n$, denoted as $(\psi_1, \dots, \psi_n) \in \mathcal{E}(X, \theta_{\phi_1}^1, \dots, \theta_{\phi_n}^n)$, if

$$\int_X \theta_{\psi_1}^1 \wedge \cdots \wedge \theta_{\psi_n}^n = \int_X \theta_{\phi_1}^1 \wedge \cdots \wedge \theta_{\phi_n}^n.$$

By Theorem 2.4, in general we only have that the left-hand side is less than the right-hand side in the above identity.

In the particular case when the potentials involved are from the same cohomology class $\{\theta\}$, and $\phi, \psi \in \mathrm{PSH}(X, \theta)$ with ϕ less singular than ψ along with $\int_X \theta_\phi^n = \int_X \theta_\psi^n$, we simply write $\psi \in \mathcal{E}(X, \theta, \phi)$, and say that ψ has *full mass relative to* θ_ϕ^n. When $\phi = V_\theta$, we recover the well-known concept of full mass currents from the literature; see [Boucksom et al. 2010].

As a consequence of Theorem 2.3, we prove a criterion for testing membership in $\mathcal{E}(X, \theta_{\phi_1}^1, \dots, \theta_{\phi_n}^n)$:

Proposition 3.1. *Let θ^j, $j \in \{1, \dots, n\}$, be smooth closed real $(1,1)$-forms on X with cohomology classes that are pseudoeffective. For all $j \in \{1, \dots, n\}$ we choose $\phi_j, \psi_j \in \mathrm{PSH}(X, \theta^j)$ such that ϕ_j is less singular than ψ_j. If $P_{\theta^j}[\psi_j](\phi_j) = \phi_j$ then $(\psi_1, \dots, \psi_n) \in \mathcal{E}(X, \theta_{\phi_1}^1, \dots, \theta_{\phi_n}^n)$.*

Proof. If $P_{\theta^j}[\psi_j](\phi_j) = \phi_j$, then $v_j^C := P_{\theta^j}(\psi_j + C, \phi_j) \nearrow \phi_j$ a.e. as $C \to \infty$. Theorem 2.3 and Remark 2.5 then imply

$$\lim_{C \to \infty} \int_X \theta_{v_1^C} \wedge \cdots \wedge \theta_{v_n^C} = \int_X \theta_{\phi_1} \wedge \cdots \wedge \theta_{\phi_n}.$$

As $P_{\theta^j}(\psi_j + C, \phi_j)$ has the same singularity type as ψ_j for any C, the result follows from Proposition 2.1. □

As a result of this simple criterion, we obtain that condition (5) in Theorem 2.3 is satisfied if the potentials u_j^k, u_j are from $\mathcal{E}(X,\theta^j)$:

Corollary 3.2. *Let θ^j, $j\in\{1,\dots,n\}$, be smooth closed real $(1,1)$-forms on X with cohomology classes that are pseudoeffective. If $\psi_j\in\mathcal{E}(X,\theta^j)$, $j\in\{1,\dots,n\}$, then*

$$\int_X \theta_{\psi_1}^1\wedge\cdots\wedge\theta_{\psi_n}^n=\int_X \theta_{V_{\theta^1}}^1\wedge\cdots\wedge\theta_{V_{\theta^n}}^n,$$

or equivalently, $(\psi_1,\dots,\psi_n)\in\mathcal{E}(X,\theta_{V_{\theta^1}}^1,\dots,\theta_{V_{\theta^n}}^n)$.

Proof. By [Darvas et al. 2018, Theorem 1.2] we have $P_{\theta^j}[\psi_j]:=P_{\theta^j}[\psi_j](V_{\theta^j})=V_{\theta^j}$. Hence Proposition 3.1 yields the conclusion. □

Remark 3.3. Unfortunately, the reverse direction in Proposition 3.1 does not hold in general. Indeed, let $X=\mathbb{CP}^1\times\mathbb{CP}^1$ with $\theta=\pi_1^*\omega_{FS}+\pi_2^\star\omega_{FS}$, where π_1,π_2 are the projections to the first and second components respectively.

Consider $\phi(z,w):=u(z)+v(w)\in\mathrm{PSH}(X,\theta)$, where $u,v\le 0$ satisfy $\omega_{FS}+i\partial\bar\partial u=\delta_{z_0}$ and $\omega_{FS}+i\partial\bar\partial v=\delta_{w_0}$, where $\delta_{z_0},\delta_{w_0}$ are Dirac masses for some $z_0,w_0\in\mathbb{CP}^1$. Clearly, $\int_X\theta_\phi^2=\int_X\theta_{\pi_2^*v}^2=0$, and since $\phi\le\pi_2^*v$, we have $\phi\in\mathcal{E}(X,\theta,\pi_2^*v)$.

On the other hand, we know that ϕ has the same Lelong number as $P_\theta[\phi]$ [Darvas et al. 2018, Theorem 1.1]. As $P_\theta[\phi](\pi_2^*v)\le P_\theta[\phi]$, it follows however that $P_\theta[\phi](\pi_2^*v)\lneq\pi_2^*v$, since at some points of $\mathbb{CP}^1\times\mathbb{CP}^1$ the Lelong number of π_2^*v is zero, but the Lelong number of ϕ is nonzero.

As we will see below (Theorem 3.14), a partial converse of Proposition 3.1 is still possible under the assumption of nonvanishing total mass.

In the remaining part of this subsection we prove basic properties of nonpluripolar products with relative full mass, which will be used later in this work.

Lemma 3.4. *Suppose $\phi_j,\psi_j\in\mathrm{PSH}(X,\theta^j)$. Then $(\psi_1,\dots,\psi_n)\in\mathcal{E}(X,\theta_{\phi_1}^1,\dots,\theta_{\phi_n}^n)$ if and only if ϕ_j is less singular than ψ_j and*

$$\int_{\bigcup_j\{\psi_j\le\phi_j-k\}}\theta_{\max(\psi_1,\phi_1-k)}^1\wedge\cdots\wedge\theta_{\max(\psi_n,\phi_n-k)}^n\to 0\quad\text{as } k\to\infty.$$

Proof. If ϕ_j is less singular than ψ_j, then $\max(\psi_j,\phi_j-k)$ has the same singularity type as ϕ_j. Consequently, Proposition 2.1 gives

$$\begin{aligned}\int_X\theta_{\phi_1}^1\wedge\cdots\wedge\theta_{\phi_n}^n&=\int_X\theta_{\max(\psi_1,\phi_1-k)}^1\wedge\cdots\wedge\theta_{\max(\psi_n,\phi_n-k)}^n\\&=\int_{\bigcap_j\{\psi_j>\phi_j-k\}}\theta_{\psi_1}^1\wedge\cdots\wedge\theta_{\psi_n}^n+\int_{\bigcup_j\{\psi_j\le\phi_j-k\}}\theta_{\max(\psi_1,\phi_1-k)}^1\wedge\cdots\wedge\theta_{\max(\psi_n,\phi_n-k)}^n.\end{aligned}$$

Since $\int_{\bigcap_j\{\psi_j>\phi_j-k\}}\theta_{\psi_1}^1\wedge\cdots\wedge\theta_{\psi_n}^n\to\int_X\theta_{\psi_1}^1\wedge\cdots\wedge\theta_{\psi_n}^n$ as $k\to\infty$, the equivalence of the lemma follows after we take the limit $k\to\infty$ in the above identity. □

As a consequence of this last lemma and the locality of the nonpluripolar product with respect to the plurifine topology we obtain the uniform estimate

$$\lim_{k\to\infty}\left|\int_B \theta_{\psi_1}^1\wedge\cdots\wedge\theta_{\psi_n}^n-\int_B \theta_{\max(\psi_1,\phi_1-k)}^1\wedge\cdots\wedge\theta_{\max(\psi_n,\phi_n-k)}^n\right| \le 2\int_{\bigcup_j\{\psi_j\le\phi_j-k\}}\theta_{\max(\psi_1,\phi_1-k)}^1\wedge\cdots\wedge\theta_{\max(\psi_n,\phi_n-k)}^n\to 0$$

for any Borel set $B\subset X$ and $(\psi_1,\dots,\psi_n)\in\mathcal{E}(X,\theta_{\phi_1}^1,\dots,\theta_{\phi_n}^n)$.

Lastly, we note the *partial comparison principle* for nonpluripolar products of relative full mass, generalizing a result of [Dinew 2009b]:

Proposition 3.5. *Suppose* $\phi_k,\psi_k\in\mathrm{PSH}(X,\theta^k)$, $k=1,\dots,j\le n$, *and* $\phi\in\mathrm{PSH}(X,\theta)$. *Assume that* $(u,\dots,u,\psi_1,\dots,\psi_j),(v,\dots,v,\psi_1,\dots,\psi_j)\in\mathcal{E}(X,\theta_\phi,\dots,\theta_\phi,\theta_{\phi_1},\dots,\theta_{\phi_j})$. *Then*

$$\int_{\{u<v\}}\theta_v^{n-j}\wedge\theta_{\psi_1}^1\wedge\cdots\wedge\theta_{\psi_j}^j\le\int_{\{u<v\}}\theta_u^{n-j}\wedge\theta_{\psi_1}^1\wedge\cdots\wedge\theta_{\psi_j}^j.$$

Proof. The proof follows the argument of [Boucksom et al. 2010, Proposition 2.2] with a vital ingredient from Theorem 2.4.

Since $\max(u,v)$ is more singular than ϕ and ψ_k is more singular than ϕ_k for $k=1,\dots,j$, it follows from the assumption and Theorem 2.4 that

$$\begin{aligned}\int_X\theta_\phi^{n-j}\wedge\theta_{\phi_1}^1\wedge\cdots\wedge\theta_{\phi_j}^j&=\int_X\theta_v^{n-j}\wedge\theta_{\psi_1}^1\wedge\cdots\wedge\theta_{\psi_j}^j\\&\le\int_X\theta_{\max(u,v)}^{n-j}\wedge\theta_{\psi_1}^1\wedge\cdots\wedge\theta_{\psi_j}^j\\&\le\int_X\theta_\phi^{n-j}\wedge\theta_{\phi_1}^1\wedge\cdots\wedge\theta_{\phi_j}^j.\end{aligned}$$

Hence the inequalities above are in fact equalities. By locality of the nonpluripolar product we then can write

$$\begin{aligned}&\int_X\theta_{\max(u,v)}^{n-j}\wedge\theta_{\psi_1}^1\wedge\cdots\wedge\theta_{\psi_j}^j\\&\quad\ge\int_{\{u>v\}}\theta_u^{n-j}\wedge\theta_{\psi_1}^1\wedge\cdots\wedge\theta_{\psi_j}^j+\int_{\{v>u\}}\theta_v^{n-j}\wedge\theta_{\psi_1}^1\wedge\cdots\wedge\theta_{\psi_j}^j\\&\quad=\int_X\theta_u^{n-j}\wedge\theta_{\psi_1}^1\wedge\cdots\wedge\theta_{\psi_j}^j-\int_{\{u\le v\}}\theta_u^{n-j}\wedge\theta_{\psi_1}^1\wedge\cdots\wedge\theta_{\psi_j}^j+\int_{\{v>u\}}\theta_v^{n-j}\wedge\theta_{\psi_1}^1\wedge\cdots\wedge\theta_{\psi_j}^j\\&\quad=\int_X\theta_{\max(u,v)}^{n-j}\wedge\theta_{\psi_1}^1\wedge\cdots\wedge\theta_{\psi_j}^j-\int_{\{u\le v\}}\theta_u^{n-j}\wedge\theta_{\psi_1}^1\wedge\cdots\wedge\theta_{\psi_j}^j+\int_{\{v>u\}}\theta_v^{n-j}\wedge\theta_{\psi_1}^1\wedge\cdots\wedge\theta_{\psi_j}^j.\end{aligned}$$

We thus get

$$\int_{\{u<v\}}\theta_v^{n-j}\wedge\theta_{\psi_1}^1\wedge\cdots\wedge\theta_{\psi_j}^j\le\int_{\{u\le v\}}\theta_u^{n-j}\wedge\theta_{\psi_1}^1\wedge\cdots\wedge\theta_{\psi_j}^j.$$

Replacing u with $u+\varepsilon$ in the above inequality, and letting $\varepsilon\searrow 0$, by the monotone convergence theorem we arrive at the result. □

In the next subsection, after we explore the class $\mathcal{E}(X,\theta,\phi)$, we will give a partial comparison principle specifically for this class, as a corollary of the above general proposition. Here we only note the following trivial consequence:

Corollary 3.6. *Suppose $\phi \in \mathrm{PSH}(X,\theta)$ and assume that $u, v \in \mathcal{E}(X,\theta,\phi)$. Then*

$$\int_{\{u<v\}} \theta_v^n \le \int_{\{u<v\}} \theta_u^n.$$

Note that the above result generalizes [Boucksom et al. 2010, Corollary 2.3].

3B. ***The envelope $P_\theta[\phi]$ and the class $\mathcal{E}(X,\theta,\phi)$.*** Let θ be a smooth closed real $(1,1)$-form on X which represents a big class and fix $\phi \in \mathrm{PSH}(X,\theta)$ such that $\phi \le 0$. In this short subsection we focus on the relative full mass class $\mathcal{E}(X,\theta,\phi)$.

Based on our previous findings, one wonders if the following set of potentials has a maximal element:

$$F_\phi := \left\{ v \in \mathrm{PSH}(X,\theta) \;\middle|\; \phi \le v \le 0 \ \text{ and } \int_X \theta_v^n = \int_X \theta_\phi^n \right\}.$$

In other words, does there exist a *least singular* potential that is less singular than ϕ but has the same full mass as ϕ. As we will see, if $\int_X \theta_\phi^n > 0$, this is indeed the case; moreover this maximal potential is equal to $P_\theta[\phi]$ (Theorem 3.12).

Linking the envelope $P_\theta[\phi]$ to the class $\mathcal{E}(X,\theta,\phi)$, observe that $\phi \le P_\theta[\phi] \le 0$ and $\int_X \theta_{P_\theta[\phi]}^n = \int_X \theta_\phi^n$; in particular $P_\theta[\phi] \in F_\phi$ and $\phi \in \mathcal{E}(X,\theta,P_\theta[\phi])$. Indeed, since $P_\theta(\phi + C, 0) \nearrow P_\theta[\phi](0) = P_\theta[\phi]$ a.e. as $C \to \infty$, using Theorems 2.4 and 2.3 we can conclude that $\int_X \theta_{P_\theta[\phi]}^n = \int_X \theta_\phi^n$.

In our study, we will need the following preliminary result, providing an estimate for the complex Monge–Ampère operator of rooftop envelopes, which builds on recent progress in [Guedj et al. 2017]:

Lemma 3.7. *Let $\varphi, \psi \in \mathrm{PSH}(X,\theta)$. If $P_\theta(\varphi,\psi) \ne -\infty$ then*

$$\theta_{P_\theta(\varphi,\psi)}^n \le \mathbb{1}_{\{P_\theta(\varphi,\psi)=\varphi\}} \theta_\varphi^n + \mathbb{1}_{\{P_\theta(\varphi,\psi)=\psi\}} \theta_\psi^n.$$

Proof. For each $t > 0$ we set $\varphi_t := \max(\varphi, V_\theta - t)$, $\psi_t := \max(\psi, V_\theta - t)$ and $v_t := P_\theta(\varphi_t, \psi_t)$. Set $v := P_\theta(\phi,\psi)$. Since φ_t, ψ_t have minimal singularities, it follows from [Guedj et al. 2017, Lemma 4.1] that

$$\theta_{v_t}^n \le \mathbb{1}_{\{v_t=\varphi_t\}} \theta_{\varphi_t}^n + \mathbb{1}_{\{v_t=\psi_t\}} \theta_{\psi_t}^n. \tag{6}$$

For $C > 0$ we introduce

$$G_C := \{v > V_\theta - C\}, \quad v^C := \max(v, V_\theta - C), \quad \text{and} \quad v_t^C := \max(v_t, V_\theta - C).$$

Since $P_\theta(\varphi,\psi) \le \varphi, \psi, v_t$, we have $G_C \subset \{V_\theta - C < \varphi\} \cap \{V_\theta - C < \psi\} \cap \{V_\theta - C < v_t\}$. For arbitrary $A > 0$ and $t > C$, this inclusion allows us to build on (6) and write

$$\begin{aligned}
\mathbb{1}_{G_C} \theta_{v_t^C}^n = \mathbb{1}_{G_C} \theta_{v_t}^n &\le \mathbb{1}_{\{v_t=\varphi_t\}\cap G_C} \theta_{\varphi_t}^n + \mathbb{1}_{\{v_t=\psi_t\}\cap G_C} \theta_{\psi_t}^n \\
&\le \mathbb{1}_{\{v_t=\varphi_t\}\cap\{\varphi>V_\theta-t\}} \theta_{\varphi_t}^n + \mathbb{1}_{\{v_t=\psi_t\}\cap\{\psi>V_\theta-t\}} \theta_{\psi_t}^n \\
&= \mathbb{1}_{\{v_t=\varphi_t\}\cap\{\varphi>V_\theta-t\}} \theta_\varphi^n + \mathbb{1}_{\{v_t=\psi_t\}\cap\{\psi>V_\theta-t\}} \theta_\psi^n \le e^{A(v_t-\varphi_t)} \theta_\varphi^n + e^{A(v_t-\psi_t)} \theta_\psi^n.
\end{aligned} \tag{7}$$

To proceed, we want to prove that

$$\liminf_{t\to\infty} \mathbb{1}_{G_C}\theta^n_{v_t^C} \geq \mathbb{1}_{G_C}\theta^n_{v^C}. \tag{8}$$

More precisely, alluding to the Banach–Alaoglu theorem, we want to show that any weak limit of $\{\mathbb{1}_{G_C}\theta^n_{v_t^C}\}_t$ is greater than $\mathbb{1}_{G_C}\theta^n_{v^C}$.

Let $U := \mathrm{Amp}(\theta)$. The potential V_θ is locally bounded on U; hence so are v_t^C and v^C. To obtain (8), we employ an idea from the proof of Theorem 2.3. For $\varepsilon > 0$ consider

$$f_\varepsilon := \frac{\max(v - V_\theta + C, 0)}{\max(v - V_\theta + C, 0) + \varepsilon},$$

and observe that $f_\varepsilon \geq 0$ is quasicontinuous on X. Moreover, the f_ε increase pointwise to $\mathbb{1}_{G_C}$ as ε goes to zero. Since $v_t^C \searrow v^C$ as $t \to \infty$, from [Guedj and Zeriahi 2017, Theorem 4.26] it follows that $f_\varepsilon\theta^n_{v_t^C}|_U \to f_\varepsilon\theta^n_{v^C}|_U$ weakly. Using this we can write

$$\liminf_{t\to\infty} \mathbb{1}_{G_C}\theta^n_{v_t^C}|_U \geq \lim_{t\to\infty} f_\varepsilon\theta^n_{v_t^C}|_U = f_\varepsilon\theta^n_{v^C}|_U.$$

Since $X \setminus U$ is pluripolar, we let $\varepsilon \to 0$ and use the monotone convergence theorem to conclude (8).

Now, letting $t \to \infty$ in (7), the estimate in (8) allows us to conclude that

$$\mathbb{1}_{G_C}\theta^n_{\max(P_\theta(\varphi,\psi),V_\theta - C)} \leq e^{A(P_\theta(\varphi,\psi)-\varphi)}\theta^n_\varphi + e^{A(P_\theta(\varphi,\psi)-\psi)}\theta^n_\psi.$$

Letting $C \to \infty$, and later $A \to \infty$, we arrive at the conclusion. □

We prove in the following that the nonpluripolar complex Monge–Ampère measure of $P_\theta[\psi](\chi)$ has bounded density with respect to θ^n_χ. This plays a crucial role in the sequel.

Theorem 3.8. *Let $\psi, \chi \in \mathrm{PSH}(X,\theta)$ be such that ψ is more singular than χ. Then $\theta^n_{P_\theta[\psi](\chi)} \leq \mathbb{1}_{\{P_\theta[\psi](\chi)=\chi\}}\theta^n_\chi$. In particular, $\theta^n_{P_\theta[\psi]} \leq \mathbb{1}_{\{P_\theta[\psi]=0\}}\theta^n$.*

This result can be thought of as a regularity result for the envelope $P_\theta[\psi](\chi)$. For a more precise regularity result on such envelopes in the particular case of potentials with algebraic singularities we refer to [Ross and Witt Nyström 2017, Theorem 1.1].

Proof. Without loss of generality we can assume that $\psi, \chi \leq 0$. For each $t > 0$ we consider $P_\theta(\psi + t, \chi)$. Since ψ is more singular than χ, we note that $P_\theta(\psi + t, \chi)$ has the same singularity type as ψ and $P_\theta(\psi + t, \chi) \nearrow P_\theta[\psi](\chi)$ a.e. It follows from Lemma 3.7 that

$$\theta^n_{P_\theta(\psi+t,\chi)} \leq \mathbb{1}_{\{P_\theta(\psi+t,\chi)=\psi+t\}}\theta^n_\psi + \mathbb{1}_{\{P_\theta(\psi+t,\chi)=\chi\}}\theta^n_\chi.$$

Since $\{P_\theta(\psi + t, \chi) = \psi + t\} \subset \{\psi + t \leq \chi\} \subset \{\psi + t \leq V_\theta\}$, and the latter decreases to a pluripolar set, the first term on the right-hand side above goes to zero as $t \to \infty$. For the second term, we observe that $\{P_\theta(\psi + t, \chi) = \chi\} \subset \{P_\theta[\psi](\chi) = \chi\}$. Hence, applying Theorem 2.3, the result follows.

For the last statement, we can apply the above argument to $\chi := V_\theta$, and note that from [Berman 2013, (1.2)], see also [Darvas et al. 2018, Theorem 2.6 (arXiv version); Guedj et al. 2017, Proposition 5.2], it follows that $\theta^n_{V_\theta} \leq \mathbb{1}_{\{V_\theta=0\}}\theta^n$. □

Using the above result, we can establish a *noncollapsing property* for the class of potentials with the same singularity type as ϕ, when $\theta_\phi^n(X) > 0$:

Corollary 3.9. *Assume that $\phi \in \mathrm{PSH}(X,\theta)$ is such that $\int_X \theta_\phi^n > 0$. If U is a Borel subset of X with positive Lebesgue measure, then there exists $\psi \in \mathrm{PSH}(X,\theta)$ having the same singularity type as ϕ such that $\theta_\psi^n(U) > 0$.*

Proof. It follows from [Boucksom et al. 2010, Theorems A, B] that there exists $h \in \mathrm{PSH}(X,\theta)$ with minimal singularities such that $\theta_h^n = c\mathbb{1}_U\omega^n$ for some normalization constant $c > 0$. For $C > 0$ consider $\varphi_C := P_\theta(\phi + C, h)$ and note that φ_C has the same singularities as ϕ. It follows from Lemma 3.7 that

$$\theta_{\varphi_C}^n \le \mathbb{1}_{\{\varphi_C = \phi + C\}}\theta_\phi^n + \mathbb{1}_{\{\varphi_C = h\}}\theta_h^n \le \mathbb{1}_{\{\phi + C \le h\}}\theta_\phi^n + c\mathbb{1}_{\{\varphi_C = h\}\cap U}\omega^n.$$

Since θ_ϕ^n is nonpluripolar, we have that $\lim_{C\to\infty}\int_{\{\phi+C\le h\}}\theta_\phi^n = 0$. Thus for $C > 0$ big enough, by the above estimate we have

$$\int_{X\setminus U}\theta_{\varphi_C}^n \le \int_{\{\phi+C\le h\}}\theta_\phi^n < \int_X \theta_{\varphi_C}^n,$$

where in the last inequality we used the fact that $\int_X \theta_{\varphi_C}^n = \int_X \theta_\phi^n > 0$. This implies that $\int_U \theta_{\varphi_C}^n > 0$ for big enough $C > 0$, finishing the argument. □

A combination of Corollary 3.9 and [Witt Nyström 2017, Corollary 4.2] immediately gives the following version of the domination principle, making the conclusion of the latter corollary more precise:

Corollary 3.10. *Assume that $u, v \in \mathrm{PSH}(X,\theta)$, u is less singular than v and $\int_X \theta_u^n > 0$. If $u \ge v$ a.e. with respect to θ_u^n, then $u \ge v$ on X.*

Proof. Assume by contradiction that $\{u < v\} \subseteq X$ has positive Lebesgue measure. Then, by Corollary 3.9 we can ensure that there exists $\psi \in \mathrm{PSH}(X,\theta)$ having the same singularity type as u such that $\theta_\psi^n(\{u < v\}) > 0$. On the other hand, since $\theta_u^n(\{u < v\}) = 0$, [Witt Nyström 2017, Corollary 4.2] gives that $\theta_\psi^n(\{u < v\}) = 0$, which is a contradiction. □

The noncollapsing mass condition $\int_X \theta_u^n > 0$ is trivially seen to be necessary. We now give the version of the *domination principle* for the relative full mass class $\mathcal{E}(X,\theta,\phi)$:

Proposition 3.11. *Suppose $\phi \in \mathrm{PSH}(X,\theta)$ satisfies $\int_X \theta_\phi^n > 0$ and $u, v \in \mathcal{E}(X,\theta,\phi)$. If $\theta_u^n(\{u < v\}) = 0$ then $u \ge v$.*

Proof. First, assume that v is less singular than u. In view of Corollary 3.9 it suffices to prove that $\theta_h^n(\{u < v\}) = 0$ for all $h \in \mathrm{PSH}(X,\theta)$ with the same singularity type as u. Let h be such a potential, and after possibly adding a constant, we can assume that $h \le u, v$. We claim that for each $t \in (0,1)$, $(1-t)v + th \in \mathcal{E}(X,\theta,\phi)$. Indeed, since $(1-t)v + th$ is less singular than u, and more singular than v, by Theorem 2.4 we can write

$$\int_X \theta_u^n \le \int_X \theta_{(1-t)v+th}^n \le \int_X \theta_v^n.$$

The comparison principle (Corollary 3.6) allows us then to write

$$t^n \int_{\{u<(1-t)v+th\}} \theta_h^n \le \int_{\{u<(1-t)v+th\}} \theta_{(1-t)v+th}^n \le \int_{\{u<v\}} \theta_u^n = 0.$$

Since $0 = \theta_h^n(\{u < (1-t)v + th\}) \nearrow \theta_h^n(\{u < v\})$ as $t \to 0$, it follows that $\theta_h^n(\{u < v\}) = 0$.

For the general case, we observe that $\theta_u^n(\{u < v\}) = \theta_u^n(\{u < \max(u, v)\})$, and the first step implies $u \ge \max(u, v) \ge v$. □

Next we show that F_ϕ, the set of potentials introduced in the beginning of this subsection, has a very specific maximal element:

Theorem 3.12. *Assume that $\phi \in \mathrm{PSH}(X, \theta)$ satisfies $\int_X \theta_\phi^n > 0$ and $\phi \le 0$. Then*

$$P_\theta[\phi] = \sup_{v \in F_\phi} v.$$

In particular, $P_\theta[\phi] = P_\theta[P_\theta[\phi]]$.

As remarked in the beginning of the subsection, $P_\theta[\phi] \in F_\phi$; hence by the above result $P_\theta[\phi]$ is the maximal element of F_ϕ.

Proof. Let $u \in F_\phi$. By Theorem 3.8 we have

$$\theta_{P_\theta[\phi]}^n(\{P_\theta[\phi] < u\}) \le \mathbb{1}_{\{P_\theta[\phi]=0\}} \theta^n(\{P_\theta[\phi] < u\}) \le \mathbb{1}_{\{P_\theta[\phi]=0\}} \theta^n(\{P_\theta[\phi] < 0\}) = 0.$$

As $\phi \le u$, and $\int_X \theta_\phi^n = \int_X \theta_u^n$, by Theorems 2.4 and 2.3 we have

$$\int_X \theta_{P_\theta[\phi]}^n = \int_X \theta_\phi^n = \int_X \theta_u^n = \int_X \theta_{P_\theta[u]}^n > 0.$$

Consequently, $P_\theta[\phi], u \in \mathcal{E}(X, \theta, P_\theta[u])$ and Proposition 3.11 now ensures that $P_\theta[\phi] \ge u$; hence $P_\theta[\phi] \ge \sup_{v \in F_\phi} v$. As $P_\theta[\phi] \in F_\phi$, it follows that $P_\theta[\phi] = \sup_{v \in F_\phi} v$.

For the last statement notice that $P_\theta[\phi] = \sup_{v \in F_\phi} v \ge \sup_{v \in F_{P_\theta[\phi]}} v = P_\theta[P_\theta[\phi]]$, since $F_\phi \supset F_{P_\theta[\phi]}$. The reverse inequality is trivial. □

Remark 3.13. The assumption $\int_X \theta_\phi^n > 0$ is necessary in the above theorem. Indeed, in the setting of Remark 3.3, it can be seen that $P_\theta[\phi] \lneq \sup_{h \in F_\phi} h$, as the potential on the right-hand side is greater than $\pi_2^* v$, since $\pi_2^* v \in F_\phi$.

As a consequence of this last result, we obtain the following characterization of membership in $\mathcal{E}(X, \theta, \phi)$, providing a partial converse to Proposition 3.1:

Theorem 3.14. *Suppose $\phi \in \mathrm{PSH}(X, \theta)$ with $\int_X \theta_\phi^n > 0$ and $\phi \le 0$. The following are equivalent*:

(i) $u \in \mathcal{E}(X, \theta, \phi)$.

(ii) *ϕ is less singular than u, and $P_\theta[u](\phi) = \phi$.*

(iii) *ϕ is less singular than u, and $P_\theta[u] = P_\theta[\phi]$.*

As a consequence of the equivalence between (i) and (iii), we see that the potential $P_\theta[u]$ stays the same for all $u \in \mathcal{E}(X,\theta,\phi)$; i.e., it is an invariant of this class. In particular, since $\mathcal{E}(X,\theta,\phi) \subset \mathcal{E}(X,\theta,P_\theta[\phi])$, by the last statement of Theorem 3.12, it seems natural to only consider potentials ϕ that are in the image of the operator $\psi \to P_\theta[\psi]$, when studying classes of relative full mass $\mathcal{E}(X,\theta,\phi)$. What is more, in the next section it will be clear that considering such a ϕ is not just more natural, but also necessary when trying to solve complex Monge–Ampère equations with prescribed singularity.

Proof. Assume that (i) holds. By Theorem 3.8 it follows that $P_\theta[u](\phi) \geq \phi$ a.e. with respect to $\theta^n_{P_\theta[u](\phi)}$. Proposition 3.11 gives $P_\theta[u](\phi) = \phi$; hence (ii) holds.

Suppose (ii) holds. We can assume that $u \leq \phi \leq 0$. Then $P_\theta[u] \geq P_\theta[u](\phi) = \phi$. By the last statement of the previous theorem, this implies

$$P_\theta[u] = P_\theta[P_\theta[u]] \geq P_\theta[\phi].$$

As the reverse inequality is trivial, (iii) follows.

Lastly, assume that (iii) holds. By Theorems 2.4 and 2.3 it follows that $\int_X \theta_u^n = \int_X \theta^n_{P_\theta[u]} = \int_X \theta^n_{P_\theta[\phi]} = \int_X \theta^n_\phi$; hence (i) holds. □

Corollary 3.15. *Suppose $\phi \in \mathrm{PSH}(X,\theta)$ such that $\int_X \theta^n_\phi > 0$. Then $\mathcal{E}(X,\theta,\phi)$ is convex. Moreover, given $\psi_1,\dots,\psi_n \in \mathcal{E}(X,\theta,\phi)$ we have*

$$\int_X \theta^{s_1}_{\psi_1} \wedge \cdots \wedge \theta^{s_n}_{\psi_n} = \int_X \theta^n_\phi, \tag{9}$$

where $s_j \geq 0$ are integers such that $\sum_{j=1}^n s_j = n$.

Proof. Let $u, v \in \mathcal{E}(X,\theta,\phi)$ and fix $t \in (0,1)$. It follows from Theorem 3.14 that $P_\theta[v](\phi) = P_\theta[u](\phi) = \phi$. This implies

$$P_\theta[tv + (1-t)u](\phi) \geq t\,P_\theta[v](\phi) + (1-t)\,P_\theta[u](\phi) = \phi.$$

As the reverse inequality is trivial, another application of Theorem 3.14 gives $tv + (1-t)u \in \mathcal{E}(X,\theta,\phi)$.

We now prove the last statement. Since $\mathcal{E}(X,\theta,\phi)$ is convex, given $\psi_1,\dots,\psi_n \in \mathcal{E}(X,\theta,\phi)$ we know that any convex combination $\psi := \sum_{j=1}^n s_j\psi_j$ with $0 \leq s_j \leq 1$ and $\sum_j s_j = n$, belongs to $\mathcal{E}(X,\theta,\phi)$. Hence

$$\int_X \left(\sum_j s_j \theta_{\psi_j}\right)^n = \int_X \theta^n_\psi = \int_X \theta^n_\phi = \int_X \left(\sum_j s_j \theta_\phi\right)^n.$$

As a result, we have an identity of two homogeneous polynomials of degree n. Therefore all the coefficients of these polynomials have to be equal, giving (9). □

Lastly, we provide another corollary, in the spirit of the partial comparison principle from Proposition 3.5:

Corollary 3.16. *Suppose $\phi \in \mathrm{PSH}(X,\theta)$ with $\int_X \theta^n_\phi > 0$. Assume that $u, v, \psi_1, \dots, \psi_j \in \mathcal{E}(X,\theta,\phi)$ for some $j \in \{0,\dots,n\}$. Then*

$$\int_{\{u<v\}} \theta_v^{n-j} \wedge \theta_{\psi_1} \wedge \cdots \wedge \theta_{\psi_j} \leq \int_{\{u<v\}} \theta_u^{n-j} \wedge \theta_{\psi_1} \wedge \cdots \wedge \theta_{\psi_j}.$$

Proof. The conclusion follows immediately from (9) together with Proposition 3.5. □

4. Complex Monge–Ampère equations with prescribed singularity type

Let θ be a smooth closed real $(1,1)$-form on X such that $\{\theta\}$ is big and $\phi \in \mathrm{PSH}(X,\theta)$. By $\mathrm{PSH}(X,\theta,\phi)$ we denote the set of θ-psh functions that are more singular than ϕ. We say that $v \in \mathrm{PSH}(X,\theta,\phi)$ has *relatively minimal singularities* if v has the same singularity type as ϕ. Clearly, $\mathcal{E}(X,\theta,\phi) \subset \mathrm{PSH}(X,\theta,\phi)$.

Let μ be a nonpluripolar positive measure on X such that $\mu(X) = \int_X \theta_\phi^n > 0$. Our aim is to study existence and uniqueness of solutions to the following equation of complex Monge–Ampère type:

$$\theta_\psi^n = \mu, \quad \psi \in \mathcal{E}(X,\theta,\phi). \tag{10}$$

It is not hard to see that this equation does not have a solution for arbitrary ϕ. Indeed, suppose for the moment that $\theta = \omega$, and choose $\phi \in \mathcal{E}(X,\omega) := \mathcal{E}(X,\omega,0)$ unbounded. It is clear that $\mathcal{E}(X,\omega,\phi) \subsetneq \mathcal{E}(X,\omega,0)$. By [Boucksom et al. 2010, Theorem A], the (trivial) equation $\omega_\psi^n = \omega^n$, $\psi \in \mathcal{E}(X,\omega,0)$, is *only* solved by potentials ψ that are constant over X; hence we cannot have $\psi \notin \mathcal{E}(X,\omega,\phi)$.

This simple example suggests that we need to be more selective in our choice of ϕ to make (10) well-posed. As it turns out, the natural choice is to take ϕ such that $P_\theta[\phi] = \phi$, as suggested by our study of currents of relative full mass in the previous subsection. Therefore, for the rest of this section we ask that ϕ additionally satisfies

$$\phi = P_\theta[\phi]. \tag{11}$$

Such a potential ϕ is called a model potential, and $[\phi]$ is a *model-type singularity*. As $V_\theta = P_\theta[V_\theta]$, one can think of such ϕ as generalizations of V_θ, the potential with minimal singularity from [Boucksom et al. 2010]. We refer to Remark 1.6 for natural constructions of model-type singularities.

As a technical assumption, we will ask that ϕ has additionally *small unbounded locus*; i.e., ϕ is locally bounded outside a closed pluripolar set $A \subset X$. This will be needed to carry out arguments involving integration by parts in the spirit of [Boucksom et al. 2010].

One wonders if maybe model-type potentials (those that satisfy (11)) always have small unbounded locus. Sadly, this is not the case, as the following simple example shows. Suppose θ is a Kähler form, and $\{x_j\}_j \subset X$ is a dense countable subset. Also let $v_j \in \mathrm{PSH}(X,\theta)$ be such that $v_j < 0$, $\int_X v_j \theta^n = 1$, and v_j has a positive Lelong number at x_j. Then $\psi = \sum_j (1/2^j) v_j \in \mathrm{PSH}(X,\theta)$ has positive Lelong numbers at all x_j. As we have argued in [Darvas et al. 2018, Theorem 1.1], the Lelong numbers of $P_\theta[\psi]$ are the same as those of ψ; hence the model-type potential $P_\theta[\psi]$ cannot have small unbounded locus.

The following convergence result is important in our later study, and it can be implicitly found in the arguments of [Boucksom et al. 2010], as well as other works:

Lemma 4.1. *Let $u_k, u_k^j \in \mathrm{PSH}(X,\theta,\phi)$ and $C > 0$ such that*

$$-C \le u_k^j - \phi \le C$$

for all $j \in \mathbb{N}$ and $k \in \{1,\dots,n\}$. Assume also that $u_k^j \to u_k$, $k \in \{1,\dots,n\}$, in capacity. Suppose also that f, f_j are uniformly bounded, quasicontinuous, such that $f_j \to f$ in capacity. Then $f_j \theta_{u_1^j} \wedge \cdots \wedge \theta_{u_n^j} \to f\theta_{u_1} \wedge \cdots \wedge \theta_{u_n}$ weakly.

Proof. Let $A \subset X$ be closed pluripolar such that $\{\phi = -\infty\} \subset A$. We set $\mu_j := \theta_{u_1^j} \wedge \cdots \wedge \theta_{u_n^j}$, and $\mu := \theta_{u_1} \wedge \cdots \wedge \theta_{u_n}$. Fix a continuous function χ on X, $\varepsilon > 0$ and U an open relatively compact subset of $X \setminus A$ such that $\mu(X \setminus U) \le \varepsilon$. Fix V a slightly larger open subset of $X \setminus A$ such that $U \Subset V \Subset X \setminus A$. Fix ρ a continuous nonnegative function on X which is supported in V and is identically 1 in U. Since all functions u_k^j are uniformly bounded in V (along with u_k) it follows from [Guedj and Zeriahi 2017, Theorem 4.26] that $\chi f_j \mu_j$ converges weakly to $\chi f \mu$ in V. Also, Bedford–Taylor theory gives that μ_j converges weakly to μ in V. Thus $\liminf_j \mu_j(U) \ge \mu(U)$; hence $\limsup_j \mu_j(X \setminus U) \le \mu(X \setminus U) \le \varepsilon$ since $\mu_j(X) = \mu(X)$. Since χ, ρ, f_j, f are uniformly bounded it follows that $\limsup_j \int_{X\setminus U} \rho|\chi f_j|\mu_j$, $\limsup_j \int_{X\setminus U} |\chi f_j|\mu_j$, $\int_{X\setminus U} \rho|\chi f|\mu$, $\int_{X\setminus U} |\chi f|\mu$ are all bounded by $C\varepsilon$ for some uniform constant $C > 0$. On the other hand, since $\chi f_j \mu_j$ converges weakly to $\chi f \mu$ in V and $\rho = 0$ outside V, we have

$$\lim_j \int_X \rho \chi f_j \, d\mu_j = \int_X \rho \chi f \, d\mu.$$

Thus,

$$\limsup_j \left| \int_X \chi f_j \, d\mu_j - \int_X \chi f \, d\mu \right| \le \limsup_j \left| \int_X \rho \chi f_j \, d\mu_j - \int_X \rho \chi f \, d\mu \right| + 4C\varepsilon.$$

It then follows that

$$\limsup_j \left| \int_X \chi f_j \, d\mu_j - \int_X \chi f \, d\mu \right| \le C'\varepsilon.$$

Letting $\varepsilon \to 0$ we arrive at the conclusion. □

4A. ***The relative Monge–Ampère capacity.*** We introduce the *relative Monge–Ampère capacity* of a Borel set $B \subset X$:

$$\mathrm{Cap}_\phi(B) := \sup\left\{ \int_B \theta_\psi^n \;\middle|\; \psi \in \mathrm{PSH}(X,\theta),\ \phi \le \psi \le \phi + 1 \right\}.$$

Note that in the Kähler case a related notion of capacity was studied in [Di Nezza and Lu 2015; 2017]. In the case when $\phi = V_\theta$ we recover the Monge–Ampère capacity used in [Boucksom et al. 2010, Section 4.1]. As is well known, the (generalized) Monge–Ampère capacity and the global relative extremal functions play a vital role in establishing uniform estimates for complex Monge–Ampère equations; see [Kołodziej 1998; Boucksom et al. 2010; Di Nezza and Lu 2015; 2017]. Along these lines the capacity Cap_ϕ will play a crucial role in proving the regularity part of Theorem 1.4.

Lemma 4.2. *The relative Monge–Ampère capacity* Cap_ϕ *is inner regular*; *i.e.*,

$$\mathrm{Cap}_\phi(E) = \sup\{\mathrm{Cap}_\phi(K) \mid K \subset E,\ K \text{ is compact}\}.$$

Proof. By definition, $\mathrm{Cap}_\phi(E) \ge \mathrm{Cap}_\phi(K)$ for any compact set $K \subset E$. Fix $\varepsilon > 0$. There exists $u \in \mathrm{PSH}(X,\theta)$ such that $\phi \le u \le \phi + 1$ and

$$\int_E \theta_u^n \ge \mathrm{Cap}_\phi(E) - \varepsilon.$$

Since θ_u^n is an inner regular Borel measure it follows that there exists a compact set $K \subset E$ such that $\int_K \theta_u^n \geq \int_E \theta_u^n - \varepsilon \geq \mathrm{Cap}_\phi(E) - 2\varepsilon$. Hence $\mathrm{Cap}_\phi(K) \geq \mathrm{Cap}_\phi(E) - 2\varepsilon$. Letting $\varepsilon \to 0$ and taking the supremum over all the compact sets $K \subset E$, we arrive at the conclusion. □

By definition, $\mathrm{Cap}_\theta(B) \leq \mathrm{Cap}_\theta(X) = \int_X \theta_\phi^n$. Next we note that if $\mathrm{Cap}_\phi(B) = 0$ then B is a very "small" set:

Lemma 4.3. *Let $B \subset X$ be a Borel set. Then $\mathrm{Cap}_\phi(B) = 0$ if and only if B is pluripolar.*

Proof. Fix ω Kähler with $\omega \geq \theta$. Recall that a Borel subset $E \subset X$ is pluripolar if and only if $\mathrm{Cap}_\omega(E) = 0$; see [Guedj and Zeriahi 2005, Corollary 3.11], which goes back to [Bedford and Taylor 1982].

If B is pluripolar then $\mathrm{Cap}_\phi(B) = 0$ by definition. Conversely, assume that $\mathrm{Cap}_\phi(B) = 0$. If B is nonpluripolar then $\mathrm{Cap}_\omega(B) > 0$. Since Cap_ω is inner regular [Berman et al. 2013, Remark 1.7], there exists a compact subset K of B such that $\mathrm{Cap}_\omega(K) > 0$. In particular K is nonpluripolar; hence the global extremal function of (K, ω), $V^*_{\omega,K}$, is bounded from above (i.e., it is not identically ∞) by [Guedj and Zeriahi 2017, Theorem 9.17]. Since $\omega \geq \theta$ we have $V^*_{\theta,K} \leq V^*_{\omega,K}$; hence $V^*_{\theta,K}$ is also bounded from above.

We recall that $\theta^n_{V^*_{\theta,K}}$ is supported on K [Guedj and Zeriahi 2017, Theorem 9.17], and we consider $u_t := P_\theta(\phi + t, V^*_{\theta,K})$, $t > 0$. By the argument of Corollary 3.9 there exists $t_0 > 0$ big enough such that $\psi := u_{t_0} \in \mathrm{PSH}(X, \theta)$ has the same singularity type as ϕ and $\int_K \theta_\psi^n > 0$. We can assume that $\phi \leq \psi \leq \phi + C$ for some $C > 0$. If $C \leq 1$ then ψ is a candidate in the definition of $\mathrm{Cap}_\phi(B)$; hence $\mathrm{Cap}_\phi(B) > 0$, which is a contradiction. In case $C > 1$, then $(1 - 1/C)\phi + (1/C)\psi$ is a candidate in the definition of $\mathrm{Cap}_\phi(K)$; hence

$$\mathrm{Cap}_\phi(B) \geq \mathrm{Cap}_\phi(K) \geq \int_K \theta^n_{(1-1/C)\phi+(1/C)\psi} > \frac{1}{C^n} \int_K \theta_\psi^n > 0,$$

a contradiction. □

4A1. *The ϕ-relative extremal function.* Recall that ϕ has small unbounded locus; i.e., ϕ is locally bounded outside a closed complete pluripolar subset $A \subset X$. Recall that by $\mathrm{PSH}(X, \theta, \phi)$ we denote the set of all θ-psh functions which are more singular than ϕ.

Let E be a Borel subset of X. The relative extremal function of (E, ϕ, θ) is defined as

$$h_{E,\phi} := \sup\{u \in \mathrm{PSH}(X, \theta, \phi) \mid u \leq \phi - 1 \text{ on } E,\ u \leq 0 \text{ on } X\}.$$

Lemma 4.4. *Let E be a Borel subset of X and $h_{E,\phi}$ be the relative extremal function of (E, ϕ, θ). Then $h^*_{E,\phi}$ is a θ-psh function such that $\phi - 1 \leq h^*_{E,\phi} \leq \phi$. Moreover, $\theta^n_{h^*_{E,\phi}}$ vanishes on $\{h^*_{E,\phi} < 0\} \setminus \bar{E}$.*

Proof. Since $\phi - 1$ is a candidate defining $h_{E,\phi}$, it follows that $\phi - 1 \leq h_{E,\phi} \leq h^*_{E,\phi}$. Any $u \in \mathrm{PSH}(X, \theta, \phi)$ with $u \leq 0$ is a candidate of $P_\theta(\phi + C, 0)$ for some $C \in \mathbb{R}$. By Theorem 3.12 we get $u \leq P_\theta[\phi] = \phi$; hence $h^*_{E,\phi} \leq \phi$.

By the above, $h^*_{E,\phi}$ is locally bounded outside the closed pluripolar set A, and a standard balayage argument, see, e.g., [Bedford and Taylor 1976; Guedj and Zeriahi 2005, Proposition 4.1; Berman et al. 2013, Lemma 1.5], gives that $\theta^n_{h^*_{E,\phi}}$ vanishes in $\{h^*_{E,\phi} < 0\} \setminus \bar{E}$. □

Theorem 4.5. *If K is a compact subset of X and $h := h^*_{K,\phi}$ then*

$$\mathrm{Cap}_\phi(K) = \int_K \theta_h^n = \int_X (\phi - h)\theta_h^n.$$

Proof. Set $h := h^*_{K,\phi}$ and observe that $h+1$ is a candidate defining Cap_ϕ. Since θ_h^n puts no mass on the set $\{h < \phi\} \setminus K$ and $h = \phi - 1$ on K modulo a pluripolar set, we thus get

$$\mathrm{Cap}_\phi(K) \geq \int_K \theta_h^n = \int_X (\phi - h)\theta_h^n.$$

Now let u be a θ-psh function such that $\phi - 1 \leq u \leq \phi$. For a fixed $\varepsilon \in (0,1)$ set $u_\varepsilon := (1-\varepsilon)u + \varepsilon\phi$. Since $h = \phi - 1$ on K modulo a pluripolar set and $\phi - 1 \leq u_\varepsilon$ it follows that $K \subset \{h < u_\varepsilon\}$ modulo a pluripolar set. By the comparison principle we then get

$$(1-\varepsilon)^n \int_K \theta_u^n \leq \int_{\{h<u_\varepsilon\}} \theta_{u_\varepsilon}^n \leq \int_{\{h<u_\varepsilon\}} \theta_h^n = \int_K \theta_h^n,$$

where in the last equality we use the fact that θ_h^n vanishes in $\{h<0\} \setminus K$. Since u was taken arbitrarily, letting $\varepsilon \to 0$ we obtain $\mathrm{Cap}_\phi(K) \leq \int_K \theta_h^n$. This together with the previous step gives the result. □

Corollary 4.6. *If (K_j) is a decreasing sequence of compact sets then*

$$\mathrm{Cap}_\phi(K) = \lim_{j\to\infty} \mathrm{Cap}_\phi(K_j),$$

where $K := \bigcap_j K_j$. In particular, for any compact set K we have

$$\mathrm{Cap}_\phi(K) = \inf\{\mathrm{Cap}_\phi(U) \mid K \subset U \subset X,\ U \text{ is open in } X\}.$$

Proof. Let $h_j := h^*_{K_j,\phi}$ be the relative extremal function of (K_j, ϕ). Then (h_j) increases almost everywhere to $h \in \mathrm{PSH}(X,\theta)$, which satisfies $\phi - 1 \leq h \leq \phi$, since $\phi - 1 \leq h_j \leq \phi$.

Next we claim that $\theta_h^n(\{h<0\} \setminus K) = 0$. Indeed, for $m \in \mathbb{N}$ fixed and for each $j > m$ we have that $\{h<0\} \setminus K_m \subset \{h_j < 0\} \setminus K_j$ and by Lemma 4.4,

$$\theta_{h_j}^n(\{h_j < 0\} \setminus K_j) = 0.$$

Using the continuity of the Monge–Ampère measure along monotone sequences (Theorem 2.3 and Remark 2.5) we have that $\theta_{h_j}^n$ converges weakly to θ_h^n. Since $\{h<0\} \setminus K_m$ is open, it follows that

$$\theta_h^n(\{h<0\} \setminus K_m) \leq \liminf_{j\to\infty} \theta_{h_j}^n(\{h<0\} \setminus K_m) = 0.$$

The claim follows as $m \to \infty$. It then follows from Theorem 4.5 and Lemma 4.1 that

$$\lim_{j\to\infty} \mathrm{Cap}_\phi(K_j) = \lim_{j\to\infty} \int_X (\phi - h_j)\theta_{h_j}^n = \int_X (\phi - h)\theta_h^n = \int_K \theta_h^n \leq \mathrm{Cap}_\phi(K).$$

As the reverse inequality is trivial, the first statement follows.

To prove the last statement, let (K_j) be a decreasing sequence of compact sets such that K is contained in the interior of K_j for all j. Then by the first part of the corollary we have

$$\begin{aligned}\mathrm{Cap}_\phi(K) = \lim_{j\to\infty} \mathrm{Cap}_\phi(K_j) &\geq \lim_{j\to\infty} \mathrm{Cap}_\phi(\mathrm{Int}(K_j))\\ &\geq \inf\{\mathrm{Cap}_\phi(U) \mid K \subset U \subset X,\ U \text{ is open in } X\},\end{aligned}$$

and hence equality. □

Corollary 4.7. *If U is an open subset of X then*

$$\mathrm{Cap}_\phi(U) = \int_X (\phi - h_{U,\phi})\theta^n_{h_{U,\phi}}.$$

Proof. Let (K_j) be an increasing sequence of compact subsets of U such that $\bigcup K_j = U$. For each j we set $h_j := h^*_{K_j,\phi}$. By Theorem 4.5 we have

$$\mathrm{Cap}_\phi(K_j) = \int_X (\phi - h_j)\theta^n_{h_j}.$$

Since h_j decreases to $h_{U,\phi}$, it follows from Lemma 4.1 that the right-hand side above converges to $\int_X (\phi - h_{U,\phi})\theta^n_{h_{U,\phi}}$. Moreover, by the argument of Lemma 4.2 we have $\lim_j \mathrm{Cap}_\phi(K_j) = \mathrm{Cap}_\phi(U)$; hence the result follows. □

4A2. *The global ϕ-extremal function.* For a Borel set $E \subset X$, we define the global *ϕ-extremal function* of (E, ϕ, θ) by

$$V_{E,\phi} := \sup\{\psi \in \mathrm{PSH}(X, \theta, \phi) \mid \psi \leq \phi \text{ on } E\}.$$

We then introduce the *relative Alexander–Taylor capacity* of E,

$$T_\phi(E) := \exp(-M_\phi(E)), \quad \text{where } M_\phi(E) := \sup_X V^*_{E,\phi}.$$

Paralleling Lemma 4.3, we have the following result:

Lemma 4.8. *Let $E \subset X$ be a Borel set. If $M_\phi(E) = \infty$, then E is pluripolar.*

Proof. Let ω be a Kähler form such that $\omega \geq \theta$. By definition we have

$$V_{E,\phi} \leq V_{E,\omega} := \sup\{\psi \in \mathrm{PSH}(X, \omega) \mid \psi \leq 0 \text{ on } E\}.$$

This clearly implies $M_\phi(E) \leq \sup_X V^*_{E,\omega}$, and so by assumption we know that $\sup_X V^*_{E,\omega} = \infty$. It then follows from [Guedj and Zeriahi 2005, Theorem 5.2] that E is pluripolar. □

If $M_\phi(E) < \infty$ then $V^*_{E,\phi} \in \mathrm{PSH}(X, \theta)$, and standard arguments give that $\theta^n_{V^*_{E,\phi}}$ does not charge $X \setminus \bar{E}$; see [Guedj and Zeriahi 2005, Theorem 5.2; 2017, Theorem 9.17]. Now, we claim that

$$\phi \leq V^*_{E,\phi} \leq P_\theta[\phi] + M_\phi(E) = \phi + M_\phi(E). \tag{12}$$

The first inequality simply follows by definition, since $\phi \leq 0$ is a candidate in the definition of $V_{E,\phi}$. If $M_\phi(E) = \infty$ then the second inequality holds trivially. Assume that $M_\phi(E) < \infty$. The inequality then holds, since $V^*_{E,\phi} - M_\phi(E) \leq 0$, and each candidate potential ψ in the definition of $V^*_{E,\phi}$ is more singular

than ϕ; i.e., $\psi - M_\phi(E)$ is a candidate in the definition of $P_\theta(\phi + C, 0)$ for some $C > 0$. Finally, the last identity follows from Theorem 3.12.

In particular, since ϕ has small unbounded locus, so does the upper semicontinuous regularization $V^*_{E,\phi}$. Also, from (12) we deduce that if $M_\phi(E) < \infty$, the θ-psh functions $V^*_{E,\phi}$ and ϕ have the same singularity type; hence Proposition 2.1 ensures that

$$\int_X \theta^n_{V^*_{E,\phi}} = \int_X \theta^n_\phi.$$

The Alexander–Taylor and Monge–Ampère capacities are related by the following estimates:

Lemma 4.9. *Suppose $K \subset X$ is a compact subset and $\mathrm{Cap}_\phi(K) > 0$. Then we have*

$$1 \le \left(\frac{\int_X \theta^n_\phi}{\mathrm{Cap}_\phi(K)}\right)^{1/n} \le \max(1, M_\phi(K)).$$

Proof. The first inequality is trivial. We now prove the second inequality. Note that we can assume that $M_\phi(K) < \infty$, since otherwise the inequality is trivially satisfied. We then consider two cases. If $M_\phi(K) \le 1$, then $V^*_{K,\phi} \le \phi + 1$; hence $V^*_{K,\phi}$ is a candidate in the definition of $\mathrm{Cap}_\phi(K)$. Since $\theta^n_{V^*_{K,\theta}}$ is supported on K, we thus have

$$\mathrm{Cap}_\phi(K) \ge \int_K \theta^n_{V^*_{K,\phi}} = \int_X \theta^n_{V^*_{K,\phi}} = \int_X \theta^n_\phi,$$

and the desired inequality holds in this case.

If $M := M_\phi(K) \ge 1$, then by (12) we have $\phi \le M^{-1} V^*_{K,\phi} + (1 - M^{-1})\phi \le \phi + 1$, and by the definition of the relative capacity we can write

$$\mathrm{Cap}_\phi(K) \ge \int_K \theta^n_{M^{-1}V^*_{K,\phi} + (1-M^{-1})\phi} \ge \frac{1}{M^n} \int_K \theta^n_{V^*_{K,\phi}} = \frac{1}{M^n} \int_X \theta^n_{V^*_{K,\phi}} = \frac{1}{M^n} \int_X \theta^n_\phi,$$

implying the desired inequality. □

4B. *The relative finite energy class $\mathcal{E}^1(X, \theta, \phi)$.* To develop the variational approach to (10), we need to understand the relative version of the Monge–Ampère energy, and its bounded locus $\mathcal{E}^1(X, \theta, \phi)$.

For $u \in \mathcal{E}(X, \theta, \phi)$ with relatively minimal singularities, we define the Monge–Ampère energy of u relative to ϕ as

$$\mathrm{I}_\phi(u) := \frac{1}{n+1} \sum_{k=0}^{n} \int_X (u - \phi)\theta^k_u \wedge \theta^{n-k}_\phi.$$

In the next theorem we collect basic properties of the Monge–Ampère energy:

Theorem 4.10. *Suppose $u, v \in \mathcal{E}(X, \theta, \phi)$ have relatively minimal singularities. The following hold*:

(i) $\mathrm{I}_\phi(u) - \mathrm{I}_\phi(v) = 1/(n+1) \sum_{k=0}^n \int_X (u - v)\theta^k_u \wedge \theta^{n-k}_v$.

(ii) *If $u \le \phi$ then, $\int_X (u - \phi)\theta^n_u \le \mathrm{I}_\phi(u) \le 1/(n+1) \int_X (u - \phi)\theta^n_u$.*

(iii) *I_ϕ is nondecreasing and concave along affine curves. Additionally, the estimates $\int_X (u - v)\theta^n_u \le \mathrm{I}_\phi(u) - \mathrm{I}_\phi(v) \le \int_X (u - v)\theta^n_v$ hold.*

Proof. Since ϕ has small unbounded locus, it is possible to repeat the arguments of [Boucksom et al. 2010, Proposition 2.8] almost word for word. As a courtesy to the reader, the detailed proof is presented here.

To start, we note that the nonpluripolar products appearing in our arguments are simply the mixed Monge–Ampère measures defined in the sense of [Bedford and Taylor 1976] on $X \setminus A$, where A is a closed complete pluripolar subset of X such that ϕ is locally bounded on $X \setminus A$ (consequently, u and v are locally bounded on $X \setminus A$). Since $u - v$ is globally bounded on X, we can perform integration by parts in our arguments below, via [Boucksom et al. 2010, Theorem 1.14].

For any fixed $k \in \{0, \dots, n-1\}$, set $T = \theta_u^k \wedge \theta_v^{n-k-1}$. Using integration by parts [Boucksom et al. 2010, Theorem 1.14], we can write

$$\begin{aligned}\int_X (u-v)\theta_u^k \wedge \theta_v^{n-k} &= \int_X (u-v)(\theta + i\partial\bar\partial v) \wedge T \\ &= \int_X (u-v) i\partial\bar\partial(v-u) \wedge T + \int_X (u-v) i\partial\bar\partial u \wedge T + \int_X (u-v)\theta \wedge T \\ &= \int_X (v-u) i\partial\bar\partial(u-v) \wedge T + \int_X (u-v)\theta_u \wedge T \\ &\geq \int_X (u-v)\theta_u \wedge T = \int_X (u-v)\theta_u^{k+1} \wedge \theta_v^{n-k-1}, \end{aligned} \tag{13}$$

where in the last inequality we used that

$$\int_X (-\varphi) i\partial\bar\partial\varphi \wedge T = i \int_X \partial\varphi \wedge \bar\partial\varphi \wedge T \geq 0$$

with $\varphi := u - v$. This shows in particular that the sequence $k \mapsto \int_X (u-\phi)\theta_u^k \wedge \theta_\phi^{n-k}$ is nonincreasing in k, verifying (ii).

Now we compute the derivative of $f(t) := \mathrm{I}_\phi(u_t)$, $t \in [0,1]$, where $u_t := tu + (1-t)v$. By the multilinearity property of the nonpluripolar product we see that $f(t)$ is a polynomial in t. Using again integration by parts [Boucksom et al. 2010, Theorem 1.14], one can check the following formula:

$$\begin{aligned} f'(t) &= \frac{1}{n+1}\left(\sum_{k=0}^n \int_X (u-v)\theta_{u_t}^k \wedge \theta_\phi^{n-k} + \sum_{k=1}^n \int_X k(u_t-\phi) i\partial\bar\partial(u-v) \wedge \theta_{u_t}^{k-1} \wedge \theta_\phi^{n-k}\right) \\ &= \frac{1}{n+1}\left(\sum_{k=0}^n \int_X (u-v)\theta_{u_t}^k \wedge \theta_\phi^{n-k} + \sum_{k=1}^n \int_X k(u-v)(\theta_{u_t} - \theta_\phi) \wedge \theta_{u_t}^{k-1} \wedge \theta_\phi^{n-k}\right) = \int_X (u-v)\theta_{u_t}^n. \end{aligned}$$

Computing one more derivative, we arrive at

$$f''(t) = n \int_X (u-v) i\partial\bar\partial(u-v) \wedge \theta_{u_t}^{n-1} = -ni \int_X \partial(u-v) \wedge \bar\partial(u-v)\theta_{u_t}^{n-1} \leq 0.$$

This shows that I_ϕ is concave along affine curves.

Now, the function $t \mapsto f'(t)$ is continuous on $[0,1]$, thanks to the convergence property of the Monge–Ampère operator (see Lemma 4.1). It thus follows that

$$\mathrm{I}_\phi(u_1) - \mathrm{I}_\phi(u_0) = \int_0^1 f'(t)\, dt = \int_0^1 \int_X (u-v)\theta_{u_t}^n\, dt.$$

Using the multilinearity of the nonpluripolar product again, we get

$$\int_0^1\int_X (u-v)\theta_{u_t}^n\,dt = \sum_{k=0}^n\left(\int_0^1 \binom{n}{k}t^k(1-t)^{n-k}\,dt\right)\int_X (u-v)\theta_u^k\wedge\theta_v^{n-k}$$
$$= \frac{1}{n+1}\sum_{k=0}^n\int_X (u-v)\theta_u^k\wedge\theta_v^{n-k}.$$

This verifies (i), and another application of (13) finishes the proof of (iii). □

Lemma 4.11. *Suppose $u_j, u\in\mathcal{E}(X,\theta,\phi)$ have relatively minimal singularities such that u_j decreases to u. Then $\mathrm{I}_\phi(u_j)$ decreases to $\mathrm{I}_\phi(u)$.*

Proof. From Theorem 4.10(iii) it follows that $|\mathrm{I}_\phi(u_j)-\mathrm{I}_\phi(u)| = \mathrm{I}_\phi(u_j)-\mathrm{I}_\phi(u)\le\int_X(u_j-u)\theta_u^n$. An application of the dominated convergence theorem finishes the argument. □

We can now define the Monge–Ampère energy for arbitrary $u\in\mathrm{PSH}(X,\theta,\phi)$ using a familiar formula:

$$\mathrm{I}_\phi(u) := \inf\{\mathrm{I}_\phi(v)\mid v\in\mathcal{E}(X,\theta,\phi),\ v \text{ has relatively minimal singularities, and } u\le v\}.$$

Lemma 4.12. *If $u\in\mathrm{PSH}(X,\theta,\phi)$ then $\mathrm{I}_\phi(u)=\lim_{t\to\infty}\mathrm{I}_\phi(\max(u,\phi-t))$.*

Proof. It follows from the above definition that $\mathrm{I}_\phi(u)\le\lim_{t\to\infty}\mathrm{I}_\phi(\max(u,\phi-t))$. Assume now that $v\in\mathrm{PSH}(X,\theta,\phi)$ is such that $u\le v$, and v has the same singularity type as ϕ (i.e., v is a candidate in the definition of $\mathrm{I}_\phi(u)$). Then for t large enough we have $\max(u,\phi-t)\le v$; hence the other inequality follows from the monotonicity of I_ϕ. □

We let $\mathcal{E}^1(X,\theta,\phi)$ denote the set of all $u\in\mathrm{PSH}(X,\theta,\phi)$ such that $\mathrm{I}_\phi(u)$ is finite. As a result of Lemma 4.12 and Theorem 4.10(iii) we observe that I_ϕ is nondecreasing in $\mathrm{PSH}(X,\theta,\phi)$. Consequently, $\mathcal{E}^1(X,\theta,\phi)$ is stable under the max operation; moreover, we have the following familiar characterization of $\mathcal{E}^1(X,\theta,\phi)$:

Lemma 4.13. *Suppose $u\in\mathrm{PSH}(X,\theta,\phi)$. Then $u\in\mathcal{E}^1(X,\theta,\phi)$ if and only if $u\in\mathcal{E}(X,\theta,\phi)$ and $\int_X(u-\phi)\theta_u^n>-\infty$.*

Proof. We can assume that $u\le\phi$. For each $C>0$ we set $u^C:=\max(u,\phi-C)$. If $\mathrm{I}_\phi(u)>-\infty$ then by the monotonicity property we have $\mathrm{I}_\phi(u^C)\ge\mathrm{I}_\phi(u)$. Since $u^C\le\phi$, an application of Theorem 4.10(ii) gives that $\int_X(u^C-\phi)\theta_{u^C}^n\ge -A$ for all C, for some $A>0$. From this we obtain that

$$\int_{\{u\le\phi-C\}}\theta_{u^C}^n\le\frac{A}{C}\to 0$$

as $C\to\infty$. Hence it follows from Lemma 3.4 that $u\in\mathcal{E}(X,\theta,\phi)$. Moreover by the plurifine property of the nonpluripolar product we have

$$\int_X(u^C-\phi)\theta_{u^C}^n\le\int_{\{u>\phi-C\}}(u-\phi)\theta_u^n.$$

Letting $C\to\infty$ we see that $\int_X(u-\phi)\theta_u^n>-A$.

To prove the reverse statement, assume that $u \in \mathcal{E}(X,\theta,\phi)$ and $\int_X (u-\phi)\theta_u^n > -\infty$. For each $C>0$, since θ_u^n and $\theta_{u^C}^n$ have the same mass and coincide in $\{u > \phi - C\}$, it follows that $\int_{\{u\le\phi-C\}} \theta_{u^C}^n = \int_{\{u\le\phi-C\}} \theta_u^n$. From this we deduce that

$$\int_X (u^C-\phi)\theta_{u^C}^n = -\int_{\{u\le\phi-C\}} C\theta_u^n + \int_{\{u>\phi-C\}} (u-\phi)\theta_u^n = \int_X (u-\phi)\theta_u^n > -A.$$

It thus follows from Theorem 4.10(ii) that $\mathrm{I}_\phi(u^C)$ is uniformly bounded. Finally, it follows from Lemma 4.12 that $\mathrm{I}_\phi(u^C) \searrow \mathrm{I}_\phi(u)$ as $C\to\infty$, finishing the proof. □

We finish this subsection with a series of small results listing various properties of the class $\mathcal{E}^1(X,\theta,\phi)$:

Lemma 4.14. *Assume that (u_j) is a sequence in $\mathcal{E}^1(X,\theta,\phi)$ decreasing to $u\in\mathcal{E}^1(X,\theta,\phi)$. Then $\mathrm{I}_\phi(u_j)$ decreases to $\mathrm{I}_\phi(u)$.*

Proof. Without loss of generality we can assume that $u_j \le \phi$ for all j. For each $C>0$ we set $u_j^C := \max(u_j, \phi - C)$ and $u^C := \max(u, \phi - C)$. Note that u_j^C, u^C have the same singularities as ϕ. Then Lemma 4.11 ensures that $\lim_j \mathrm{I}_\phi(u_j^C) = \mathrm{I}_\phi(u^C)$. The monotonicity of I_ϕ gives now that $\mathrm{I}_\phi(u) \le \lim_j \mathrm{I}_\phi(u_j) \le \lim_j \mathrm{I}_\phi(u_j^C) = \mathrm{I}_\phi(u^C)$. Letting $C\to\infty$, the result follows. □

Lemma 4.15. *Assume that (u_j) is a decreasing sequence in $\mathcal{E}^1(X,\theta,\phi)$ such that $\mathrm{I}_\phi(u_j)$ is uniformly bounded. Then the limit $u := \lim_j u_j$ belongs to $\mathcal{E}^1(X,\theta,\phi)$ and $\mathrm{I}_\phi(u_j)$ decreases to $\mathrm{I}_\phi(u)$.*

Proof. We can assume that $u_j \le \phi$ for all j. Since $\mathrm{I}_\phi(u_j) \le \int_X (u_j-\phi)\theta_\phi^n$, $\mathrm{I}_\phi(u_j)$ is uniformly bounded and θ_ϕ^n has bounded density with respect to ω^n, it follows that $\int_X u_j\omega^n$ is uniformly bounded; hence $u \not\equiv -\infty$.

By continuity along decreasing sequences (Lemma 4.14) we have $\lim_{j\to\infty} \mathrm{I}_\phi(\max(u_j,\phi-C)) = \mathrm{I}_\phi(\max(u,\phi-C))$. It follows that $\mathrm{I}_\phi(\max(u,\phi-C))$ is uniformly bounded. Lemma 4.12 then ensures that $\mathrm{I}_\phi(u)$ is finite; i.e., $u\in\mathcal{E}^1(X,\theta,\phi)$. □

Corollary 4.16. *I_ϕ is concave along affine curves in $\mathrm{PSH}(X,\theta,\phi)$. In particular, the set $\mathcal{E}^1(X,\theta,\phi)$ is convex.*

Proof. Let $u, v \in \mathrm{PSH}(X,\theta,\phi)$ and $u_t := tu + (1-t)v$, $t\in(0,1)$. If one of u, v is not in $\mathcal{E}^1(X,\theta,\phi)$ then the conclusion is obvious. So, we can assume that both u and v belong to $\mathcal{E}^1(X,\theta,\phi)$. For each $C>0$ we set $u_t^C := t\max(u,\phi-C) + (1-t)\max(v,\phi-C)$. By Theorem 4.10(iii), $t\to \mathrm{I}_\phi(u_t^C)$ is concave. Since u_t^C decreases to u_t as $C\to\infty$, Lemma 4.15 gives the conclusion. □

4C. ***The variational method.*** Recall that ϕ is a θ-psh function with small unbounded locus such that $\phi = P_\theta[\phi]$, and $\int_X \theta_\phi > 0$. For this subsection we additionally normalize our class so that $\int_X \theta_\phi^n = 1$.

We adapt the variational method of [Berman et al. 2013] to solve the complex Monge–Ampère equations in our more general setting:

$$\theta_u^n = e^{\lambda u}\mu, \quad u\in\mathcal{E}(X,\theta,\phi), \tag{14}$$

where $\lambda \ge 0$ and μ is a positive nonpluripolar measure on X. If $\lambda = 0$ then we also assume that $\mu(X)=1$, which is a necessary condition for the equation to be solvable.

We introduce the following functionals on $\mathcal{E}^1(X,\theta,\phi)$:

$$F_\lambda(u) := F_{\lambda,\mu}(u) := \mathrm{I}_\phi(u) - L_{\lambda,\mu}(u), \quad u \in \mathcal{E}^1(X,\theta,\phi),$$

where $L_{\lambda,\mu}(u) := (1/\lambda)\int_X e^{\lambda u}\,d\mu$ if $\lambda > 0$ and $L_\mu(u) := L_{0,\mu}(u) := \int_X (u-\phi)\,d\mu$. Note that when $\lambda > 0$, F_λ is finite on $\mathcal{E}^1(X,\theta,\phi)$. It is no longer the case if $\lambda = 0$, in which case we will restrict ourselves to the following set of measures. For each constant $A \geq 1$ we let $\mathcal{M}_A$ denote the set of all probability measures μ on X such that

$$\mu(E) \leq A \cdot \mathrm{Cap}_\phi(E) \quad \text{for all Borel subsets } E \subset X.$$

Lemma 4.17. *$\mathcal{M}_A$ is a compact convex subset of the set of probability measures on X.*

Proof. The convexity is obvious. We now prove that $\mathcal{M}_A$ is closed. Assume that $(\mu_j) \subset \mathcal{M}_A$ is a sequence converging weakly to a probability measure μ. Then for any open set U we have

$$\mu(U) \leq \liminf_j \mu_j(U) \leq A\,\mathrm{Cap}_\phi(U).$$

Now, let $K \subset X$ be a compact subset. Taking the infimum over all open sets $U \supset K$ in the above inequality, it follows from Corollary 4.6 that $\mu(K) \leq A\,\mathrm{Cap}_\phi(K)$. Since μ and Cap_ϕ are inner regular (Lemma 4.2) it follows that the inequality holds for all Borel sets, finishing the proof. □

Lemma 4.18. *If $\mu \in \mathcal{M}_A$ then $F_{0,\mu}$ is finite on $\mathcal{E}^1(X,\theta,\phi)$. Moreover, there is a constant $B > 0$ depending on A such that for all $u \in \mathrm{PSH}(X,\theta,\phi)$ with $\sup_X u = 0$ we have*

$$\int_X (u-\phi)^2\,d\mu \leq B(|\mathrm{I}_\phi(u)| + 1).$$

The proof given below is inspired by [Berman et al. 2013, Lemma 2.9].

Proof. Fix $u \in \mathrm{PSH}(X,\theta,\phi)$ such that $\sup_X u = 0$. By considering $u_k := \max(u, \phi - k)$ and then letting $k \to \infty$, we can assume that $u - \phi$ is bounded. We first prove that

$$\int_1^\infty t\,\mathrm{Cap}_\phi(u < \phi - 2t)\,dt \leq C(-\mathrm{I}_\phi(u) + 1) \tag{15}$$

for some uniform constant $C := C(n) > 0$.

Indeed, for each $t > 1$ we set $u_t := t^{-1}u + (1-t^{-1})\phi$. We also fix $\psi \in \mathrm{PSH}(X,\theta)$ such that $\phi - 1 \leq \psi \leq \phi$. Observe that $u_t, \psi \in \mathcal{E}(X,\theta,\phi)$ and that the following inclusions hold:

$$(u < \phi - 2t) \subset (u_t < \psi - 1) \subset (u < \phi - t), \quad t > 1.$$

It thus follows that

$$\theta_\psi^n(u < \phi - 2t) \leq \theta_\psi^n(u_t < \psi - 1) \leq \theta_{u_t}^n(u_t < \psi - 1) \leq \theta_{u_t}^n(u < \phi - t), \tag{16}$$

where in the second inequality we used the comparison principle (see Corollary 3.6). Expanding $\theta_{u_t}^n$ we see that

$$\theta_{u_t}^n \leq Ct^{-1}\sum_{k=1}^n \theta_u^k \wedge \theta_\phi^{n-k} + \theta_\phi^n \quad \text{for all } t > 1, \tag{17}$$

for a uniform constant $C = C(n)$. Since θ_ϕ^n has bounded density with respect to Lebesgue measure (see Theorem 3.8), using [Guedj and Zeriahi 2017, Theorem 2.50] we infer that

$$\theta_\phi^n(u < \phi - t) \le A \int_{\{u \le -t\}} \omega^n \le Ae^{-at} \tag{18}$$

for some uniform constants $a, A > 0$ depending only on n, ω, X. Combining (18) with (16) and (17) we get that

$$\begin{aligned}\int_1^\infty t\theta_\psi^n(u < \phi - 2t)\,dt &\le \int_1^\infty t\theta_{u_t}^n(u < \phi - t)\,dt \\ &\le C\int_1^\infty \sum_{k=0}^n \theta_u^k \wedge \theta_\phi^{n-k}(u < \phi - t)\,dt + \int_1^\infty t\theta_\phi^n(u < \phi - t)\,dt \\ &\le C(n+1)|\mathrm{I}_\phi(u)| + C'.\end{aligned}$$

Taking the supremum over all candidates $\psi + 1$ we arrive at

$$\int_1^\infty t\,\mathrm{Cap}_\phi(u < \phi - 2t)\,dt \le C(n+1)|\mathrm{I}_\phi(u)| + C',$$

proving (15). Finally, we can write

$$\begin{aligned}\int_X (u-\phi)^2\,d\mu = 2\int_0^\infty t\mu(u < \phi - t)\,dt &\le 4 + 8\int_1^\infty t\mu(u < \phi - 2t)\,dt \\ &\le 4 + 8\int_1^\infty At\,\mathrm{Cap}_\phi(u < \phi - 2t)\,dt \le B(|\mathrm{I}_\phi(u)| + 1),\end{aligned}$$

where $B > 0$ is a uniform constant depending on n, C, C'. □

Observe that Lemma 4.18 together with Hölder's inequality give that $F_{0,\mu}$ is finite on $\mathcal{E}^1(X, \theta, \phi)$ whenever $\mu \in \mathcal{M}_A$ for some $A \ge 1$. Indeed

$$\int_X |u - \phi|\,d\mu \le \left(\int_X (u-\phi)^2\,d\mu\right)^{1/2} \mu(X)^{1/2} \le C(|\mathrm{I}_\phi(u)|^{1/2} + 1) \tag{19}$$

for a suitable $C > 0$.

4C1. *Maximizers are solutions.*

Proposition 4.19. *$\mathrm{I}_\phi : \mathcal{E}^1(X, \theta, \phi) \to \mathbb{R}$ is upper semicontinuous with respect to the weak L^1 topology of potentials.*

Proof. Assume that (u_j) is a sequence in $\mathcal{E}^1(X, \theta, \phi)$ converging in L^1 to $u \in \mathcal{E}^1(X, \theta, \phi)$. We can assume that $u_j \le 0$ for all j. For each $k, \ell \in \mathbb{N}$ we set $v_{k,\ell} := \max(u_k, \dots, u_{k+\ell})$. As $\mathcal{E}^1(X, \theta, \phi)$ is stable under the max operation, we have $v_{k,\ell} \in \mathcal{E}^1(X, \theta, \phi)$.

Moreover $v_{k,\ell} \nearrow \varphi_k := (\sup_{j \ge k} u_j)^*$; hence by the monotonicity property we get $\mathrm{I}_\phi(\varphi_k) \ge \mathrm{I}_\phi(v_{k,\ell}) \ge \mathrm{I}_\phi(u_k) > -\infty$. As a result, $\varphi_k \in \mathcal{E}^1(X, \theta, \phi)$. By Hartogs' lemma, $\varphi_k \searrow u$ as $k \to \infty$. By Lemma 4.14 it follows that $\mathrm{I}_\phi(\varphi_k)$ decreases to $\mathrm{I}_\phi(u)$. Thus, using the monotonicity of I_ϕ we get $\mathrm{I}_\phi(u) = \lim_{k\to\infty} \mathrm{I}_\phi(\varphi_k) \ge \limsup_{k\to\infty} \mathrm{I}_\phi(u_k)$, finishing the proof. □

Next we describe the first-order variation of I_ϕ, shadowing a result from [Berman and Boucksom 2010]:

Proposition 4.20. *Let $u \in \mathcal{E}^1(X,\theta,\phi)$ and χ be a continuous function on X. For each $t>0$ set $u_t := P_\theta(u+t\chi)$. Then $u_t \in \mathcal{E}^1(X,\theta,\phi)$, $t \mapsto \mathrm{I}_\phi(u_t)$, is differentiable, and its derivative is given by*

$$\frac{d}{dt}\mathrm{I}_\phi(u_t) = \int_X \chi\theta^n_{u_t}, \quad t\in\mathbb{R}.$$

Proof. Note that $u + t\inf_X \chi$ is a candidate in each envelope; hence $u + t\inf_X \chi \le u_t$. The monotonicity of I_ϕ now implies that $u_t \in \mathcal{E}^1(X,\theta,\phi)$.

As the singularity type of each u_t is the same, we can apply Lemma 4.21 below and conclude

$$\int_X (u_{t+s}-u_t)\theta^n_{u_{t+s}} \le \mathrm{I}_\phi(u_{t+s}) - \mathrm{I}_\phi(u_t) \le \int_X (u_{t+s}-u_t)\theta^n_{u_t}.$$

It follows from [Darvas et al. 2018, Proposition 2.13] that $\theta^n_{u_t}$ is supported on $\{u_t = u+t\chi\}$. We thus have

$$\int_X (u_{t+s}-u_t)\theta^n_{u_t} = \int_X (u_{t+s}-u-t\chi)\theta^n_{u_t} \le \int_X s\chi\theta^n_{u_t},$$

since $u_{t+s} \le u + (t+s)\chi$. Similarly we have

$$\int_X (u_{t+s}-u_t)\theta^n_{u_{t+s}} = \int_X (u+(t+s)\chi-u_t)\theta^n_{u_{t+s}} \ge \int_X s\chi\theta^n_{u_{t+s}}.$$

Since u_{t+s} converges uniformly to u_t as $s\to 0$, by Theorem 2.3 it follows that $\theta^n_{u_{t+s}}$ converges weakly to $\theta^n_{u_t}$. As χ is continuous, dividing by $s>0$ and letting $s\to 0^+$ we see that the right derivative of $\mathrm{I}_\phi(u_t)$ at t is $\int_X \chi\theta^n_{u_t}$. The same argument applies for the left derivative. □

Lemma 4.21. *Suppose $u, v \in \mathcal{E}^1(X,\theta,\phi)$ have the same singularity type. Then*

$$\int_X (u-v)\theta^n_u \le \mathrm{I}_\phi(u) - \mathrm{I}_\phi(v) \le \int_X (u-v)\theta^n_v.$$

Proof. First, note that these estimates hold for $u^C := \max(u, \phi - C)$ and $v^C := \max(v,\phi-C)$, by Theorem 4.10(iii). It is easy to see that $u^C - v^C$ is uniformly bounded and converges to $u-v$. Also, by the comments after Lemma 3.4 it follows that the measures $\theta^n_{v^C}$ converge uniformly to θ^n_v (not just weakly!). Putting these last two facts together, the dominated convergence theorem gives

$$\left|\int_X (u^C-v^C)\theta^n_{v^C} - \int_X (u-v)\theta^n_v\right| \le \left|\int_X (u^C-v^C)(\theta^n_{v^C}-\theta^n_v)\right| + \left|\int_X (u^C-v^C)\theta^n_v - \int_X (u-v)\theta^n_v\right| \to 0$$

as $C\to\infty$. A similar convergence statement holds for the left-hand side of our double estimate as well, and using Lemma 4.12, the result follows. □

Theorem 4.22. *Assume that $L_{\lambda,\mu}$ is finite on $\mathcal{E}^1(X,\theta,\phi)$ and $u \in \mathcal{E}^1(X,\theta,\phi)$ maximizes $F_{\lambda,\mu}$ on $\mathcal{E}^1(X,\theta,\phi)$. Then u solves (14).*

Proof. First, let's assume that $\lambda \neq 0$. Let χ be an arbitrary continuous function on X and set $u_t := P_\theta(u+t\chi)$. It follows from Proposition 4.20 that $u_t \in \mathcal{E}^1(X,\theta,\phi)$ for all $t\in\mathbb{R}$, that the function

$$g(t) := \mathrm{I}_\phi(u_t) - L_{\lambda,\mu}(u+t\chi)$$

is differentiable on $\mathbb{R}$, and its derivative is given by $g'(t) = \int_X \chi\theta_{u_t}^n - \int_X \chi e^{\lambda(u+t\chi)}\,d\mu$. Moreover, as $u_t \le u + t\chi$, we have $g(t) \le F_{\lambda,\mu}(u_t) \le \sup_{\mathcal{E}^1(X,\theta,\phi)} F_{\lambda,\mu} = F(u) = g(0)$. This means that g attains a maximum at 0; hence $g'(0) = 0$. Since χ was taken to be arbitrary, it follows that $\theta_u^n = e^{\lambda u}\mu$. When $\lambda = 0$, similar arguments give the conclusion. □

4C2. *The case* $\lambda > 0$. Having computed the first-order variation of the Monge–Ampère energy, we establish the following existence and uniqueness result.

Theorem 4.23. *Assume that* μ *is a positive nonpluripolar measure on* X *and* $\lambda > 0$. *Then there exists a unique* $\varphi \in \mathcal{E}^1(X,\theta,\phi)$ *such that*

$$\theta_\varphi^n = e^{\lambda\varphi}\mu. \tag{20}$$

Proof. We use the variational method as above; see also [Darvas et al. 2018]. It suffices to treat the case $\lambda = 1$ as the other cases can de done similarly. Consider

$$F(u) := \mathrm{I}_\phi(u) - \int_X e^u\,d\mu, \quad u \in \mathcal{E}^1(X,\theta,\phi).$$

Let (φ_j) be a sequence in $\mathcal{E}^1(X,\theta,\phi)$ such that $\lim_j F(\varphi_j) = \sup_{\mathcal{E}^1(X,\theta,\phi)} F > -\infty$. We claim that $\sup_X \varphi_j$ is uniformly bounded from above. Indeed, assume that it were not the case. Then by relabeling the sequence we can assume that $\sup_X \varphi_j$ increases to ∞. By the compactness property [Guedj and Zeriahi 2005, Proposition 2.7] it follows that the sequence $\psi_j := \varphi_j - \sup_X \varphi_j$ converges in $L^1(X,\omega^n)$ to some $\psi \in \mathrm{PSH}(X,\theta)$ such that $\sup_X \psi = 0$. In particular $\int_X e^\psi\,d\mu > 0$. It thus follows that

$$\int_X e^{\varphi_j}\,d\mu = e^{\sup_X \varphi_j}\int_X e^{\psi_j}\,d\mu \ge c e^{\sup_X \varphi_j} \tag{21}$$

for some positive constant c. Note also that $\psi_j \le \phi$ since $\psi_j \in \mathcal{E}(X,\theta,\phi)$ and $\psi_j \le 0$ and ϕ is the maximal function with these properties (see Theorem 3.12). It then follows that

$$\mathrm{I}_\phi(\varphi_j) = \mathrm{I}_\phi(\psi_j) + \sup_X \varphi_j \le \sup_X \varphi_j. \tag{22}$$

From (21) and (22) we arrive at

$$\lim_{j\to\infty} F(\varphi_j) \le \lim_{j\to\infty} (\sup_X \varphi_j - c e^{\sup_X \varphi_j}) = -\infty,$$

which is a contradiction. Thus $\sup_X \varphi_j$ is bounded from above as claimed. Since $F(\varphi_j) \le \mathrm{I}_\phi(\varphi_j) \le \sup_X \varphi_j$, it follows that $\mathrm{I}_\phi(\varphi_j)$ and hence $\sup_X \varphi_j$ is also bounded from below. It follows again from [Guedj and Zeriahi 2005, Proposition 2.7] that a subsequence of φ_j (still denoted by φ_j) converges in $L^1(X,\omega^n)$ to some $\varphi \in \mathrm{PSH}(X,\theta)$. Since I_ϕ is upper semicontinuous it follows that $\varphi \in \mathcal{E}^1(X,\theta,\phi)$. Moreover, by continuity of $u \mapsto \int_X e^u\,d\mu$ we get that $F(\varphi) \ge \sup_{\mathcal{E}^1(X,\theta,\phi)} F$. Hence φ maximizes F on $\mathcal{E}^1(X,\theta,\phi)$. Now Theorem 4.22 shows that φ solves the desired complex Monge–Ampère equation. The next lemma address the uniqueness question. □

Lemma 4.24. *Let* $\lambda > 0$. *Assume that* $\varphi \in \mathcal{E}(X,\theta,\phi)$ *is a solution of* (20) *while* $\psi \in \mathcal{E}(X,\theta,\phi)$ *satisfies* $\theta_\psi^n \ge e^{\lambda\psi}\mu$. *Then* $\varphi \ge \psi$ *on* X.

Proof. By the comparison principle for the class $\mathcal{E}(X,\theta,\phi)$ (Corollary 3.6) we have

$$\int_{\{\varphi<\psi\}}\theta_\psi^n \le \int_{\{\varphi<\psi\}}\theta_\varphi^n.$$

As φ is a solution and ψ is a subsolution to (20) we also have

$$\int_{\{\varphi<\psi\}} e^{\lambda\psi}\,d\mu \le \int_{\{\varphi<\psi\}}\theta_\psi^n \le \int_{\{\varphi<\psi\}}\theta_\varphi^n = \int_{\{\varphi<\psi\}} e^{\lambda\varphi}\,d\mu \le \int_{\{\varphi<\psi\}} e^{\lambda\psi}\,d\mu.$$

It follows that all inequalities above are equalities; hence $\varphi\ge\psi$ μ-almost everywhere on X. Since $\mu = e^{-\lambda\varphi}\theta_\varphi^n$, it follows that $\theta_\varphi^n(\{\varphi<\psi\})=0$. By the domination principle (Proposition 3.11) we get that $\varphi\ge\psi$ everywhere on X. □

4C3. *The case* $\lambda = 0$.

Theorem 4.25. *Assume that* $\mu\in\mathcal{M}_A$ *for some* $A\ge 1$. *Then there exists* $u\in\mathcal{E}^1(X,\theta,\phi)$ *such that* $\theta_u^n=\mu$.

Proof. In view of Theorem 4.22 it suffices to find a maximizer in $\mathcal{E}^1(X,\theta,\phi)$ of the functional $F := F_{0,\mu}$ defined by

$$F(u) := \mathrm{I}_\phi(u) - \int_X (u-\phi)\,d\mu, \quad u\in\mathcal{E}^1(X,\theta,\phi).$$

Note that $F(u)$ is finite for all $u\in\mathcal{E}^1(X,\theta,\phi)$ since $\mu\in\mathcal{M}_A$ (see Lemma 4.18). Let (u_j) be a sequence in $\mathcal{E}^1(X,\theta,\phi)$ such that $\sup_X u_j = 0$ and $F(u_j)$ increases to $\sup_{\mathcal{E}^1(X,\theta,\phi)} F > -\infty$. By the compactness property [Guedj and Zeriahi 2005], a subsequence of (u_j) converges to $u\in\mathrm{PSH}(X,\theta,\phi)$, and $\sup_X u = 0$. Moreover, since $\mu\in\mathcal{M}_A$, by (19) we have

$$F(u_j) \le \mathrm{I}_\phi(u_j) + C|\mathrm{I}_\phi(u_j)|^{1/2} + C \quad \text{for all } j.$$

It thus follows that $\mathrm{I}_\phi(u_j)$ is uniformly bounded. Since I_ϕ is upper semicontinuous it follows that $u\in\mathcal{E}^1(X,\theta,\phi)$. Also, since $\int_X(u_j-\phi)^2\,d\mu$ is uniformly bounded (Lemma 4.18) it follows from the same arguments as [Guedj and Zeriahi 2017, Lemma 11.5] that $\int_X(u_j-\phi)\,d\mu$ converges to $\int_X(u-\phi)\,d\mu$. Since I_ϕ is upper semicontinuous, we obtain that $F(u)\ge\limsup_j F(u_j)$. Hence u maximizes F on $\mathcal{E}^1(X,\theta,\phi)$, and the result follows. □

Lemma 4.26. *If* μ *is a positive nonpluripolar measure on* X *and* $A\ge 1$ *then there exists* $\nu\in\mathcal{M}_A$ *and* $0\le f\in L^1(X,\nu)$ *such that* $\mu = f\nu$.

The short proof given below is due to Cegrell [1998].

Proof. It follows from Lemma 4.17 that $\mathcal{M}_A$ is a convex compact subset of $\mathcal{M}(X)$, the space of probability measures on X. It follows from [König and Seever 1969, Lemma 1] that we can write

$$\mu = \nu + \sigma,$$

where ν,σ are nonnegative Borel measures on X such that ν is absolutely continuous with respect to an element in $\mathcal{M}_A$ and σ is singular with respect to any element of $\mathcal{M}_A$; i.e., $\sigma\perp m$ for any $m\in\mathcal{M}_A$. It then follows from [Rainwater 1969, Theorem] that σ is supported on a Borel set E such that $m(E)=0$ for

all $m \in \mathcal{M}_A$. If u is a candidate defining the capacity $\mathrm{Cap}_\phi(E)$, then clearly $\theta_u^n \in \mathcal{M}_A$; hence $\int_E \theta_u^n = 0$. It follows that $\mathrm{Cap}_\phi(E) = 0$; hence by Lemma 4.3 E is pluripolar. Therefore, $\sigma = 0$ since μ does not charge pluripolar sets. □

To prove the main existence result in this subsection we also need the following lemma. The argument uses the locality of nonpluripolar Monge–Ampère measures with respect to the plurifine topology, and is identical to the proof of [Guedj and Zeriahi 2007, Corollary 1.10].

Lemma 4.27. *Assume that ν is a positive nonpluripolar Borel measure on X and $u, v \in \mathrm{PSH}(X, \theta)$. If $\theta_u^n \geq \nu$ and $\theta_v^n \geq \nu$ then $\theta^n_{\max(u,v)} \geq \nu$.*

Theorem 4.28. *Assume that μ is a positive nonpluripolar measure on X such that $\mu(X) = \int_X \theta_\phi^n$. Then there exists $u \in \mathcal{E}(X, \theta, \phi)$ (unique up to a constant) such that $\theta_u^n = \mu$.*

Proof. It follows from Lemma 4.26 that $\mu = f\nu$, where $\nu \in \mathcal{M}_1$ and $0 \leq f \in L^1(X, \nu)$. For each j it follows from Theorem 4.25 that there exists $u_j \in \mathcal{E}^1(X, \theta, \phi)$ such that $\sup_X u_j = 0$ and

$$\theta_{u_j}^n = c_j \min(f, j)\nu.$$

Here, c_j is a normalization constant and $c_j \to 1$ as $j \to \infty$. We can assume that $1 \leq c_j \leq 2$ for all j. By compactness [Guedj and Zeriahi 2017, Proposition 8.5], a subsequence of (u_j) converges in $L^1(X, \omega^n)$ to $u \in \mathrm{PSH}(X, \theta, \phi)$ with $\sup_X u = 0$. We will show that $u \in \mathcal{E}(X, \theta, \phi)$. For each $k \in \mathbb{N}$ we set $v_k := (\sup_{j \geq k} u_j)^*$. Then $v_k \in \mathcal{E}^1(X, \theta, \phi)$ and (v_k) decreases pointwise to u. For each k fixed, and for all $j > k$ we have $\theta_{u_j}^n \geq \min(f, k)\nu$. Thus for all $\ell \in \mathbb{N}$ it follows from Lemma 4.27 that $\theta^n_{w_{k,\ell}} \geq \min(f, k)\nu$, where $w_{k,\ell} := \max(u_k, \ldots, u_{k+\ell})$. Since $(w_{k,\ell})$ increases almost everywhere to v_k as $\ell \to \infty$, it follows from Theorem 2.3 and Remark 2.5 that

$$\theta_{v_k}^n \geq \min(f, k)\nu.$$

Thus for each $C > 0$, setting $v_k^C := \max(v_k, V_\theta - C)$, using the plurifine property of the Monge–Ampère measure and observing that $\{u > V_\theta - C\} \subseteq \{v_k > V_\theta - C\}$, we have

$$\theta^n_{v_k^C} \geq \mathbb{1}_{\{v_k > V_\theta - C\}} \theta^n_{v_k} \geq \mathbb{1}_{\{v_k > V_\theta - C\}} \min(f, k)\nu \geq \mathbb{1}_{\{u > V_\theta - C\}} \min(f, k)\nu.$$

Since (v_k^C) decreases to $u^C := \max(u, V_\theta - C)$ and $v_k^C, u^C \in \mathcal{E}(X, \theta)$, it follows from Theorem 2.3 that $\theta^n_{v_k^C}$ converges weakly to $\theta^n_{u^C}$; hence

$$\theta^n_{u^C} \geq \mathbb{1}_{\{u > V_\theta - C\}} \mu.$$

Since μ is nonpluripolar, by letting $C \to \infty$ it follows that

$$\theta_u^n = \lim_{C \to \infty} \mathbb{1}_{\{u > V_\theta - C\}} \theta^n_{u^C} \geq \lim_{C \to \infty} \mathbb{1}_{\{u > V_\theta - C\}} \mu = \mu.$$

Moreover by [Witt Nyström 2017, Theorem 1.2] the total mass of θ_u^n is smaller than $\int_X \theta_\phi^n = \mu(X)$ since $u \leq \phi$. Hence $\int_X \theta_\phi^n = \mu(X) = \int_X \theta_u^n$. It thus follows that $u \in \mathcal{E}(X, \theta, \phi)$ and $\theta_u^n = \mu$. Uniqueness is addressed in the next theorem. □

Theorem 4.29. *Assume $u, v \in \mathcal{E}(X, \theta, \phi)$ are such that $\theta_u^n = \theta_v^n$. Then $u - v$ is constant.*

The proof of this uniqueness result rests on the adaptation of the mass concentration technique of Kołodziej and Dinew [2009b] to our more general setting; see also [Boucksom et al. 2010; Dinew and Lu 2015]. The arguments carry over almost verbatim, but as a courtesy to the reader we provide a detailed account.

Proof. Set $\mu := \theta_u^n = \theta_v^n$. We will prove that there exists a constant C such that μ is supported on $\{u = v + C\}$. This will allow us to apply the domination principle (Proposition 3.11) to ensure the conclusion. Assume that it is not the case. Arguing exactly as in [Boucksom et al. 2010, Section 3.3] we can assume that $0 < \mu(U) < \mu(X) = \int_X \theta_\phi^n$ and $\mu(\{u = v\}) = 0$, where $U := \{u < v\}$. Let $c > 1$ be a normalization constant such that $\int_{\{u<v\}} c^n \, d\mu = \mu(X)$. It follows from Theorem 4.28 that there exists $h \in \mathcal{E}(X, \theta, \phi)$, $\sup_X h = 0$, such that $\theta_h^n = c^n \mathbb{1}_U \mu$. In particular, $h \le \phi$. For each $t \in (0, 1)$ we set $U_t := \{(1-t)u + t\phi < (1-t)v + th\}$ and note that, since $h \le \phi$, the sets U_t increase as $t \to 0^+$ to $U \setminus \{h = -\infty\}$.

By the mixed Monge–Ampère inequalities [Boucksom et al. 2010, Proposition 1.11], which go back to [Dinew 2009a; Kołodziej 2003], we have

$$\theta_u^{n-1} \wedge \theta_h \ge \mathbb{1}_U c\mu, \quad \theta_u^k \wedge \theta_v^{n-k} \ge \mu, \quad k = 0, \dots, n. \tag{23}$$

Moreover, since $u, v, h \in \mathcal{E}(X, \theta, \phi)$, it follows from Corollary 3.15 that all the above nonpluripolar products have the same mass. Consequently, $\theta_u^k \wedge \theta_v^{n-k} = \mu$, $k = 0, \dots, n$. Using the partial comparison principle (Proposition 3.5) we can write

$$\int_{U_t} \theta_u^{n-1} \wedge \theta_{(1-t)v+th} \le \int_{U_t} \theta_u^{n-1} \wedge \theta_{(1-t)u+t\phi}.$$

Expanding, and using the fact that $\theta_u^n = \theta_u^{n-1} \wedge \theta_v$ we get

$$\int_{U_t} \theta_u^{n-1} \wedge \theta_h \le \int_{U_t} \theta_u^{n-1} \wedge \theta_\phi. \tag{24}$$

Combining (23) and (24) we have $c\mu(U_t) \le \int_{U_t} \theta_u^{n-1} \wedge \theta_h \le \int_{U_t} \theta_u^{n-1} \wedge \theta_\phi$. Letting $t \to 0$, and noting that μ is nonpluripolar (hence μ puts no mass on the set $\{h = -\infty\}$) we obtain

$$c\mu(U) \le \int_U \theta_u^{n-1} \wedge \theta_\phi.$$

Now, applying the same arguments for $V := \{u > v\}$ we obtain

$$b\mu(V) \le \int_V \theta_u^{n-1} \wedge \theta_\phi,$$

where $b > 1$ is a constant such that $b^n \mu(V) = \mu(X)$. Using that $\mu(\{u = v\}) = 0$, we can sum up the last two inequalities and obtain

$$0 < \min(b, c)\mu(X) \le \int_X \theta_u^{n-1} \wedge \theta_\phi = \mu(X),$$

where the last equality follows again from Corollary 3.15. This is a contradiction since $\min(b, c) > 1$. □

4D. ***Regularity of solutions.*** Recall that we work with $\phi \in \mathrm{PSH}(X,\theta)$ with small unbounded locus such that $P_\theta[\phi]=\phi$, and $\int_X \theta_\phi^n > 0$. Let $f \in L^p(\omega^n)$ with $f \geq 0$. In the previous subsection we have shown that the equation

$$\theta_\psi^n = f\omega^n, \quad \psi \in \mathcal{E}^1(X,\theta,\phi),$$

has a unique solution. In this subsection we will show that this solution has the same singularity type as ϕ. This generalizes [Boucksom et al. 2010, Theorem B], which treats the particular case of solutions with minimal singularities in a big class. Analogous results will be obtained for the solutions of (20) as well.

Our arguments will closely follow the path laid out in [Boucksom et al. 2010, Section 4.1], which builds on fundamental work of Kołodziej [1998; 2003] in the Kähler case. As we shall see, the fact that ϕ has model-type singularity plays a vital role in making sure that the methods of [Boucksom et al. 2010] work in our more general context as well.

We first prove that any measure with $L^{1+\varepsilon}$, $\varepsilon > 0$, density is dominated by the relative capacity:

Proposition 4.30. *Let $f \in L^p(\omega^n)$, $p > 1$, with $f \geq 0$. Then there exists $C > 0$ depending only on θ, ω, p and $\|f\|_{L^p}$ such that*

$$\int_E f\omega^n \leq \frac{C}{\left(\int_X \theta_\phi^n\right)^2} \cdot \mathrm{Cap}_\phi(E)^2$$

for all Borel sets $E \subset X$.

Proof. Since Cap_ϕ is inner regular we can assume that E is compact. Thanks to Lemma 4.8 we can also assume that $M_\phi(E) < \infty$.

We introduce $\nu_\theta := \sup_{T,x} \nu(T,x)$, where $x \in X$, T is any closed positive $(1,1)$-current cohomologous with θ, and $\nu(T,x)$ denotes the Lelong number of T at x. As a result, the uniform version of Skoda's integrability theorem [Guedj and Zeriahi 2017, Theorem 2.50] yields a constant $C > 0$, only depending on θ and ω such that $\int_X \exp(-\nu_\theta^{-1}\psi)\omega^n \leq C$ for all $\psi \in \mathrm{PSH}(X,\theta)$ with $\sup_X \psi = 0$. Applying this to $V_{E,\phi}^* - M_\phi(E)$ we get

$$\int_X \exp(-\nu_\theta^{-1} V_{E,\phi}^*)\omega^n \leq C \cdot \exp(-\nu_\theta^{-1} M_\phi(E)).$$

On the other hand, $V_{E,\phi}^* \leq 0$ on E a.e. with respect to Lebesgue measure; hence

$$\mathrm{Vol}_\omega(E) := \int_E \omega^n \leq C \cdot \exp(-\nu_\theta^{-1} M_\phi(E)). \tag{25}$$

An application of Hölder's inequality gives

$$\int_E f\omega^n \leq \|f\|_{L^p} \mathrm{Vol}_\omega(E)^{(p-1)/p}. \tag{26}$$

At this point we may assume that $M_\phi(E) \geq 1$. Indeed, if this were not the case, then Lemma 4.9 would imply that $\mathrm{Cap}_\phi(E) = \int_X \theta_\phi^n$, yielding the desired estimate of the proposition. Putting together Lemma 4.9,

(25) and (26) we get

$$\int_E f\omega^n \le C^{p-1/p}\cdot\|f\|_{L^p}\cdot\exp\left(-\frac{p-1}{pv_\theta}\left(\frac{\mathrm{Cap}_\phi(E)}{\int_X\theta_\phi^n}\right)^{-1/n}\right).$$

The result now follows, as $\exp(-t^{-1/n}) = O(t^2)$ when $t\to 0_+$. □

Before we state the main result of this subsection, we need one last lemma, which is a simple consequence of our comparison principle:

Lemma 4.31. *Let $u\in\mathcal{E}(X,\theta,\phi)$. Then for all $t>0$ and $\delta\in(0,1]$ we have*

$$\mathrm{Cap}_\phi\{u<\phi-t-\delta\}\le\frac{1}{\delta^n}\int_{\{u<\phi-t\}}\theta_u^n.$$

Proof. Let $\psi\in\mathrm{PSH}(X,\theta,\phi)$ be such that $\phi\le\psi\le\phi+1$. In particular, note that $\psi\in\mathcal{E}(X,\theta,\phi)$. We then have

$$\{u<\phi-t-\delta\}\subset\{u<\delta\psi+(1-\delta)\phi-t-\delta\}\subset\{u<\phi-t\}.$$

Since $\delta^n\theta_\psi^n\le\theta_{\delta\psi+(1-\delta)\phi}^n$, u has relative full mass and $\mathcal{E}(X,\theta,\phi)$ is convex, Corollary 3.6 yields

$$\delta^n\int_{\{u<\phi-t-\delta\}}\theta_\psi^n\le\int_{\{u<\delta\psi+(1-\delta)\phi-t-\delta\}}\theta_{\delta\psi+(1-\delta)\phi}^n\le\int_{\{u<\delta\psi+(1-\delta)\phi-t-\delta\}}\theta_u^n\le\int_{\{u<\phi-t\}}\theta_u^n.$$

Since ψ is an arbitrary candidate in the definition of Cap_ϕ, the proof is complete. □

We arrive at the main results of this subsection:

Theorem 4.32. *Suppose $\phi=P_\theta[\phi]$ has small unbounded locus and $\int_X\theta_\phi^n>0$. Let also $\psi\in\mathcal{E}(X,\theta,\phi)$ with $\sup_X\psi=0$. If $\theta_\psi^n=f\omega^n$ for some $f\in L^p(\omega^n)$, $p>1$, then ψ has the same singularity type as ϕ; more precisely,*

$$\phi-C\left(\|f\|_{L^p},p,\omega,\theta,\int_X\theta_\phi^n\right)\le\psi\le\phi.$$

Proof. To begin, we introduce the function

$$g(t):=(\mathrm{Cap}_\phi\{\psi<\phi-t\})^{1/n},\quad t\ge 0.$$

We will show that $g(M)=0$ for some M under control. By Lemma 4.3 we will then have $\psi\ge\phi-M$ a.e. with respect to ω^n, which then implies $\psi\ge\phi-M$ on X.

Since $\theta_\psi^n=f\omega^n$, it follows from Proposition 4.30 and Lemma 4.31 that

$$g(t+\delta)\le\frac{C^{1/n}}{\delta}g(t)^2,\quad t>0,\quad 0<\delta<1.$$

Consequently, we can apply [Eyssidieux et al. 2009, Lemma 2.3] to conclude that $g(M)=0$ for $M:=t_0+2$. As an important detail, the constant $t_0>0$ has to be chosen so that

$$g(t_0)<\frac{1}{2C^{1/n}}.$$

On the other hand, Lemma 4.31 (with $\delta = 1$) implies

$$g(t+1)^n \le \int_{\{\psi < \phi - t - 1\}} f\omega^n \le \frac{1}{t+1}\int_X |\phi - \psi| f\omega^n \le \frac{1}{t+1}\|f\|_{L^p}(\|\psi\|_{L^q} + \|\phi\|_{L^q}),$$

where in the last estimate we used Hölder's inequality with $q = p/(p-1)$. Since ψ and ϕ both belong to the compact set of θ-psh functions normalized by $\sup_X u = 0$, their L^q norms are bounded by an absolute constant only depending on θ, ω and p. Consequently, it is possible to choose t_0 to be only dependent on $\|f\|_{L^p}$, θ, ω, $\int_X \theta_\phi^n$ and p, finishing the proof. □

Corollary 4.33. *Suppose $\phi = P_\theta[\phi]$ has small unbounded locus and $\int_X \theta_\phi^n > 0$. If $\lambda > 0$ and, $\psi \in \mathcal{E}(X, \theta, \phi)$, $\theta_\psi^n = e^{\lambda\psi} f\omega^n$ for some $f \in L^p(\omega^n)$, $p > 1$, then ψ has the same singularity type as ϕ.*

Proof. Since ψ is bounded from above on X and $\lambda > 0$, it follows that $e^{\lambda\psi} f \in L^p(X, \omega^n)$, $p > 1$. The result follows from Theorem 4.32. □

4E. ***Naturality of model-type singularities and examples.*** Our readers may still wonder if our choice of model potentials is a natural one in the discussion of complex Monge–Ampère equations with prescribed singularity. We hope to address the doubts in the next result.

Theorem 4.34. *Suppose $\psi \in \mathrm{PSH}(X, \theta)$ has small unbounded locus and the equation*

$$\theta_u^n = f\omega^n$$

has a solution $u \in \mathrm{PSH}(X, \theta)$ with the same singularity type as ψ for all $f \in L^\infty$, $f \ge 0$, satisfying $\int_X \theta_\psi^n = \int_X f\omega^n > 0$. Then ψ has model-type singularity.

Proof. Our simple proof follows the guidelines of the example described in the beginning of Section 4. Indeed, suppose that $[\psi]$ is not of model type. Then $P_\theta[\psi]$ is strictly less singular than ψ, but of course $\mathcal{E}(X, \theta, \psi) \subset \mathcal{E}(X, \theta, P_\theta[\psi])$, as $\int_X \theta_\psi^n = \int_X \theta_{P_\theta[\psi]}^n$.

By Theorem 3.8, there exists $g \in L^\infty$ such that $\theta_{P_\theta[\psi]}^n = g\omega^n$. By the uniqueness theorem (Theorem 4.29), $P_\theta[\psi]$ is the only solution of this last equation inside $\mathcal{E}(X, \theta, P_\theta[\psi])$.

Since $\mathcal{E}(X, \theta, \psi) \subset \mathcal{E}(X, \theta, P_\theta[\psi])$, but $P_\theta[\psi] \notin \mathcal{E}(X, \theta, \psi)$, we get that $\theta_u^n = g\omega^n$ cannot have any solution that has the same singularity type as ψ. □

Next we point out a simple way to construct model singularity types:

Proposition 4.35. *Suppose that $\psi \in \mathrm{PSH}(X, \theta)$ has small unbounded locus and $\theta_\psi^n = f\omega^n$ for some $f \in L^p(\omega^n)$, $p > 1$, with $\int_X f\omega^n > 0$. Then ψ has model-type singularity.*

Proof. We first observe that $\psi \in \mathcal{E}(X, \theta, P_\theta[\psi])$. Since θ_ψ^n has L^p density with $p > 1$, it thus follows from Theorem 4.32 that $\psi - P_\theta[\psi]$ is bounded on X; hence $[\psi] = [P_\theta[\psi]]$, implying that ψ has model-type singularity. □

Using this simple proposition, one can show that all analytic singularity types are of model type, which was previously known to be true using algebraic methods; see [Ross and Witt Nyström 2014; Rashkovskii and Sigurdsson 2005]:

Proposition 4.36. *Suppose $\psi \in \mathrm{PSH}(X,\theta)$ has analytic singularity type; i.e., ψ can be locally written as $c\log\big(\sum_j |f_j|^2\big)+g$, where f_j are holomorphic, $c>0$ and g is smooth. Then $[\psi]$ is of model type.*

Proof. We can assume that our fixed Kähler form ω satisfies $\omega \geq 2\theta$. Since $P_\theta[\psi] \leq P_\omega[\psi]$, it suffices to prove that $\psi - P_\omega[\psi]$ is globally bounded on X. In fact we will prove the following stronger result:

$$\rho := \frac{\omega_\psi^n}{\omega^n} \in L^p(\omega^n) \quad \text{for some } p>1. \tag{27}$$

As $\omega/2 \geq \theta$ it follows that $\int_X \omega_\psi^n \geq 2^{-n}\int_X \omega^n > 0$; hence Proposition 4.35 will imply that $\psi - P_\omega[\psi]$ is globally bounded on X.

We now prove (27). Since X is compact it suffices to prove that there exists a small open neighborhood U around a given point $x \in X$ (which will be fixed) such that $\rho \in L^p(U, dV)$ for some $p>1$. Since ψ has analytic singularities we can find a holomorphic coordinate chart Ω around x such that

$$\psi = c\log\sum_{j=1}^N |f_j|^2 + g$$

in a neighborhood of Ω, where $c>0$ is a constant, f_j are holomorphic functions in Ω and g is a smooth real-valued function in Ω. Let $A>0$ be large enough so that $(A-1)\omega + i\partial\bar\partial g \geq 0$ in Ω.

In $X\setminus\{\psi=-\infty\}$, since ψ is smooth we can write $\omega_\psi^n = \rho\omega^n$, where $\rho\geq 0$ is smooth. We extend ρ to be 0 over the set $\{\psi=-\infty\}$. Then $\rho\omega^n$ is the nonpluripolar Monge–Ampère measure of ψ with respect to ω as follows from [Boucksom et al. 2010]; hence

$$\int_\Omega \rho\omega^n \leq \int_X \rho\omega^n \leq \int_X \omega^n.$$

Similarly we can write $(A\omega + i\partial\bar\partial\psi)^n = \rho_A\omega^n$ in $\Omega\setminus\{\psi=-\infty\}$, where $0\leq\rho_A\in L^1(\Omega, dV)$.

Now, we carry out the computation in $\Omega\setminus\{\psi=-\infty\}$. For notational convenience we set $h := \sum_{j=1}^N |f_j|^2$, $\varphi := \log\sum_{j=1}^N |f_j|^2$ and we compute $i\partial\bar\partial\varphi$:

$$i\partial\bar\partial\varphi = \frac{\sum_{j=1}^N i\partial f_j\wedge\overline{\partial f}_j}{h} - \frac{i\big(\sum_{j=1}^N \bar f_j\partial f_j\big)\wedge\big(\sum_{j=1}^N f_j\overline{\partial f}_j\big)}{h^2}.$$

For each $1\leq j<k\leq N$ we set $\alpha_{j,k} := f_j\partial f_k - f_k\partial f_j$. Then we obtain

$$i\partial\bar\partial\varphi = h^{-2}\sum_{j<k} i\alpha_{j,k}\wedge\bar\alpha_{j,k}. \tag{28}$$

Let $C>0$ be large enough such that $C^{-1}\beta \leq A\omega + i\partial\bar\partial g \leq C\beta$ in Ω, where β is the standard Kähler form in $\mathbb{C}^n$. For each $\ell=0,\dots,n$, set $\gamma_l := (i\partial\bar\partial\varphi)^\ell\wedge\beta^{n-\ell}$. Then there exists a constant $B>1$ (depending on c, $C>0$) such that in $\Omega\setminus\{\psi=-\infty\}$ one has

$$\frac{1}{B}\sum_{\ell=0}^n\gamma_\ell = \frac{1}{B}\sum_{\ell=0}^n (i\partial\bar\partial\varphi)^\ell\wedge\beta^{n-\ell} \leq (A\omega+i\partial\bar\partial\psi)^n \leq B\sum_{p=0}^n (i\partial\bar\partial\varphi)^\ell\wedge\beta^{n-\ell} = B\sum_{\ell=0}^n\gamma_\ell. \tag{29}$$

By the definition of $\alpha_{j,k}$ it follows that the $(\ell, 0)$-forms $\alpha_{j_1,k_1} \wedge \cdots \wedge \alpha_{j_\ell,k_\ell}$ are of the type $\sum F_k\, dz_{I_k}$, where $|I_k| = \ell$, and each F_k is holomorphic in Ω. By the above identity in (28), each γ_ℓ is the sum of (n, n)-forms of type $|F|^2 h^{-2\ell}\beta^n$, where F is holomorphic in Ω. By the first estimate in (29) it follows that for each ℓ,

$$\int_\Omega |F|^2 h^{-2\ell}\beta^n \le B\int_\Omega \rho_A \omega^n < \infty;$$

hence $|F|^2 e^{-2\ell \log h}$ is integrable in Ω. From the resolution of Demailly's strong openness conjecture [2001] due to Guan and Zhou [2015] (see also [Hiep 2014] for an alternative proof) it follows that each $|F|^2 h^{-2\ell}$ is in $L^p(U, dV)$ for some $p > 1$ and a smaller neighborhood $U \subset \Omega$ of x. Finally, from the second estimate in (29) we see that $\omega_\psi^n/\omega^n \in L^p(U, dV)$, which what we wanted. □

5. Log-concavity of nonpluripolar products

Theorem 5.1. *Let $T_1, \ldots, T_n$ be positive $(1, 1)$-currents on a compact Kähler manifold X. Assume that each T_j has potential with small unbounded locus. Then*

$$\int_X \langle T_1 \wedge \cdots \wedge T_n\rangle \ge \left(\int_X \langle T_1^n\rangle\right)^{1/n} \cdots \left(\int_X \langle T_n^n\rangle\right)^{1/n}.$$

Proof. We can assume that the classes of T_j are big and their masses are nonzero. Otherwise the right-hand side of the inequality to be proved is zero. Consider smooth closed real $(1, 1)$-forms θ^j, and $u_j \in \mathrm{PSH}(X, \theta^j)$ with small unbounded locus such that $T_j = \theta_{u_j}^j$.

For each $j = 1, \ldots, n$, Theorem 4.28 ensures that there exists a normalizing constant $c_j > 0$ and $\varphi_j \in \mathcal{E}(X, \theta^j, P_\theta[u_j])$ such that $(\theta_{\varphi_j}^j)^n = c_j\omega^n$.

We can assume that $\int_X \omega^n = 1$; thus we can write

$$c_j = \int_X (\theta_{\varphi_j}^j)^n = \int_X (\theta_{P_\theta[u_j]}^j)^n = \int_X (\theta_{u_j}^j)^n = \int_X \langle T_j^n\rangle.$$

A combination of Proposition 2.1 and Theorem 2.3 then gives

$$\int_X \theta_{\varphi_1}^1 \wedge \cdots \wedge \theta_{\varphi_n}^n = \int_X \theta_{P_\theta[u_1]}^1 \wedge \cdots \wedge \theta_{P_\theta[u_n]}^n = \int_X \theta_{u_1}^1 \wedge \cdots \wedge \theta_{u_n}^n = \int_X \langle T_1 \wedge \cdots \wedge T_n\rangle.$$

An application of [Boucksom et al. 2010, Proposition 1.11] gives that $\theta_{\varphi_1}^1 \wedge \cdots \wedge \theta_{\varphi_n}^n \ge c_1^{1/n} \cdots c_n^{1/n}\omega^n$. The result follows after we integrate this estimate. □

Acknowledgements

Darvas was partially supported by BSF grant 2012236 and NSF grant DMS–1610202. Di Nezza was supported by a Marie Skłodowska Curie individual fellowship 660940–KRF–CY (MSCA–IF).

We would like to thank Robert Berman, Mattias Jonsson and László Lempert for their insightful comments that improved the presentation of the paper. We warmly thank the referees for many suggestions improving the presentation of the paper.

References

[Aubin 1978] T. Aubin, "Équations du type Monge–Ampère sur les variétés kählériennes compactes", *Bull. Sci. Math.* (2) **102**:1 (1978), 63–95. MR Zbl

[Bedford and Demailly 1988] E. Bedford and J.-P. Demailly, "Two counterexamples concerning the pluri-complex Green function in $\mathbb{C}^n$", *Indiana Univ. Math. J.* **37**:4 (1988), 865–867. MR Zbl

[Bedford and Taylor 1976] E. Bedford and B. A. Taylor, "The Dirichlet problem for a complex Monge–Ampère equation", *Invent. Math.* **37**:1 (1976), 1–44. MR Zbl

[Bedford and Taylor 1982] E. Bedford and B. A. Taylor, "A new capacity for plurisubharmonic functions", *Acta Math.* **149**:1-2 (1982), 1–40. MR Zbl

[Bedford and Taylor 1987] E. Bedford and B. A. Taylor, "Fine topology, Šilov boundary, and $(dd^c)^n$", *J. Funct. Anal.* **72**:2 (1987), 225–251. MR Zbl

[Berman 2013] R. J. Berman, "From Monge-Ampere equations to envelopes and geodesic rays in the zero temperature limit", preprint, 2013. arXiv

[Berman and Boucksom 2010] R. Berman and S. Boucksom, "Growth of balls of holomorphic sections and energy at equilibrium", *Invent. Math.* **181**:2 (2010), 337–394. MR Zbl

[Berman et al. 2013] R. J. Berman, S. Boucksom, V. Guedj, and A. Zeriahi, "A variational approach to complex Monge–Ampère equations", *Publ. Math. Inst. Hautes Études Sci.* **117** (2013), 179–245. 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

[Cegrell 1998] U. Cegrell, "Pluricomplex energy", *Acta Math.* **180**:2 (1998), 187–217. MR Zbl

[Coman and Guedj 2009] D. Coman and V. Guedj, "Quasiplurisubharmonic Green functions", *J. Math. Pures Appl.* (9) **92**:5 (2009), 456–475. MR Zbl

[Darvas 2017] T. Darvas, "Weak geodesic rays in the space of Kähler potentials and the class $\mathcal{E}(X,\omega)$", *J. Inst. Math. Jussieu* **16**:4 (2017), 837–858. MR Zbl

[Darvas and Rubinstein 2017] T. Darvas and Y. A. Rubinstein, "Tian's properness conjectures and Finsler geometry of the space of Kähler metrics", *J. Amer. Math. Soc.* **30**:2 (2017), 347–387. MR Zbl

[Darvas et al. 2018] T. Darvas, E. Di Nezza, and C. H. Lu, "On the singularity type of full mass currents in big cohomology classes", *Compos. Math.* **154**:2 (2018), 380–409. MR Zbl arXiv

[Demailly 2001] J.-P. Demailly, "Multiplier ideal sheaves and analytic methods in algebraic geometry", pp. 1–148 in *School on Vanishing Theorems and Effective Results in Algebraic Geometry* (Trieste, 2000), edited by J. P. Demailly et al., ICTP Lect. Notes **6**, Abdus Salam Int. Cent. Theoret. Phys., Trieste, 2001. MR Zbl

[Di Nezza and Lu 2015] E. Di Nezza and C. H. Lu, "Generalized Monge–Ampère capacities", *Int. Math. Res. Not.* **2015**:16 (2015), 7287–7322. MR Zbl

[Di Nezza and Lu 2017] E. Di Nezza and C. H. Lu, "Complex Monge–Ampère equations on quasi-projective varieties", *J. Reine Angew. Math.* **727** (2017), 145–167. MR Zbl

[Dinew 2009a] S. Dinew, "An inequality for mixed Monge–Ampère measures", *Math. Z.* **262**:1 (2009), 1–15. MR Zbl

[Dinew 2009b] S. Dinew, "Uniqueness in $\mathcal{E}(X,\omega)$", *J. Funct. Anal.* **256**:7 (2009), 2113–2122. MR Zbl

[Dinew and Lu 2015] S. Dinew and C. H. Lu, "Mixed Hessian inequalities and uniqueness in the class $\mathcal{E}(X,\omega,m)$", *Math. Z.* **279**:3-4 (2015), 753–766. MR Zbl

[Eyssidieux et al. 2009] P. Eyssidieux, V. Guedj, and A. Zeriahi, "Singular Kähler–Einstein metrics", *J. Amer. Math. Soc.* **22**:3 (2009), 607–639. MR Zbl

[Guan 1998] B. Guan, "The Dirichlet problem for complex Monge–Ampère equations and regularity of the pluri-complex Green function", *Comm. Anal. Geom.* **6**:4 (1998), 687–703. MR Zbl

[Guan and Zhou 2015] Q. Guan and X. Zhou, "A proof of Demailly's strong openness conjecture", *Ann. of Math.* (2) **182**:2 (2015), 605–616. MR Zbl

[Guedj and Zeriahi 2005] V. Guedj and A. Zeriahi, "Intrinsic capacities on compact Kähler manifolds", *J. Geom. Anal.* **15**:4 (2005), 607–639. MR Zbl

[Guedj and Zeriahi 2007] V. Guedj and A. Zeriahi, "The weighted Monge–Ampère energy of quasiplurisubharmonic functions", *J. Funct. Anal.* **250**:2 (2007), 442–482. MR Zbl

[Guedj and Zeriahi 2017] V. Guedj and A. Zeriahi, *Degenerate complex Monge–Ampère equations*, EMS Tracts in Mathematics **26**, European Mathematical Society, Zürich, 2017. MR Zbl

[Guedj et al. 2017] V. Guedj, C. H. Lu, and A. Zeriahi, "Plurisubharmonic envelopes and supersolutions", preprint, 2017. To appear in *J. Differential Geom.* arXiv

[Hiep 2014] P. H. Hiep, "The weighted log canonical threshold", *C. R. Math. Acad. Sci. Paris* **352**:4 (2014), 283–288. MR Zbl

[Kołodziej 1998] S. Kołodziej, "The complex Monge–Ampère equation", *Acta Math.* **180**:1 (1998), 69–117. MR Zbl

[Kołodziej 2003] S. Kołodziej, "The Monge–Ampère equation on compact Kähler manifolds", *Indiana Univ. Math. J.* **52**:3 (2003), 667–686. MR Zbl

[König and Seever 1969] H. König and G. L. Seever, "The abstract F. and M. Riesz theorem", *Duke Math. J.* **36** (1969), 791–797. MR Zbl

[Lempert 1983] L. Lempert, "Solving the degenerate complex Monge–Ampère equation with one concentrated singularity", *Math. Ann.* **263**:4 (1983), 515–532. MR Zbl

[Phong and Sturm 2010a] D. H. Phong and J. Sturm, "The Dirichlet problem for degenerate complex Monge–Ampere equations", *Comm. Anal. Geom.* **18**:1 (2010), 145–170. MR Zbl

[Phong and Sturm 2010b] D. H. Phong and J. Sturm, "Regularity of geodesic rays and Monge–Ampère equations", *Proc. Amer. Math. Soc.* **138**:10 (2010), 3637–3650. MR Zbl

[Phong and Sturm 2014] D. H. Phong and J. Sturm, "On the singularities of the pluricomplex Green's function", pp. 419–435 in *Advances in analysis: the legacy of Elias M. Stein*, edited by C. Fefferman et al., Princeton Math. Ser. **50**, Princeton Univ. Press, 2014. MR Zbl

[Rainwater 1969] J. Rainwater, "A note on the preceding paper", *Duke Math. J.* **36** (1969), 799–800. MR Zbl

[Rashkovskii and Sigurdsson 2005] A. Rashkovskii and R. Sigurdsson, "Green functions with singularities along complex spaces", *Internat. J. Math.* **16**:4 (2005), 333–355. MR Zbl

[Ross and Witt Nyström 2014] J. Ross and D. Witt Nyström, "Analytic test configurations and geodesic rays", *J. Symplectic Geom.* **12**:1 (2014), 125–169. MR Zbl

[Ross and Witt Nyström 2017] J. Ross and D. Witt Nyström, "Envelopes of positive metrics with prescribed singularities", *Ann. Fac. Sci. Toulouse Math.* (6) **26**:3 (2017), 687–728. MR Zbl

[Witt Nyström 2017] D. Witt Nyström, "Monotonicity of non-pluripolar Monge–Ampère masses", preprint, 2017. arXiv

[Xing 1996] Y. Xing, "Continuity of the complex Monge–Ampère operator", *Proc. Amer. Math. Soc.* **124**:2 (1996), 457–467. MR Zbl

[Xing 2009] Y. Xing, "Continuity of the complex Monge–Ampère operator on compact Kähler manifolds", *Math. Z.* **263**:2 (2009), 331–344. 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 27 Jun 2017. Revised 4 Feb 2018. Accepted 10 Apr 2018.

TAMÁS DARVAS: tdarvas@math.umd.edu
Department of Mathematics, University of Maryland, College Park, MD, United States

ELEONORA DI NEZZA: dinezza@ihes.fr
Institut des Hautes Études Scientifiques, Université Paris-Saclay, Bures sur Yvette, France

CHINH H. LU: hoang-chinh.lu@u-psud.fr
Laboratoire de Mathématiques d'Orsay, Université Paris-Sud, CNRS, Université Paris-Saclay, Orsay, France

ANALYSIS AND PDE
Vol. 11, No. 8, 2018
dx.doi.org/10.2140/apde.2018.11.2089

msp

ON WEAK WEIGHTED ESTIMATES OF THE MARTINGALE TRANSFORM AND A DYADIC SHIFT

FEDOR NAZAROV, ALEXANDER REZNIKOV, VASILY VASYUNIN AND ALEXANDER VOLBERG

We consider weak-type estimates for several singular operators using the Bellman-function approach. In particular, we consider a concrete dyadic shift. We disprove the A_1 conjecture for those operators, which stayed open after Muckenhoupt and Wheeden's conjecture was disproved by Reguera and Thiele.

1. End-point estimates: notation and facts

The end-point estimates play an important part in the theory of singular integrals (weighted and unweighted). They are usually the most difficult estimates in the theory, and the most interesting of course. It is a general principle that one can extrapolate the estimate from the end-point situation to all other situations. We refer the reader to [Cruz-Uribe et al. 2011], which treats this subject of extrapolation in depth.

On the other hand, it happens quite often that the singular integral estimates exhibit a certain "blow-up" near the end point. Catching this blow-up can be a difficult task. We demonstrate this hunt for blow-ups by examples of weighted dyadic singular integrals and their behavior in $L^p(w)$. The end-point p will be naturally 1 (and sometimes slightly unnaturally 2) depending on the martingale singular operator. The singular integrals in this article are the easiest possible. They are dyadic martingale operators on the σ-algebra generated by the usual homogeneous dyadic lattice on the real line. We do not consider any nonhomogeneous situations, and this standard σ-algebra generated by a dyadic lattice $\mathcal{D}$ will be provided with Lebesgue measure.

Our goal will be to show how the Bellman-function technique gives the proof of the blow-up of the weighted estimates of the corresponding weighted dyadic singular operators. This blow-up will be demonstrated by certain estimates from below of the Bellman function of a dyadic problem.

The Bellman-function part will be reduced to the task of finding the lower estimate for the solutions of the concrete Monge–Ampère differential equation with concrete first-order terms (drift).

We will get a logarithmic blow-up not only for the martingale transform but also for a concrete dyadic shift; see our main result, Theorem 2.2.

Nazarov is partially supported by the NSF grant DMS 1265623. Vasyunin is partially supported by the RFBR grant 16-01-00635. Volberg is partially supported by the NSF grants DMS 1265549, DMS 1600065, and by the Hausdorff Institute for Mathematics, Bonn, Germany. Volberg and Vasyunin are partially supported by the Oberwolfach Institute for Mathematics, Germany. Reznikov is partially supported by NSF grant DMS 1764398.

MSC2010: primary 42A45, 42A61, 42B20, 42B35, 42B37, 47A30; secondary 42A50, 49L20, 49L25.

Keywords: martingale transform, weak weighted estimate.

2. End-point estimates for the martingale transform

Our measure space throughout this article will be $(X, \mathfrak{A}, dx)$, where the σ-algebra $\mathfrak{A}$ is generated by a standard dyadic filtration $\mathcal{D} = \bigcup_k \mathcal{D}_k$ on $\mathbb{R}$. We consider the martingale transform and dyadic shifts related to this homogeneous dyadic filtration.

As always, the symbol $\langle f \rangle_I$ denotes the average value of f over the set I; i.e., $\langle f \rangle_I = (1/|I|) \int_I f\, dx$. We consider martingale differences (recall that the symbol $\mathrm{ch}(J)$ denotes the dyadic children of J)

$$\Delta_J f := \sum_{I \in \mathrm{ch}(J)} \mathbf{1}_I (\langle f \rangle_I - \langle f \rangle_J).$$

For our case of the dyadic lattice on the line we have that $|\Delta_J f|$ is constant on J, and

$$\Delta_J f = \tfrac{1}{2}[(\langle f \rangle_{J_+} - \langle f \rangle_{J_-})\mathbf{1}_{J_+} + (\langle f \rangle_{J_-} - \langle f \rangle_{J_+})\mathbf{1}_{J_-}].$$

In this section and in the next one we consider the dyadic A_1 and A_2 classes of weights, but we skip the word dyadic, because we consider here only dyadic operators. We consider a positive function $w(x)$, and as before we call it an A_2 weight if

$$Q := [w]_{A_2} := \sup_{J \in \mathcal{D}} \langle w \rangle_J \langle w^{-1} \rangle_J < \infty. \tag{2-1}$$

We call w an A_1 weight if

$$Q := [w]_{A_1} := \sup_{J \in \mathcal{D}} \frac{\langle w \rangle_J}{\inf_J w} < \infty. \tag{2-2}$$

By Mw we will denote the martingale maximal function of w; that is, $Mw(x) = \sup_{x \in J,\, J \in \mathcal{D}} \langle w \rangle_J$. Then $w \in A_1$ with "norm" Q means that

$$Mw \le Q \cdot w \quad a.e.,$$

and $Q = [w]_{A_1}$ is the best constant in this inequality.

Recall that the martingale transform is the operator given by $T\varphi = \sum_{J \in \mathcal{D}} \varepsilon_J \Delta_J \varphi$. It is convenient to use the Haar function h_J associated with the dyadic interval J,

$$h_J(x) := \begin{cases} 1/|J|^{1/2}, & x \in J_+, \\ -1/|J|^{1/2}, & x \in J_-. \end{cases}$$

In this notation, the martingale transform is

$$T\varphi = \sum_{J \in cD} \varepsilon_J (\varphi, h_J) h_J,$$

where we (1) always assume the sum has an unspecified but finite number of terms, and (2) $|\varepsilon_J| \le 1$.

We are interested in several weak-type estimates.

We first consider the weak estimate for the martingale transform T in the weighted space $L^1(\mathbb{R}, w\, dx)$, where $w \in A_1$. The end-point exponent is naturally $p = 1$, and we wish to understand the order of

magnitude of the constant $A([w]_{A_1})$ in the weak-type inequality for the dyadic martingale transform:

$$\frac{1}{|I|}w\left\{x \in I : \sum_{J\in\mathcal{D}(I)} \varepsilon_J(\varphi, h_J)h_J(x) > \lambda\right\} \le C_{[w]_{A_1}} \frac{\langle|\varphi|w\rangle_I}{\lambda}. \tag{2-3}$$

Here φ runs over all functions such that $\operatorname{supp}\varphi \subset I$ and $\varphi \in L^1(I, w\,dx)$, $w \in A_1$. This section will be devoted to the study of the "sharp" order of magnitude of constants $C_{[w]_{A_1}}$ in terms of $[w]_{A_1}$ if $[w]_{A_1}$ is large. We are primarily interested in the estimate of $C_{[w]_{A_1}}$ from below, that is, in finding the worst possible A_1 weight in terms of weak-type estimates (of course this involves also finding the worst test function φ as well).

We will prove the following result.

Theorem 2.1. *There is a positive absolute constant c and a weight $w \in A_1$ with $[w]_{A_1}$ as large as we wish such that constant $C_{[w]_{A_1}}$ from (2-3) satisfies*

$$C_{[w]_{A_1}} \ge c[w]_{A_1}(\log[w]_{A_1})^{1/3}.$$

In fact, we will prove a sharper result. We will consider a particular dyadic shift, and we will prove the estimate $\ge c[w]_{A_1}(\log[w]_{A_1})^{1/3}$ for one particular dyadic shift. Ours is the following dyadic singular operator on $L^1(I, w\,dx)$, $I = [0, 1]$:

$$S\mathbf{1}_I = 0, \quad Sh_J = h_{J_-} - h_{J_+}, \quad J \in D(I).$$

Our main result is the following theorem.

Theorem 2.2. *There is a positive absolute constant c and a weight $w \in A_1$ such that*

$$\|S\|_{L^1(w)\to L^{1,\infty}(w)} \ge c[w]_{A_1}(\log[w]_{A_1})^{1/3}.$$

In [Lerner et al. 2009] the following estimate from above was proved:

Theorem 2.3. *There is a positive absolute constant C such that for any weight $w \in A_1$ the constant $C_{[w]_{A_1}}$ from (2-3) satisfies*

$$C_{[w]_{A_1}} \le c[w]_{A_1}\log[w]_{A_1}.$$

Remark 2.4. The sharp power remained enigmatic for quite a while. Very recently it was proved that for the Hilbert transform the exponent turns out to be 1 [Lerner et al. 2017]. However, it seems to be very probable that at the end-point of the scale, all operators behave differently, and the estimate for the dyadic shift S or the martingale transform might be different from the one for the Hilbert transform. A recent preprint [Ivanisvili and Volberg 2017] shows that the sharp power is actually 1 for the martingale transform as well.

Remark 2.5. This note is based on two preprints [Nazarov et al. 2015; 2016], but Theorem 2.2 was not formulated in these preprints; however, as the attentive reader can notice, it was proved there.

2A. ***Bellman approach: the Bellman function of the weak weighted estimate of the martingale transform and its properties.*** To find the "optimal" $C_{[w]_{A_1}}$ we use again the Bellman-function technique. The idea is to reformulate the infinite-dimensional problem of optimization of $C_{[w]_{A_1}}$, that is, finding the

"smallest" $C_{[w]_{A_1}}$ that works for all inequalities (2-3), in terms of the growth estimate on a certain function of only a finite number of variables (five in this case).

The Bellman function will depend on the number $Q \geq 1$ and is given by

$$\boldsymbol{B}(F, w, m, f, \lambda) := \boldsymbol{B}_Q(F, w, m, f, \lambda) := \sup \frac{1}{|I|}\omega\left\{x \in I : \sum_{J \subseteq I, J \in D} \varepsilon_J(\varphi, h_J)h_J(x) > \lambda\right\}, \tag{2-4}$$

where the sup is taken over all ε_J, $|\varepsilon_J| \leq 1$, $J \in D(I)$, and over all $\varphi \in L^1(I, \omega\, dx)$ such that $F := \langle|\varphi|\,\omega\rangle_I$, $f := \langle\varphi\rangle_I$, $w = \langle\omega\rangle_I$, $m \leq \inf_I \omega$, and ω are all dyadic A_1 weights such that $[w]_{A_1} \leq Q$. This function is obviously defined in the convex subdomain of $\mathbb{R}^5$

$$\Omega := \{(F, w, m, f, \lambda) \in \mathbb{R}^5 : F \geq |f|\, m,\ m \leq w \leq Qm\}. \tag{2-5}$$

Remark 2.6. We warn the reader that emotional attachment to the notation F, f, w for functions should be forgotten. These symbols in this and the following sections stand for numbers.

2A1. *The properties of* $\boldsymbol{B}_Q$. The first property: homogeneity. By definition, it is clear that

$$s\boldsymbol{B}\left(\frac{F}{s}, \frac{w}{s}, \frac{m}{s}, f, \lambda\right) = \boldsymbol{B}(F, w, m, f, \lambda), \quad \boldsymbol{B}(tF, w, m, tf, t\lambda) = \boldsymbol{B}(F, w, m, f, \lambda).$$

Choosing $s = m$ and $t = \lambda^{-1}$ and introducing new variables

$$\alpha = \frac{F}{m\lambda}, \quad \beta = \frac{w}{m}, \quad \gamma = \frac{f}{\lambda}$$

we can see that

$$\frac{1}{m}\boldsymbol{B}(F, w, m, f, \lambda) = B\left(\frac{F}{m\lambda}, \frac{w}{m}, \frac{f}{\lambda}\right) =: B(\alpha, \beta, \gamma), \tag{2-6}$$

where $B(\alpha, \beta, \gamma) = \boldsymbol{B}(\alpha, \beta, 1, \gamma, 1)$.

Obviously B is defined in the domain

$$G := \{(\alpha, \beta, \gamma) : |\gamma| \leq \alpha,\ 1 \leq \beta \leq Q\}. \tag{2-7}$$

The second property: special form of concavity. We formulate this property as the following theorem.

Theorem 2.7. *Let* P, P_+, $P_- \in \Omega$ *and, for* $0 \leq t \leq 1$,

$$P = (F, w, \min(m_+, m_-), f, \lambda),$$
$$P_+ = (F + A, w + u, m_+, f + a, \lambda + ta),$$
$$P_- = (F - A, w - u, m_-, f - a, \lambda - ta).$$

Then

$$\boldsymbol{B}(P) - \tfrac{1}{2}(\boldsymbol{B}(P_+) + \boldsymbol{B}(P_-)) \geq 0. \tag{2-8}$$

At the same time, if P, P_+, $P_- \in \Omega$, *and, for* $0 \leq t \leq 1$,

$$P = (F, w, \min(m_+, m_-), f, \lambda),$$
$$P_+ = (F + A, w + u, m_+, f + a, \lambda - ta),$$
$$P_- = (F - A, w - u, m_-, f - a, \lambda + ta),$$

then

$$\boldsymbol{B}(P) - \tfrac{1}{2}(\boldsymbol{B}(P_+) + \boldsymbol{B}(P_-)) \geq 0. \tag{2-9}$$

In particular, with fixed m, and with all points being inside Ω, we get for all $t \in [0, 1]$

$$\begin{aligned}\boldsymbol{B}(F, w, m, f, \lambda) \geq \tfrac{1}{4}\big(&\boldsymbol{B}(F - dF, w - dw, m, f - d\lambda, \lambda - t d\lambda)\\ &+\boldsymbol{B}(F - dF, w - dw, m, f - d\lambda, \lambda + t d\lambda)\\ &+\boldsymbol{B}(F + dF, w + dw, m, f + d\lambda, \lambda - t d\lambda)\\ &+\boldsymbol{B}(F + dF, w + dw, m, f + d\lambda, \lambda + t d\lambda)\big).\end{aligned} \tag{2-10}$$

Remark 2.8. (1) The differential notation, i.e., dF, dw, $d\lambda$, just means small numbers. (2) In (2-10) we lose a bit of information in comparison with (2-8), (2-9), but this is exactly (2-10), which we are going to use in the future.

Before proving this theorem, let us explain a bit more about what kind of concavity is represented by inequalities (2-8), (2-9), and thus by their consequence (2-10). We can use different notation for coordinates P_+, P_-, $P_\pm := (F_\pm, w_\pm, m_\pm, f_\pm, \lambda_\pm)$. We require all P, $P_\pm$ to belong to Ω and it is evident that

$$F = \frac{F_+ + F_-}{2}, \quad w = \frac{w_+ + w_-}{2}, \quad m = m_+ \wedge m_-, \quad f = \frac{f_+ + f_-}{2}, \quad \lambda = \frac{\lambda_+ + \lambda_-}{2},$$

but also "jumps" in the fourth and the fifth coordinates must be dependent on each other, namely,

$$t\Delta f := t(f_+ - f_-) = (\lambda_+ - \lambda_-) =: \Delta\lambda \quad \text{or} \quad t\Delta f = -\Delta\lambda, \quad 0 \leq t \leq 1.$$

So the function $\boldsymbol{B}$ (as we will now see) possesses such sophisticated concavity as encoded by jumps from any point $P \in \Omega$ to P_+, $P_- \in \Omega$, where P is almost the average of $P_\pm$, but not quite: the difference is that (1) the third coordinate is not an arithmetic average of the third coordinates of $P_\pm$, but their minimum, and (2) that the jumps in the fourth and the fifth coordinates are interdependent as above.

Proof. Fix P, P_+, $P_- \in \Omega$ as in (2-8). Let φ_+, φ_-, ω_+, ω_- be functions and weights giving the supremum in $\boldsymbol{B}(P_+)$, $\boldsymbol{B}(P_-)$ respectively up to a small number $\eta > 0$. Using the fact that $\boldsymbol{B}$ does not depend on I, we assume φ_+, ω_+ are on I_+ and φ_-, ω_- are on I_-. Consider

$$\varphi(x) := \begin{cases}\varphi_+(x), & x \in I_+,\\ \varphi_-(x), & x \in I_-,\end{cases} \qquad \omega(x) := \begin{cases}\omega_+(x), & x \in I_+\\ \omega_-(x), & x \in I_-.\end{cases}$$

Notice that then

$$(\varphi, h_I) \cdot \frac{1}{\sqrt{|I|}} = \Delta_I\varphi = \tfrac{1}{2}(P_{+,4} - P_{-,4}) =: a. \tag{2-11}$$

We denote the i-th coordinate of a point P by P_i. Then it is easy to see that $P_3 = \min(P_{3,-}, P_{3,+}) = \min(\min_{I_-}\omega_-, \min_{I_+}\omega_+)$, $P_5 = \lambda$,

$$\langle|\varphi|\omega\rangle_I = F = P_1, \quad \langle\omega\rangle_I = w = P_2, \quad \langle\varphi\rangle_I = f = P_4. \tag{2-12}$$

Notice that for $x \in I_\pm$ using (2-11), we get if $\varepsilon_I = -t,\ 0 \le t \le 1$,

$$\begin{aligned}\frac{1}{|I|}\omega_\pm\left\{x \in I_\pm : \sum_{J\subseteq I, J\in D} \varepsilon_J(\varphi, h_J)h_J(x) > \lambda\right\} &= \frac{1}{|I|}\omega_\pm\left\{x \in I_\pm : \sum_{J\subseteq I_\pm, J\in D} \varepsilon_J(\varphi, h_J)h_J(x) > \lambda \pm ta\right\}\\ &= \frac{1}{2|I_\pm|}\omega_\pm\left\{x \in I_\pm : \sum_{J\subseteq I_\pm, J\in D} \varepsilon_J(\varphi_\pm, h_J)h_J(x) > P_{\pm,5}\right\}\\ &\ge \tfrac{1}{2}B(P_\pm) - \eta.\end{aligned}$$

Combining the two options for the left-hand side we obtain for $\varepsilon_I = -1$

$$\frac{1}{|I|}\omega\left\{x \in I : \sum_{J\subseteq I, J\in D} \varepsilon_J(\varphi, h_J)h_J(x) > \lambda\right\} \ge \tfrac{1}{2}(B(P_+) + B(P_-)) - 2\eta.$$

Let us use now the simple information (2-12): if we take the supremum in the left-hand side above over all functions φ such that $\langle|\varphi|\,\omega\rangle_I = F,\ \langle\varphi\rangle_I = f,\ \langle\omega\rangle_I = w$, and weights ω such that $\langle\omega\rangle_I = w$ in dyadic A_1 with A_1-norm at most Q, and supremum over all $\varepsilon_J = \pm s,\ s \in [0, 1]$ (only ε_I stays fixed), we get a quantity smaller than or equal to the one where we have the supremum over all functions φ such that $\langle|\varphi|\,\omega\rangle = F,\ \langle\varphi\rangle_I = f,\ \langle\omega\rangle = w$, and weights ω such that $\langle\omega\rangle = w$ in dyadic A_1 with A_1-norm at most Q, and an unrestricted supremum over all $\varepsilon_J = \pm s,\ s \in [0, 1],\ \varepsilon_I = -t,\ 0 \le t \le 1$. The latter quantity is of course $\boldsymbol{B}(F, w, m, f, \lambda)$. So we proved (2-8).

To prove (2-9) we repeat verbatim the same reasoning, only keeping now $\varepsilon_I = t,\ 0 \le t \le 1$. □

Remark 2.9. This theorem is a sort of "fancy" concavity property; the attentive reader will see that (2-8), (2-9) include a biconcavity property entirely similar to the one demonstrated by the celebrated Burkholder function. We will use the consequence of biconcavity encompassed by (2-10). This is still another concavity. Let us also remark that it can be shown that $\boldsymbol{B}$ is a supersolution of a certain degenerate elliptic equation (but this fact does not help us in estimating $\boldsymbol{B}$ below).

The third property: $\boldsymbol{B}$ decreases in m. The function $\boldsymbol{B}$ is obviously decreasing in m. In fact, if m decreases (all other coordinates being fixed) then the collection of weights increases, and the supremum increases. It is not difficult to see that $\boldsymbol{B}$ is also continuous.

The fourth property: the function B from (2-6) is concave. Recall that by (2-6)

$$B\left(\frac{F}{\lambda}, w, \frac{f}{\lambda}\right) = \boldsymbol{B}(F, w, 1, f, \lambda). \tag{2-13}$$

Choosing $t = 0$ in Theorem 2.7 we see that $\boldsymbol{B}(F, w, 1, f, \lambda)$ is concave when λ is fixed. This proves the fourth property, which we formulated intentionally in terms of B and not $\boldsymbol{B}$.

The fifth property: the function $t \to (1/t)B(t\alpha, t\beta, \gamma)$ is increasing. This is the combination of (2-6) and the third property above.

The sixth property: the domain of definition of B is $G = \{(\alpha, \beta, \gamma) \in \mathbb{R}^3 : 1 \le \beta \le Q,\ |\gamma| \le \alpha\}$.

The seventh property: symmetry and monotonicity in γ. It is easy to see from the definition of $\boldsymbol{B}$ that it is even in its variable f. Therefore,

$$B(\alpha, \beta, \gamma) = B(\alpha, \beta, -\gamma).$$

Notice that the concavity of B (in γ) and this symmetry together imply that $\gamma \to B(\cdot, \cdot, \gamma)$ is decreasing on $\gamma \in [0, \alpha]$.

2B. ***The goal and the idea of the proof.*** In this section we are going to prove the following estimate from below on the function B.

Theorem 2.10. *There is an absolute positive constant c such that for some point $(\alpha, \beta, \gamma) \in G$*

$$B(\alpha, \beta, \gamma) \geq c\, Q (\log Q)^{1/3} \alpha. \tag{2-14}$$

Remark 2.11. It is a subtle result and it will take some space below to prove. Recall that Muckenhoupt conjectured that for the Hilbert transform H and any weight $w \in A_1$ the following two estimates hold on a unit interval I:

$$w\{x \in I : |Hf(x)| > \lambda\} \leq \frac{C}{\lambda} \int_I |f| M w \, dx, \tag{2-15}$$

$$w\{x \in I : |Hf(x)| > \lambda\} \leq \frac{C\,[w]_{A_1}}{\lambda} \int_I |f| w \, dx, \tag{2-16}$$

Obviously if (2-15) holds then (2-16) is valid as well. It took many years to *disprove* (2-15). This was done by Maria Reguera and Christoph Thiele [Reguera 2011; Reguera and Thiele 2012]. The constructions involve a very irregular (almost a sum of delta measures) weight w, so there was a hope that such an effect cannot appear when the weight is regular in the sense that $w \in A_1$. Theorem 2.10 gives a counterexample to this hope for the case when the Hilbert transform is replaced by the martingale transform on a usual homogeneous dyadic filtration. The reader can consult [Nazarov et al. 2015] to see that for the Hilbert transform a counterexample also exists, and so (2-16) fails as well. The counterexample for the Hilbert transform is the transference of a counterexample we build here for the martingale transform. Notice that Theorem 2.10 implicitly gives a certain counterexample for the Hilbert transform.

Now a couple of words about the idea of the proof of Theorem 2.10. Ideally we would like to find the formula for B, and therefore for $\boldsymbol{B}$ because of (2-6). To proceed we rewrite the second property of $\boldsymbol{B}$ as a PDE on B. Then we try to find the boundary conditions on B on ∂G, and then we may hope to solve this PDE. Unfortunately there are many roadblocks on this path, starting with the fact that the second property of $\boldsymbol{B}$ is not a PDE; it is rather a partial differential inequality in discrete form. We will write it down as a pointwise partial differential inequality, but for that we will need a subtle result of Aleksandrov. We also can find boundary values of B; see some of them in Section 2B1 below. However, the main difficulty is that our partial differential expression is in three dimensions.

2B1. *Unweighted case.* We first consider the simplest case of $m = \omega = 1$ identically. Then we are left with function $\mathcal{B}el(F, f, \lambda) = \boldsymbol{B}(F, 1, 1, f, \lambda)$, which is defined in a convex domain $\Omega_0 \subset \mathbb{R}^3$, $\Omega_0 := \{(F, f, \lambda) \in \mathbb{R}^3 : |f| \le F\}$, and whose concavity properties are described in:

Theorem 2.12. *Let* $P, P_+, P_- \in \Omega_0$,

$$P = (F, f, \lambda), \quad P_+ = (F + A, f + a, \lambda + a), \quad P_- = (F - A, f - a, \lambda - a).$$

Then

$$\mathcal{B}el(P) - \tfrac{1}{2}(\mathcal{B}el(P_+) + \mathcal{B}el(P_-)) \ge 0. \tag{2-17}$$

At the same time, if $P, P_+, P_- \in \Omega_0$,

$$P = (F, f, \lambda), \quad P_+ = (F + A, f + a, \lambda - a), \quad P_- = (F - A, f - a, \lambda + a),$$

then

$$\mathcal{B}el(P) - \tfrac{1}{2}(\mathcal{B}el(P_+) + \mathcal{B}el(P_-)) \ge 0. \tag{2-18}$$

Let us make the change of variables $(F, f, \lambda) \to (F, y_1, y_2)$:

$$y_1 := \tfrac{1}{2}(\lambda + f), \quad y_2 := \tfrac{1}{2}(\lambda - f).$$

Define

$$M(F, y_1, y_2) := B(F, y_1 - y_2, y_1 + y_2) = \mathcal{B}el(F, f, \lambda).$$

In terms of the function M, Theorem 2.12 reads as follows:

Theorem 2.13. *The function* M *is defined in the domain* $G := \{(F, y_1, y_2) : |y_1 - y_2| \le F\}$, *and for each fixed* y_2, $M(F, y_1, y_2)$ *is concave in* (F, y_1) *and for each fixed* y_1, $M(F, y_1, y_2)$ *is concave in* (F, y_2).

The properties of M are strongly reminiscent of the properties of the Burkholder function.

In the unweighted situation we can find $\boldsymbol{B}$ (or M) precisely. Here is the result proved in [Reznikov et al. 2013]:

Theorem 2.14.
$$\mathcal{B}el(F, f, \lambda) = \begin{cases} 1 & \text{if } \lambda \le F, \\ 1 - (\lambda - F)^2/(\lambda^2 - f^2) & \text{if } \lambda > F. \end{cases} \tag{2-19}$$

This result means that we found a boundary value of the Bellman function $\boldsymbol{B}(F, w, m, f, \lambda)$ of the weighted problem on the part of its boundary; namely we found this function of five variables on $\{P \in \partial\Omega : w = P_2 = P_3 = m\}$:

$$\boldsymbol{B}(F, m, m, f, \lambda) = m \begin{cases} 1 & \text{if } \lambda \le F, \\ 1 - (\lambda - F)^2/(\lambda^2 - f^2) & \text{if } \lambda > F. \end{cases} \tag{2-20}$$

In terms of the function B from (2-6), we have the following boundary values of B:

$$B(\alpha, 1, \gamma) = \begin{cases} 1 & \text{if } \alpha \ge 1, \\ 1 - (1 - \alpha)^2/(1 - \gamma^2) & \text{if } 0 \le |\gamma| \le \alpha < 1. \end{cases} \tag{2-21}$$

2C. ***From discrete inequality to differential inequality via Aleksandrov's theorem.*** By the fourth property of Section 2A1 the function B is concave on its domain of definition G. By the result of Aleksandrov, see Theorem 6.9 of [Evans and Gariepy 1992], B has all second derivatives almost everywhere; this means that for a.e. $x \in G^\circ$ and all small vectors $h \in \mathbb{R}^3$,

$$B(x+h) = B(x) + \nabla B(x)\cdot h + \langle H_B(x)\cdot h, h\rangle + o(|h|^2), \tag{2-22}$$

where H_B is the Hessian matrix of B. On the other hand the second property of Section 2A1 can be rewritten in terms of B as

$$\begin{aligned} B\left(\frac{F}{\lambda}, \beta, \frac{f}{\lambda}\right) - \frac{1}{4}\Bigg[& B\left(\frac{F-dF}{\lambda-d\lambda}, \beta-d\beta, \frac{f-d\lambda}{\lambda-d\lambda}\right) \\ & + B\left(\frac{F-dF}{\lambda-d\lambda}, \beta-d\beta, \frac{f+d\lambda}{\lambda-d\lambda}\right) \\ & + B\left(\frac{F+dF}{\lambda+d\lambda}, \beta+d\beta, \frac{f-d\lambda}{\lambda+d\lambda}\right) \\ & + B\left(\frac{F+dF}{\lambda+d\lambda}, \beta+d\beta, \frac{f+d\lambda}{\lambda+d\lambda}\right)\Bigg] \geq 0. \end{aligned} \tag{2-23}$$

Here $(F/\lambda, \beta, f/\lambda) \in G^\circ$ and $(dF, d\beta, d\lambda)$ is just any small vector in $\mathbb{R}^3$.

Theorem 2.15. *For almost every point $P=(\alpha,\beta,\gamma)=:(F/\lambda,\beta,f/\lambda)\in G^\circ$ and every vector $(dF,d\beta,d\lambda)\in\mathbb{R}^3$ we have*

$$\begin{aligned} & -\alpha^2 B_{\alpha\alpha}(P)\left(\frac{dF}{F}-\frac{d\lambda}{\lambda}\right)^2 - \beta^2 B_{\beta\beta}(P)\left(\frac{d\beta}{\beta}\right)^2 - (1+\gamma^2)B_{\gamma\gamma}(P)\left(\frac{d\lambda}{\lambda}\right)^2 \\ & \quad - 2\alpha\beta B_{\alpha\beta}(P)\left(\frac{dF}{F}-\frac{d\lambda}{\lambda}\right)\frac{d\beta}{\beta} + 2\beta\gamma B_{\beta\gamma}(P)\frac{d\beta}{\beta}\frac{d\lambda}{\lambda} + 2\alpha\gamma B_{\alpha\gamma}(P)\left(\frac{dF}{F}-\frac{d\lambda}{\lambda}\right)\frac{d\lambda}{\lambda} \\ & \quad + 2\alpha B_\alpha(P)\left(\frac{dF}{F}-\frac{d\lambda}{\lambda}\right)\frac{d\lambda}{\lambda} - 2\gamma B_\gamma(P)\left(\frac{d\lambda}{\lambda}\right)^2 \geq 0. \end{aligned} \tag{2-24}$$

Remark 2.16. We can mollify $\boldsymbol{B}$ to make it smooth and still have its "fancy concavity properties". But then we lose homogeneity and cannot reduce $\boldsymbol{B}$ to B. We can mollify $\boldsymbol{B}$ to keep its homogeneity — just choose the mollifier depending on the point — but then we loose its "fancy concavity property". In short, we have a problem with the mollification. This is why Aleksandrov's theorem is very useful now.

Proof. Fix a point $P \in G^\circ$, where Aleksandrov's identity (2-22) holds. Fix an arbitrary $(dx, dy, d\lambda) \in \mathbb{R}^3$. Let us use (2-23) by expanding the fractions

$$\frac{x \pm \varepsilon\, dx}{\lambda \pm \varepsilon d\lambda}, \quad \frac{f \pm \varepsilon d\lambda}{\lambda \pm \varepsilon d\lambda}$$

up to the second order in small parameter ε, and combining with the identity (2-22) after that. All terms with ε^0, ε^1 will disappear identically. Only the terms with ε^2 and smaller stay. After division by ε^2 we let ε tend to zero and get (2-24) for a.e. point $P \in G^\circ$. □

Of course we need something else from positive B to be able to prove that B satisfying this partial differential inequality (2-24) in the domain $G^\circ = \{P = (\alpha, \beta, \gamma) : 1 < \beta < Q,\ 0 < |\gamma| < \alpha\}$ has the estimate (2-14) from below. We actually have this "something else" in the form of an obstacle condition, which we will introduce in Section 2E.

But let us first simplify (2-24). Let us call by $\mathcal{N}$ the matrix of the quadratic form in (2-24). After a rather straightforward operation $\mathcal{N} \to \mathcal{M}_1 := A^* \mathcal{N} A$ with a certain invertible matrix A, we can write down the nonnegativity of the differential form in (2-24) as the a.e.-in-G° nonnegativity of the matrix

$$\mathcal{M}_1 := \begin{bmatrix} -\alpha^2 B_{\alpha\alpha} & -\alpha\beta B_{\alpha\beta} & \alpha\gamma B_{\alpha\gamma} + \alpha B_\alpha \\ -\alpha\beta B_{\alpha\beta} & -\beta^2 B_{\beta\beta} & \beta\gamma B_{\beta\gamma} \\ \alpha\gamma B_{\alpha\gamma} + \alpha B_\alpha & \beta\gamma B_{\beta\gamma} & -(1+\gamma^2) B_{\gamma\gamma} - 2\gamma B_\gamma \end{bmatrix} \geq 0. \tag{2-25}$$

However, we saw already that $B(\alpha, \beta, \gamma)$ is concave, which implies the nonnegativity of yet another matrix:

$$\mathcal{M}_2 := \begin{bmatrix} -\alpha^2 B_{\alpha\alpha} & -\alpha\beta B_{\alpha\beta} & -\alpha\gamma B_{\alpha\gamma} \\ -\alpha\beta B_{\alpha\beta} & -\beta^2 B_{\beta\beta} & -\beta\gamma B_{\beta\gamma} \\ -\alpha\gamma B_{\alpha\gamma} & -\beta\gamma B_{\beta\gamma} & -\gamma^2 B_{\gamma\gamma} \end{bmatrix} \geq 0. \tag{2-26}$$

Taking the half-sum of (2-25) and (2-26), we obtain the nonnegativity

$$\mathcal{M} := \begin{bmatrix} -\alpha^2 B_{\alpha\alpha} & -\alpha\beta B_{\alpha\beta} & \frac{1}{2}\alpha B_\alpha \\ -\alpha\beta B_{\alpha\beta} & -\beta^2 B_{\beta\beta} & 0 \\ \frac{1}{2}\alpha B_\alpha & 0 & -\left(\frac{1}{2}+\gamma^2\right) B_{\gamma\gamma} - \gamma B_\gamma \end{bmatrix} \geq 0. \tag{2-27}$$

It is now natural to restrict the quadratic form of this matrix on certain two-dimensional hyperplanes in the three-dimensional tangent space Tan_p of the graph $\Gamma := \{p := (P, B(P)),\ P \in G^\circ\}$ at a given point p. Namely, let us consider the quadratic form of the matrix $\mathcal{M}$ in (2-25) on vectors of the form

$$(\xi, \xi, \eta). \tag{2-28}$$

Then, using the notation

$$\psi(\alpha, \beta, \gamma) := \psi_B(\alpha, \beta, \gamma) := -\alpha^2 B_{\alpha\alpha} - 2\alpha\beta B_{\alpha\beta} - \beta^2 B_{\beta\beta}, \tag{2-29}$$

we get the a.e.-in-G° nonnegativity of the matrix

$$\begin{bmatrix} \psi(\alpha, \beta, \gamma) & \frac{1}{2}\alpha B_\alpha \\ \frac{1}{2}\alpha B_\alpha & -\left(\frac{1}{2}+\gamma^2\right) B_{\gamma\gamma} - \gamma B_\gamma \end{bmatrix} \geq 0. \tag{2-30}$$

Definition 2.17. Consider a subdomain of G,

$$G_1 := \left\{(\alpha, \beta, \gamma) \in G : |\gamma| < \tfrac{1}{2}\alpha, 2 < \beta < Q\right\}.$$

Fix now $(\alpha, \beta, \gamma) \in G_1$ and a parameter $t \in \left[\frac{1}{2}, 1\right]$. Replace in the previous inequality (α, β, γ) by $(t\alpha, t\beta, \gamma)$. Denote temporarily

$$P_t := (t\alpha, t\beta, \gamma), \quad (\alpha, \beta, \gamma) \in G_1, \quad \tfrac{1}{2} \leq t \leq 1.$$

Then we get for every such t and every point P_t the following inequality for all $(\xi, \eta) \in \mathbb{R}^2$:

$$\xi^2[\psi(P_t)] + \xi\eta(\alpha t B_\alpha(P_t)) + \eta^2(-\gamma B_\gamma(P_t) - \left(\tfrac{1}{2} + \gamma^2\right) B_{\gamma\gamma}(P_t)) \geq 0. \tag{2-31}$$

Consider a new function H, which is a certain averaging of B; namely, for any $P = (\alpha, \beta, \gamma) \in G_1$, let

$$H(P) = 2\int_{1/2}^{1} B(P_t)\, dt.$$

Notice several simple facts. First of all

$$\alpha H_\alpha = 2\int_{1/2}^{1} \alpha t B(t\alpha, t\beta, \gamma)\, dt, \quad \alpha^2 H_{\alpha\alpha} = 2\int_{1/2}^{1} (\alpha t)^2 B_{\alpha\alpha}(t\alpha, t\beta, \gamma)\, dt.$$

Similarly, if for every function F we introduce the notation

$$\psi_F(\alpha, \beta, \gamma) := -\alpha^2 F_{\alpha\alpha} - 2\alpha\beta F_{\alpha\beta} - \beta^2 F_{\beta\beta}, \tag{2-32}$$

we get

$$\psi_H = 2\int_{1/2}^{1} \psi_B(t\alpha, t\beta, \gamma)\, dt.$$

Now integrate (2-31) on the interval $t \in \left[\frac{1}{2}, 1\right]$. The previous simple observations allow us now to rewrite this as a pointwise inequality for function H on domain G_1 introduced in Definition 2.17:

$$\xi^2[\psi_H(P)] + \xi\eta(\alpha H_\alpha(P)) + \eta^2(-\gamma H_\gamma(P) - (1/2 + \gamma^2) H_{\gamma\gamma}(P)) \geq 0. \tag{2-33}$$

The reader may wonder why we are so keen to replace (2-31) by the virtually identical (2-33)? The answer is because we can give a very good pointwise estimate on $\psi_H(P)$, $P \in G_1$. Unfortunately we cannot give any pointwise estimate on $\psi(P)$, $P \in G$.

Now we deduce the desired pointwise estimate on ψ_H; we will use below its consequences. First, let us define

$$R := \sup \frac{B(P)}{\alpha}, \quad P = (\alpha, \beta, \gamma) \in G. \tag{2-34}$$

Our goal formulated in (2-14) is to prove $R \geq cQ(\log Q)^\varepsilon$. We are still not too close, but notice that automatically $B(P) \leq R\alpha$, $P = (\alpha, \beta, \gamma) \in G$.

Lemma 2.18. *If $P = (\alpha, \beta, \gamma)$ is such that $|\gamma| \leq \frac{1}{8}\alpha$ and $\beta > 100$ then*

$$\psi_H(P) = 2\int_{1/2}^{1} \psi(t\alpha, t\beta, \gamma)\, dt \leq CR\left(|\gamma| + \frac{\alpha}{\beta}\right),$$

where C is an absolute constant.

Proof. Consider the function

$$\varphi(t) := B(t\alpha, t\beta, \gamma) \tag{2-35}$$

for a.e. $(\alpha, \beta, \gamma) \in G_1$. It is concave.

Let us first prove that

$$\int_{1/2}^{1} -\varphi''(t)\,dt \le CR\left(|\gamma| + \frac{\alpha}{\beta}\right). \tag{2-36}$$

This would imply

$$\int_{1/2}^{1} \psi(t\alpha, t\beta, \gamma)\,dt \le CR\left(|\gamma| + \frac{\alpha}{\beta}\right),$$

because by the definitions (2-29), (2-35) of ψ and φ we have

$$\psi(t\alpha, t\beta, \gamma) = -t^2\varphi''(t).$$

To prove (2-36) let us consider an auxiliary function $r(t) := \varphi(1)t - \varphi(t)$. It is defined for $t \in [\max(|\gamma|/\alpha, 1/\beta), 1]$. At 1 it vanishes, it is convex, and it attains its maximum on its left end-point $t_0 = \max(|\gamma|/\alpha, 1/\beta)$. The last statement follows from the fact that $\varphi(t)/t$ is increasing; this is the fifth property of Section 2A1 of B.

So on $[t_0, 1]$,

$$r(t) \le r(t_0) \le \varphi(1)t_0 \le R\alpha t_0 \le R\alpha\left(\frac{|\gamma|}{\alpha} + \frac{1}{\beta}\right). \tag{2-37}$$

As $\varphi(t)/t$ is increasing, we have $t\varphi'(t) - \varphi(t) \ge 0$, and thus $r'(1) \le 0$. Let us write down the Taylor formula for the convex function $r(t)$ in integral form, keeping in mind that $r(1) = 0$, $r'(1) \le 0$:

$$r(t_0) = (t_0 - 1)r'(1) + \int_{t_0}^{1} dt \int_{t}^{1} r''(s)\,ds.$$

Fubini's theorem, (2-37), and $r'(1) \le 0$ imply

$$\int_{t_0}^{1} (s - t_0)r''(s)\,ds \le R\alpha\left(\frac{|\gamma|}{\alpha} + \frac{1}{\beta}\right).$$

But $t_0 \le \frac{1}{8}$ by the assumptions of the lemma. So $\int_{1/2}^{1} r''(s)\,ds \le \frac{8}{3}R\alpha(|\gamma|/\alpha + 1/\beta)$. Hence, as $r'' = -\varphi''$,

$$\int_{1/2}^{1} -\varphi''(s)\,ds \le \tfrac{8}{3}R\alpha\left(\frac{|\gamma|}{\alpha} + \frac{1}{\beta}\right).$$

The proof of (2-36) is finished and this, as we saw at the beginning of the proof, gives Lemma 2.18. □

2D. ***Logarithmic blow-up.*** Recall that

$$G_3 = \left\{P \in G : |\gamma| \le \tfrac{1}{1000}\alpha,\ \beta > 100\right\}.$$

By Lemma 2.18 we conclude that for any $P = (\alpha, \beta, \gamma) \in G_3$

$$[\psi_H] \cdot \left[-\gamma H_\gamma - \left(\tfrac{1}{2} + \gamma^2\right)H_{\gamma\gamma}\right] \ge \tfrac{1}{4}\alpha^2 H_\alpha^2. \tag{2-38}$$

We will consider only points P such that

$$0 < \gamma \ll \alpha \ll \beta, \quad \alpha \le 1.$$

The absolute constants C, c will vary from line to line.

Let us temporarily take for granted the following inequality, where c_1, c_2 are absolute positive constants:

$$\alpha \le c_2 \frac{\beta}{R} \quad \Longrightarrow \quad H_\alpha(\alpha, \beta, \gamma) \ge c_1 \beta, \quad \beta \in \big(1, \tfrac{1}{2} Q\big]. \tag{2-39}$$

Using Lemma 2.18 we obtain

$$\psi_H \le C R \left(\gamma + \frac{\alpha}{\beta}\right).$$

Now we combine this inequality with inequalities and (2-39), (2-38) to obtain

$$-\gamma H_\gamma - \big(\tfrac{1}{2} + \gamma^2\big) H_{\gamma\gamma} \ge c_3 \frac{\alpha^2 \beta^2}{R(\alpha/\beta + \gamma)}. \tag{2-40}$$

Using the fact that we consider only $0 < \gamma \le \alpha \le 1$, we can rewrite (2-40) as

$$-\frac{2\gamma}{(1+2\gamma^2)} H_\gamma - H_{\gamma\gamma} \ge c_4 \frac{\alpha^2 \beta^2}{R(\alpha/\beta + \gamma)}.$$

Using the integrating factor, we get

$$-[\mu(\gamma) H_\gamma]_\gamma \ge c_5 \frac{\alpha \beta^3}{R(1 + (\beta/\alpha)\gamma)}.$$

We integrate this inequality from 0 to γ to produce (we use that $\mu(\gamma) \approx 1$ when γ is small)

$$-H_\gamma \ge c_6 \frac{\alpha^2 \beta^2}{R} \log\left(1 + \frac{\beta}{\alpha}\gamma\right). \tag{2-41}$$

From now on let us fix α as follows:

$$\alpha = c_2 \frac{\beta}{R}, \tag{2-42}$$

where c_2 is from (2-39).

We integrate (2-41) from 0 to γ and use the positivity of H to produce

$$H(\alpha, \beta, 0) - H(\alpha, \beta, \gamma) \ge c_6 \frac{\alpha^3 \beta}{R} \left[\left(1 + \frac{\beta}{\alpha}\gamma\right) \log\left(1 + \frac{\beta}{\alpha}\gamma\right) - \frac{\beta}{\alpha}\gamma\right] \ge c_7 \frac{\alpha^2 \beta^2}{R} \gamma \log\left(\frac{\beta}{\alpha}\gamma\right); \tag{2-43}$$

the last inequality holds true because $\beta/\alpha = cR$, and because from now on we will fix γ and β:

$$\beta = \frac{Q}{4}, \quad \gamma = c_8 \frac{\beta}{R}, \tag{2-44}$$

where an absolute positive constant c_8 is much smaller than c_2 from (2-42). In particular, $(\beta/\alpha)\gamma \asymp \beta = \frac{1}{4} Q$ and so it is much bigger than 1. This justifies the last inequality in (2-43). This also gives

$$\gamma \ll \alpha.$$

We just obtained the inequality

$$\frac{\alpha^2 \beta^2}{R} \gamma \log\left(\frac{\beta}{\alpha}\gamma\right) \le C(H(\alpha, \beta, 0) - H(\alpha, \beta, \gamma)). \tag{2-45}$$

Let us use the fact that $B(\alpha,\beta,\gamma)$ is concave in γ (it is concave in all three variables) and that by its definition it is even in γ. See the seventh property of Section 2A1. The same then holds for the function H, which is just some averaging of B in the first two variables. Being even in γ on $\gamma\in[-\alpha,\alpha]$ and concave, it automatically decreases for $\gamma\in[0,\alpha]$; concavity and nonnegativity of H give $H(\alpha,\beta,\gamma)\ge(1-\gamma/\alpha)H(\alpha,\beta,0)$. This allows us to estimate the right-hand side of (2-45), and we have

$$\frac{\alpha^2\beta^2}{R}\gamma\log\left(\frac{\beta}{\alpha}\gamma\right)\le C(H(\alpha,\beta,0)-H(\alpha,\beta,\gamma))\le C\frac{\gamma}{\alpha}H(\alpha,\beta,0).$$

Taking into consideration one more time that $H(\alpha,\beta,\gamma)\le R\alpha$ by the definition of R in (2-34) and by the construction of H, we get

$$\frac{\alpha^2\beta^2}{R}\gamma\log\left(\frac{\beta}{\alpha}\gamma\right)\le C(H(\alpha,\beta,0)-H(\alpha,\beta,\gamma))\le CR\gamma, \tag{2-46}$$

or

$$\frac{Q^4}{R^4}\log\left(\frac{\beta}{\alpha}\gamma\right)\le C. \tag{2-47}$$

As $\beta/\alpha=cR$ and $\gamma\asymp Q/R$, we can see that $\log((\beta/\alpha)\gamma)\ge\log(cQ)$, from which it follows that

$$R\ge cQ(\log Q)^{1/4} \tag{2-48}$$

with a positive absolute c. Theorem 2.10 gets proved with $\delta=\frac14$.

We are left to prove (2-39).

Lemma 2.19. *Suppose $H(1,\beta,\gamma)\ge A$. Then the following holds*:

$$H_\alpha\left(\frac{A}{2R},\beta,\gamma\right)\ge\frac{A}{2}.$$

Proof. Suppose not, then $H_\alpha(A/(2R),\beta,\gamma)\le\frac12 A$. Then $H_\alpha(\alpha,\beta,\gamma)\le\frac12 A$ for all $\alpha\in[A/(2R),1]$ by the fact that H_α decreases in α as H is concave.

But

$$H(1,\beta,\gamma)-H\left(\frac{A}{2R},\beta,\gamma\right)\ge A-R\frac{A}{2R}=\frac{A}{2}$$

by the definition of R in (2-34) and the fact that H is a certain averaging of B.

On the other hand,

$$H(1,\beta,\gamma)-H\left(\frac{A}{2R},\beta,\gamma\right)=H_\alpha(\theta,\beta,\gamma)\left(1-\frac{A}{2R}\right),$$

$\theta\in[A/(2R),1]$. We obtain (combining the last inequalities)

$$\frac{A}{2}\le H(1,\beta,\gamma)-H\left(\frac{A}{2R},\beta,\gamma\right)<H_\alpha(\theta,\beta,\gamma)\le\frac{A}{2}.$$

We come to a contradiction, so the lemma is proved. □

The combination of Lemma 2.19 and (2-53) proves inequality (2-39).

2E. ***An obstacle condition on functions B and H.*** Now we want to show the following obstacle condition for B, which we already used:

$$\text{If } |\gamma| < \tfrac{1}{4}, \text{ then } B(1, \beta, \gamma) \geq \tfrac{1}{3}\beta. \tag{2-49}$$

Let $I := [0, 1]$. Given numbers $(F, \beta, m, f, \lambda)$ such that $|f| < \frac{1}{4}\lambda,\ F/m = \lambda,\ m \leq \beta \leq Qm$, it is enough to construct functions φ, ψ, w on I such that:

(1) Each of these functions has constant values on grandchildren of I.

(2) If $\varphi = \langle\varphi\rangle_I + (\varphi, h_{I_-})h_{I_-} + (\varphi, h_{I_+})h_{I_+}$, then

$$\psi = -\lambda + (\varphi, h_{I_-})h_{I_-} - (\varphi, h_{I_+})h_{I_+}.$$

(3) $\langle w\rangle_I = \beta,\ \min_I w = m$.

(4) The w-measure of the subset of I, where

$$\psi \geq 0 \tag{2-50}$$

is at least $c\beta$, where c is an absolute positive constant. Notice that (2-50) is the same as $(\varphi, h_{I_-})h_{I_-} - (\varphi, h_{I_+})h_{I_+} \geq \lambda$.

Here is the construction of such a triple (φ, ψ, w). Fix $\beta \in (1, Q]$. Put $\varphi = -a$ on I_{--}, $\varphi = b$ on I_{++}, and $\varphi = 0$ otherwise. And $w = 1$ on $I_{--} \cup I_{++}$, and $w = \beta$ otherwise. Then put

$$\psi := -\lambda + (\varphi, h_{I_-})h_{I_-} - (\varphi, h_{I_+})h_{I_+}.$$

Let $0 < a < b$ and a be close to b. Put $\lambda = \frac{1}{4}(a + b)$. Then the average of φ is $\frac{1}{4}(b - a)$. It is small with respect to λ and we can prescribe it to be any number smaller than $\frac{1}{4}\lambda$. $F = \frac{1}{4}(a + b),\ m = 1$.

On the other hand, the function $\lambda + \psi$ (which is a martingale transform of $\varphi - \langle\varphi\rangle_I$) is at least $-(\varphi, h_{I_+})h_{I_+} \geq \frac{1}{2}b \geq \lambda$ on I_{+-}, whose w-measure is more than $\frac{1}{3}w(I)$. So

$$B\big(1, \tfrac{1}{2}(1 + \beta), \gamma\big) \geq \tfrac{1}{3}\beta \tag{2-51}$$

for all sufficiently small γ.

By concavity and positivity of B we see immediately

$$B(\alpha, \beta, \gamma) \geq c\beta, \quad \alpha \geq \tfrac{1}{100}, \tag{2-52}$$

with absolute positive c and for $\beta \in \big(1, \frac{1}{2}Q\big]$.

Now, from the definition of functions H we conclude that the following obstacle condition holds for the function H:

$$H(1, \beta, \gamma) \geq \tfrac{1}{3}\beta \tag{2-53}$$

for all sufficiently small γ and for $\beta \in \big(1, \frac{1}{2}Q\big]$.

2F. ***Improving the exponent*** $\frac{1}{3}$**.** From (2-53) we know that (this is for all $\gamma,\ 0\le\gamma\le 1$)

$$H\big(1,\tfrac14 Q,\gamma\big)\ge \tfrac{1}{12}Q.$$

As $H(\alpha,\beta,\gamma)\le R\alpha$ we immediately conclude that

$$H\left(\frac{Q}{24R},\frac{Q}{4},\gamma\right)\le\frac{Q}{24}$$

(this is for all $\gamma,\ 0\le\gamma\le\alpha:=Q/(24R)$). Combined with the previous displayed inequality above this gives us (2-39),

$$H_\alpha\left(\frac{Q}{24R},\frac{Q}{4},\gamma\right)\ge\frac{Q}{24}. \tag{2-54}$$

But there may be a better point $\tilde\alpha\gg\alpha:=Q/(24R)$, where $H\big(\tilde\alpha,\frac14 Q,\gamma\big)\le\frac{1}{24}Q$. Then automatically we have the same estimate for H_α at this point:

$$H_\alpha\big(\tilde\alpha,\tfrac14 Q,\gamma\big)\ge\tfrac{1}{24}Q. \tag{2-55}$$

So let us consider the largest $\tilde\alpha\in[\alpha,1]$ where $\alpha=Q/(24R)$ such that the following holds:

$$\begin{gathered} H\big(\tilde\alpha,\tfrac14 Q,0\big)=\tfrac{1}{24}Q.\\ \text{Then}\quad H\big(\tilde\alpha,\tfrac14 Q,\gamma\big)\le\tfrac{1}{24}Q,\quad \gamma\in[0,\tilde\alpha]. \end{gathered} \tag{2-56}$$

Two cases may occur:

Case 1: $\tilde\alpha\ge Q^{1/2}/(24R^{1/2})$. Then in (2-40) we can use $\tilde\alpha\ge Q^{1/2}/(24R^{1/2})$ and $\beta=\frac14 Q$. We just follow (2-45) and (2-46) with these new data, but with one small change; γ in (2-46) can be between 0 and $\tilde\alpha$, so in particular, it can be chosen to be $\gamma=Q^{1/2}/(24R^{1/2})$. Then instead of (2-47) we get

$$c\frac{Q^3}{R^3}\log\left(\frac{cQ}{\tilde\alpha}\gamma\right)=c\frac{Q^3}{R^3}\log\left(\frac{cQR^{1/2}}{Q^{1/2}}\cdot\frac{cQ^{1/2}}{R^{1/2}}\right)\le C. \tag{2-57}$$

This implies

$$R\ge cQ\log^{1/3}Q. \tag{2-58}$$

Case 2: $\tilde\alpha\le Q^{1/2}/(24R^{1/2})$. At $\alpha_1:=\min\big(Q/(48R),\frac23\tilde\alpha\big)$ we have

$$H\big(\alpha_1,\tfrac14 Q,\gamma\big)\le\tfrac{1}{48}Q.$$

But we saw that $\tilde\alpha\ge Q/(24R)$ by its definition. Hence, $\alpha_1=Q/(48\alpha)$. Comparing the last displayed inequality with (2-56) we conclude that

$$\begin{aligned}\tilde\alpha H_\alpha\left(\alpha_1,\frac{Q}{4},\gamma\right)&\ge(\tilde\alpha-\alpha_1)H_\alpha\left(\alpha_1,\frac{Q}{4},\gamma\right)\\ &\ge H\left(\tilde\alpha,\frac{Q}{4},\gamma\right)-H\left(\alpha_1,\frac{Q}{4},\gamma\right)\ge\left(1-\frac{\gamma}{\tilde\alpha}\right)H\left(\tilde\alpha,\frac{Q}{4},0\right)-\frac{Q}{48}\\ &\ge\left(1-\frac{\gamma}{\tilde\alpha}\right)H\left(\tilde\alpha,\frac{Q}{4},0\right)-\frac{Q}{48}\ge\left(1-\frac{\gamma}{\tilde\alpha}\right)\frac{Q}{24}-\frac{Q}{48}=\frac{Q}{144}\end{aligned}$$

if $\gamma \in \left[0, \frac{2}{3}\alpha_1\right]$. Hence, using that $\tilde{\alpha} \le Q^{1/2}/(24R^{1/2})$, we obtain the improved estimate on the derivative

$$\text{for all } \gamma \in \left[0, \tfrac{2}{3}\alpha_1\right], \quad H_\alpha\left(\alpha_1, \tfrac{1}{4}Q, \gamma\right) \ge cQ^{1/2}R^{1/2}. \tag{2-59}$$

Then in (2-40) we can use $\alpha := \alpha_1 = \frac{1}{48}Q$, $\beta = cQ^{1/2}R^{1/2}$, and $\gamma = \frac{2}{3}\alpha_1$.

And now we have a new estimate from below, namely (2-59). We just follow (2-45) and (2-46) with these new data, but with one small change; γ in (2-46) can be between 0 and $\alpha_1 = Q/(48R)$. Then instead of (2-47) we get

$$c\frac{Q^2}{R^2}\frac{QR}{R}\log\left(\frac{cQ}{\alpha_1}\gamma\right) \le CR;$$

so again, having $\gamma = \frac{2}{3}\alpha_1$, we obtain

$$R \ge cQ\log^{1/3} Q.$$

2G. ***Our Bellman function B as a viscosity supersolution of a degenerate elliptic equation.*** Let us remind the reader that we defined in (2-4) the function $\boldsymbol{B}$ on the domain Ω introduced in (2-5). We want to demonstrate in this short subsection that $\boldsymbol{B}$ is a supersolution in the viscosity sense of a certain degenerate elliptic equation.

We haven't used this before, but this knowledge might happen to be important. In particular, it may happen to be true that the reader more familiar with viscosity (super)solutions can simplify a bit our proof of Theorem 2.10, which we just finished proving. In this section D^2u denotes the Hessian matrix of u.

Definition 2.20. An equation $H(x, u, Du, D^2u) = 0$, $x \in \Omega \subset \mathbb{R}^d$, on a function u defined in a domain Ω is called degenerate elliptic if the function H satisfies the following condition: for any point $(x, u, p) \in \Omega \times \mathbb{R} \times \mathbb{R}^d$ and any two $d \times d$ real symmetric matrices X and Y, we have that from $Y \ge X$ it follows that $H(x, u, p, X) \ge H(x, u, p, Y)$.

For example, $H(x, u, p, X) = -\operatorname{trace} X$ gives a degenerate elliptic equation $-\Delta u = 0$. Many examples of degenerate elliptic operators can be found in the first sections of [Nadirashvili et al. 2014]; our example below can be found there too.

Definition 2.21. A lower semicontinuous function u is called a viscosity supersolution of (a degenerate elliptic equation) $H(x, u, Du, D^2u) = 0$ if for every point $x_0 \in \Omega$ and for every C^2 function φ such that (1) $\varphi(x_0) = u(x_0)$ and (2) $\varphi(x) \le u(x)$ for x in a small neighborhood of x_0 inside Ω, one has the inequality $H(x_0, \varphi(x_0), D\varphi(x_0), D^2\varphi(x_0)) \ge 0$.

To define viscosity subsolution one changes lower to upper semicontinuous, requires $\varphi(x) \ge u(x)$ for x in a small neighborhood of x_0 inside Ω, and gets the conclusion that $H(x_0, \varphi(x_0), D\varphi(x_0), D^2\varphi(x_0)) \le 0$.

To define the degenerate elliptic equation whose viscosity supersolution is $\boldsymbol{B}$ in Ω from (2-5), we consult Theorem 2.7 and especially inequality (2-10).

Our function $H(x, u, p, X)$ will depend only on matrices X that run over 5×5 real symmetric matrices. A vector v in $\mathbb{R}^5$ is called adapted if $v = (v_1, v_2, 0, 0, v_5)$, $\|v\| = 1$. The set of adapted vectors is called $\mathcal{A}$.

Let us consider the following H_{wmt}, where the subscript stands for "weak martingale transform":

$$H_{\text{wmt}}(X) := -\sup_{v\in\mathcal{A}}[(Xv, v) + X_{44}v_5^2].$$

It is very easy to check that if $Y \geq X$ are two real symmetric matrices, then $H(X) \geq H(Y)$.

Let us see that $\boldsymbol{B}$ from (2-4) satisfies all conditions of a viscosity supersolution of $H_{\text{wmt}}(D^2u) = 0$ in Ω from (2-5). The lower semicontinuity of $\boldsymbol{B}$ follows easily from its definition. Now let us fix $x_0 = (F, w, m, f, \lambda)$. If a smooth φ satisfies $\varphi(x) \leq \boldsymbol{B}(x)$ in a neighborhood of this x_0, then

$$\varphi(F \pm dF, w \pm dw, m, f \pm df, \lambda + \pm d\lambda) \leq \boldsymbol{B}(F \pm dF, w \pm dw, m, f \pm df, \lambda + \pm d\lambda)$$

for all sufficiently small real numbers $dF, dw, df, d\lambda$. Of course we also have $\varphi(F, w, m, f, \lambda) = \boldsymbol{B}(F, w, m, f, \lambda)$. Automatically, (2-10) gives us now that for all sufficiently small real numbers $dF, dw, df, d\lambda$ the following holds:

$$\begin{aligned}\varphi(F, w, m, f, \lambda) \geq \tfrac{1}{4}\big(&\varphi(F - dF, w - dw, m, f - d\lambda, \lambda - d\lambda)\\ &+\varphi(F - dF, w - dw, m, f + d\lambda, \lambda - d\lambda)\\ &+\varphi(F + dF, w + dw, m, f - d\lambda, \lambda + d\lambda)\\ &+\varphi(F + dF, w + dw, m, f + d\lambda, \lambda + d\lambda)\big).\end{aligned} \tag{2-60}$$

The function φ is smooth. Let us use Taylor's formula for all terms in the right-hand side of (2-60). We can easily see that $\varphi(F, w, m, f, \lambda)$ will disappear together with all terms having the first derivatives of φ. After simple algebra, which we leave to the reader, we can see that (2-60) implies an "infinitesimal" version of itself, which holds for any triple $(dF, dw, d\lambda)$:

$$-\big(\varphi_{FF}(dF)^2 + 2\varphi_{Fw}dFdw + \varphi_{ww}(dw)^2 + \varphi_{\lambda\lambda}(d\lambda)^2 + 2\varphi_{F\lambda}dFd\lambda + 2\varphi_{w\lambda}dwd\lambda + \varphi_{ff}(d\lambda)^2\big) \geq 0. \tag{2-61}$$

The reader can immediately see by the definition of H_{wmt} that we just proved

$$H_{\text{wmt}}(D^2\varphi(x_0)) \geq 0.$$

This means exactly that $\boldsymbol{B}$ is a viscosity supersolution of a degenerate elliptic equation $H_{\text{wmt}}(D^2u) = 0$.

3. Random-walk interpretation

In this section we want to prepare the ground for proving our main result, Theorem 2.2. We consider again the domain (2-5), namely,

$$\Omega^s := \{(F, w, m, f, \lambda) \in \mathbb{R}^5 : F \geq |f|\, m,\ m \leq w \leq Q\, m\}. \tag{3-1}$$

We consider special random walks in this domain. From the point (F, w, m, f, λ) in Ω^s we move with equal probability $\frac{1}{4}$ to the following four points (they have to be in this same domain Ω^s):

$$\begin{aligned}&(F - dF, w - dw, m_1, f - d\lambda, \lambda - td\lambda), \quad (F - dF, w - dw, m_2, f - d\lambda, \lambda + td\lambda),\\ &(F + dF, w + dw, m_3, f + d\lambda, \lambda - td\lambda), \quad (F + dF, w + dw, m_4, f + d\lambda, \lambda + td\lambda),\end{aligned}$$

where $\min(m_1, m_2, m_3, m_4) = m$. The symbols $dF, dw, d\lambda$ are just real numbers and $t \in [0, 1]$. The only condition on them is that we stay in Ω^s after performing this one-step random walk.

Now we make another random step from the four points listed above. We consider such random walks which also satisfy two conditions: (a) There is only finite number of steps. (b) The last step brings us to the following part of the boundary of Ω^s:

$$\partial_w \Omega^s = \{(F, w, m, f, \lambda) : w = m\}. \tag{3-2}$$

Let us call such random walks $\mathcal{W}^s$.

Every random walk is the collection of four martingales $\boldsymbol{F}, \boldsymbol{w}, \boldsymbol{f}, \Lambda$ and the "martingale-with-respect-to-minimum" M. Martingales $\boldsymbol{f}$ and Λ are strongly dependent. The random vector $(\boldsymbol{F}, \boldsymbol{w}, M, \boldsymbol{f}, \Lambda)$ should stay in Ω^s and should finish at $\partial_w \Omega^s$. We have a natural probability measure on $\mathcal{W}^s$; the expectation will be called $\mathbb{E}$.

If we start with the point (F, w, m, f, λ) in Ω^s, let us denote by

$$(\boldsymbol{F}(\omega), \boldsymbol{w}(\omega), M(\omega), \boldsymbol{f}(\omega), \Lambda(\omega))$$

a vector function we get after the walk ends at $\partial_w \Omega^s$. In particular, $\boldsymbol{w}(\omega) = M(\omega)$ identically.

We introduce

$$\mathbb{V}(F, w, m, f, \lambda) := \mathbb{V}_Q(F, w, m, f, \lambda) := \sup \mathbb{E}\, \mathbf{1}_{\Lambda(\omega) \le 0},$$

where the supremum is taken over all walks in $\mathcal{W}^s$ started at (F, w, m, f, λ).

Theorem 3.1. *The function $V = V_Q$ satisfies all the same properties as $\boldsymbol{B}_Q$ from Section 2A1 with one change; instead of properties (2-8), (2-9) it satisfies the analog of (2-10), namely,*

$$\begin{aligned} \mathbb{V}(F, w, m, f, \lambda) \ge \tfrac{1}{4}\big(&\mathbb{V}(F - dF, w - dw, m, f - d\lambda, \lambda - t d\lambda) \\ &+ \mathbb{V}(F - dF, w - dw, m, f - d\lambda, \lambda + t d\lambda) \\ &+ \mathbb{V}(F + dF, w + dw, m, f + d\lambda, \lambda - t d\lambda) \\ &+ \mathbb{V}(F + dF, w + dw, m, f + d\lambda, \lambda + t d\lambda)\big). \end{aligned} \tag{3-3}$$

The proof is the same as the proof of Theorem 2.7; it is based on the same trick of concatenation.

We can introduce the function V starting with the function $\mathbb{V}$ in the same manner as in (2-6), namely,

$$\frac{1}{m}\mathbb{V}(F, w, m, f, \lambda) = V\left(\frac{F}{m\lambda}, \frac{w}{m}, \frac{f}{\lambda}\right) =: V(\alpha, \beta, \gamma), \tag{3-4}$$

defined in the same domain $G = \{(\alpha, \beta, \gamma) : |\gamma| \le \alpha,\ 1 \le \beta \le Q\}$.

Theorem 3.2. *If $|\gamma| < \frac{1}{4}$, then $V(1, \beta, \gamma) \ge \frac{1}{3}\beta$.* (3-5)

To show this we just notice that in Section 2E we constructed a one-step random walk from $\mathcal{W}^s$ such that (3-5) is ensured by item (4) of Section 2E.

In the proof of Theorem 2.10 we used only the properties of $\boldsymbol{B}$ and B that were listed in Section 2A1 and the obstacle condition (2-49). More precisely, we never used properties (2-8), (2-9); only (2-10) was used.

But we have all those ingredients now ready for $\mathbb{V}$ and V. Therefore, we have already proved the following result.

Theorem 3.3. *There exists an absolute positive constant c_V such that*

$$\sup_{(F,w,m,f,\lambda)\in\Omega^s} \frac{|\lambda|\mathbb{V}_Q(F,w,m,f,\lambda)}{F} = \sup_{(\alpha,\beta,\gamma)\in G} \frac{V(\alpha,\beta,\gamma)}{\alpha} \geq c_V Q(\log Q)^{1/3}. \tag{3-6}$$

4. A particular martingale transform and the lower estimate of its norm from $L^1(w)$ to $L^{1,\infty}(w)$: the proof of Theorem 2.2

Let us consider a concrete dyadic shift S and prove Theorem 2.2 for it. Theorem 3.3 claims that we can choose a point $(F_0, w_0, m_0, f_0, \lambda_0) \in \Omega^s$ such that some random walk from $\mathcal{W}^s$ (in particular having finitely many steps and finishing at $\partial_w\Omega^s = \{(F, w, m, f, \lambda) \in \partial\Omega^s : w = m\}$) will have the property that

$$\mathbb{E}\mathbf{1}_{\Lambda(\omega)\leq 0} > \frac{c_V}{2} Q(\log Q)^{1/3} \frac{F_0}{\lambda_0}, \tag{4-1}$$

where $(\boldsymbol{F}(\omega), M(\omega), M(\omega), \boldsymbol{f}(\omega), \Lambda(\omega))$ are the final values of the walk. We can now establish the correspondence between ω and points of the interval $I = [0, 1]$. We assume that $(F_0, w_0, m_0, f_0, \lambda_0)$ are starting values of our "martingales" on I. But our random walk also generates by its first step certain numbers $dF, dw, d\lambda, m_1, m_2, m_3, m_4$, and $t \in [0, 1]$, $m_0 = \min(m_i)$.

We call $dF, df, d\lambda$ martingale differences, m_i, $i = 1, \dots, 4$; we call them splittings of m_0.

We associate:

- $(F_0 - dF, w_0 - dw, m_1, f_0 - d\lambda, \lambda_0 - td\lambda)$ with values of our "martingales" on I_{--}.
- $(F_0 - dF, w_0 - dw, m_2, f_0 - d\lambda, \lambda_0 + td\lambda)$ with values of our "martingales" on I_{-+}.
- $(F_0 + dF, w_0 + dw, m_3, f_0 + d\lambda, \lambda_0 - td\lambda)$ with values of our "martingales" on I_{++}.
- $(F_0 + dF, w_0 + dw, m_4, f_0 + d\lambda, \lambda_0 + td\lambda)$ with values of our "martingales" on I_{+-}.

If one or several of these points are already on $\partial_w\Omega^s$ we do not touch them anymore. For the rest of points we have the second step, which is given by new martingale differences (and new splittings, now of each of m_i, $i = 1, \dots, 4$). We continue to associate the points with now grandchildren of $I_{\sigma,\sigma'}$, $\sigma, \sigma' = \pm$. We continue this process for finitely many times, until all the points of the walk hit $\partial_w\Omega^s$, where the process stops.

By our association process we constructed functions with finitely many values, constant on some small dyadic intervals of $D(I)$. These are the functions $\varphi(x)$, $\psi(x)$, $W(x)$, $\Phi(x) = |\varphi(x)|W(x)$, $x \in I = [0, 1]$, $m_0 = \min_I W(x)$, $\langle W\rangle_J \leq Q \min_J W$ for all dyadic intervals of $J \in D(I)$. Moreover, it is easy to check by our construction that we have

$$S(-\varphi) = -\psi + \lambda_0.$$

We established the correspondence between ω and the points x of the interval $I = [0, 1]$. Under this correspondence $\varphi(x)$ is $\boldsymbol{f}(\omega)$, $\psi(x)$ is $\Lambda(\omega)$, $\Phi(x)$ is $\boldsymbol{F}(\omega)$, $W(x)$ is $\boldsymbol{w}(\omega)$, and $m(\omega)$ corresponds to minimums of W on small final dyadic intervals.

Now we use (4-1). It becomes the inequality

$$W\{x \in I : S(-\varphi)(x) \geq \lambda_0\} \geq \frac{c_V}{2} Q(\log Q)^{1/3} \frac{\int |\varphi| W \, dx}{\lambda_0},$$

which proves Theorem 2.2.

References

[Cruz-Uribe et al. 2011] D. V. Cruz-Uribe, J. M. Martell, and C. Pérez, *Weights, extrapolation and the theory of Rubio de Francia*, Operator Theory: Advances and Applications **215**, Springer, 2011. MR Zbl

[Evans and Gariepy 1992] L. C. Evans and R. F. Gariepy, *Measure theory and fine properties of functions*, CRC, Boca Raton, FL, 1992. MR Zbl

[Ivanisvili and Volberg 2017] P. Ivanisvili and A. Volberg, "Martingale transform and square function: some weak and restricted weak sharp weighted estimates", preprint, 2017. arXiv 1711.10578

[Lerner et al. 2009] A. K. Lerner, S. Ombrosi, and C. Pérez, "A_1 bounds for Calderón–Zygmund operators related to a problem of Muckenhoupt and Wheeden", *Math. Res. Lett.* **16**:1 (2009), 149–156. MR Zbl

[Lerner et al. 2017] A. K. Lerner, F. Nazarov, and S. Ombrosi, "On the sharp upper bound related to the weak Muckenhoupt–Wheeden conjecture", preprint, 2017. arXiv 1710.07700

[Nadirashvili et al. 2014] N. Nadirashvili, V. Tkachev, and S. Vlăduţ, *Nonlinear elliptic equations and nonassociative algebras*, Mathematical Surveys and Monographs **200**, American Mathematical Society, Providence, RI, 2014. MR Zbl

[Nazarov et al. 2015] F. Nazarov, A. Reznikov, V. Vasyunin, and A. Volberg, "A Bellman function counterexample to the A_1 conjecture: the blow-up of the weak norm estimates of weighted singular operators", preprint, 2015. arXiv 1506.04710v1

[Nazarov et al. 2016] F. Nazarov, A. Reznikov, V. Vasyunin, and A. Volberg, "On weak weighted estimates of martingale transform", preprint, 2016. arXiv 1612.03958

[Reguera 2011] M. C. Reguera, "On Muckenhoupt–Wheeden conjecture", *Adv. Math.* **227**:4 (2011), 1436–1450. MR Zbl

[Reguera and Thiele 2012] M. C. Reguera and C. Thiele, "The Hilbert transform does not map $L^1(Mw)$ to $L^{1,\infty}(w)$", *Math. Res. Lett.* **19**:1 (2012), 1–7. MR Zbl

[Reznikov et al. 2013] A. Reznikov, V. Vasyunin, and A. Volberg, "Extremizers and Bellman function for martingale weak type inequality", preprint, 2013. arXiv 1311.2133

Received 31 Jul 2017. Revised 10 Feb 2018. Accepted 10 Apr 2018.

FEDOR NAZAROV: nazarov@math.kent.edu
Department of Mathematical Sciences, Kent State University, Kent, OH, United States

ALEXANDER REZNIKOV: reznikov@math.fsu.edu
Department of Mathematics, Michigan State University, East Lansing, MI, United States

VASILY VASYUNIN: vasyunin@pdmi.ras.ru
V. A. Steklov Mathematical Institute, St. Petersburg, Russia

ALEXANDER VOLBERG: volberg@math.msu.edu
Department of Mathematics, Michigan State University, East Lansing, MI, United States

ANALYSIS AND PDE
Vol. 11, No. 8, 2018
dx.doi.org/10.2140/apde.2018.11.2111

msp

TWO-MICROLOCAL REGULARITY OF QUASIMODES ON THE TORUS

FABRICIO MACIÀ AND GABRIEL RIVIÈRE

We study the regularity of stationary and time-dependent solutions to strong perturbations of the free Schrödinger equation on two-dimensional flat tori. This is achieved by performing a second microlocalization related to the size of the perturbation and by analyzing concentration and nonconcentration properties at this new scale. In particular, we show that sufficiently accurate quasimodes can only concentrate on the set of critical points of the average of the potential along closed geodesics.

1. Introduction

The high-frequency analysis of eigenfunctions of elliptic operators on a compact Riemannian manifold has been the subject of intensive study in the past fifty years. To this day, many questions remain open, even in the simplest cases. Here we focus on eigenfunctions of Schrödinger operators on $\mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$, the standard torus endowed with its canonical metric. Eigenfunctions of a Schrödinger operator on $\mathbb{T}^d$ are the solutions to the equation

$$-\Delta u_\lambda(x) + V(x)\, u_\lambda(x) = \lambda^2 u_\lambda(x), \quad x \in \mathbb{T}^d, \qquad \|u_\lambda\|_{L^2(\mathbb{T}^d)} = 1, \tag{1}$$

where the potential V is real-valued and essentially bounded. In the free case $V = 0$, a straightforward computation shows that eigenfunctions of eigenvalue λ^2 are linear combinations of complex exponentials $e^{2i\pi k.x}$ with frequencies $k \in \mathbb{Z}^d$ lying on a circle of radius $\lambda/(2\pi) > 0$ centered at the origin. However, extracting from this exact representation formula an asymptotic description of eigenfunctions in the high-frequency limit $\lambda \to +\infty$ is a hard problem, due to the fact that multiplicities of large eigenvalues can also be very big. Instead, one can try to describe particular features of high-frequency eigenfunctions, such as formation of (asymptotic) singularities.

A natural way to quantify these singularities is through the scale of L^p spaces. This has been a classical topic in harmonic analysis, that originates with the seminal result of [Zygmund 1974] showing that, for $d = 2$ and in the free case, there exists some universal constant C such that any solution u_λ of (1) satisfies $\|u_\lambda\|_{L^4(\mathbb{T}^2)} \leq C$. Later on, Bourgain [1993] conjectured that, again for the free case and when $d \geq 3$, one must have $\|u_\lambda\|_{L^{2d/(d-2)}(\mathbb{T}^d)} \leq C_\delta \lambda^\delta$ for every $\delta > 0$. We refer the reader to [Bourgain 2013; Bourgain and Demeter 2015] for recent progress towards this conjecture. Note that the problem of showing the existence of an index $p > 2$ such that $\|u_\lambda\|_{L^p(\mathbb{T}^d)}$ is uniformly bounded remains open for $d \geq 3$.

Macià takes part in the visiting faculty program of ICMAT and is partially supported by ERC Starting Grant 277778 and the grants MTM2013-41780-P and MTM2017-85934-C3-3-P (MECD). Rivière is partially supported by the Agence Nationale de la Recherche through the Labex CEMPI (ANR-11-LABX-0007-01) and the ANR project GeRaSic (ANR-13-BS01- 0007-01).
MSC2010: 58J51, 35P20, 35Q41, 58J50.
Keywords: quasimodes, Schrödinger operator, semiclassical measures, time-dependent Schrödinger equation.

There are alternative ways to describe the asymptotic structure of the solutions of (1). For instance, notice that a direct corollary of Zygmund's result is that, in the free case, any accumulation point of the sequence of probability measures,

$$\nu_\lambda(dx) = |u_\lambda(x)|^2\, dx,$$

is a probability measure which is absolutely continuous with respect to the Lebesgue measure on $\mathbb{T}^2$ (it has in fact an L^2 density). This result was refined by Jakobson [1997] who showed that the density has to be a trigonometric polynomial whose frequencies enjoy certain geometric constraints. It is natural to try to understand what happens when $d \geq 3$, where no analogue to Zygmund's result is known to hold, or when the Laplacian is perturbed by a lower-order term, such as a potential. Note that the problem of identifying accumulation points of sequences of moduli squares of eigenfunctions has a long history and it is connected to fundamental questions in quantum mechanics.

In dimension $d \geq 3$ and for $V = 0$, Bourgain proved that any accumulation point has to be absolutely continuous even if we do not know a priori that the L^p norms of eigenfunctions are uniformly bounded for small $p > 2$; this result was reported in [Jakobson 1997]. In the same reference, Jakobson obtained partial results on the structure of the densities of accumulation points. These results are based on harmonic analysis techniques and arguments on the geometry of lattice points. Absolute continuity of accumulation points also holds in the case of a nonzero potential $V \in L^\infty(\mathbb{T}^d)$, as was proved by Anantharaman and the first author [Anantharaman and Macià 2014]. The proof of that result is based on methods from semiclassical analysis for the time-dependent Schrödinger equation that were introduced for the particular case $d = 2$ in [Macià 2010]. In fact, the results in [Anantharaman and Macià 2014] apply to the more general problem

$$\widehat{P}_\epsilon(\hbar) u_\hbar = \tfrac{1}{2} u_\hbar + o(\hbar\epsilon_\hbar), \quad \|u_\hbar\|_{L^2(\mathbb{T}^d)} = 1, \tag{2}$$

where $\hbar \to 0^+$ is some semiclassical parameter, and where

$$\widehat{P}_\epsilon(\hbar) := -\tfrac{1}{2}\hbar^2\Delta + \epsilon_\hbar^2 V, \tag{3}$$

with $0 \leq \epsilon_\hbar \leq \hbar$ for $\hbar$ small enough.[1] Among the main ingredients used in this approach are the two-microlocal techniques developed in [Nier 1996; Miller 1996; Fermanian-Kammerer 2000; 2005; Fermanian-Kammerer and Gérard 2002] in a different context. The results in [Anantharaman and Macià 2014] were further extended to treat the case of more general completely integrable systems in [Anantharaman et al. 2015]. This approach can also be used in order to analyze the Schrödinger equation on the planar disk [Anantharaman et al. 2016a; 2016b]. Note that studying the regularity of the solutions to (2) is also related to problems arising in control theory, as was shown by Burq and Zworski [2004]. We refer the reader to [Anantharaman and Léautaud 2014; Anantharaman et al. 2016b; Anantharaman and Macià 2014; Bourgain et al. 2013; Burq and Zworski 2004; 2012; Macià 2011] for perspectives from the point of view of control theory.

A different but related approach consists in studying the wavefront set $\mathrm{WF}_\hbar(u_\hbar)$ of solutions to (2). This was done in a series of works by Wunsch [2008; 2012] and Vasy and Wunsch [2009] dealing

[1] Note that, when $\hbar = \epsilon_\hbar = \lambda^{-1}$, equation (2) is essentially equation (1).

with completely integrable systems in dimension $d = 2$. In these articles, the authors investigated the properties of the semiclassical wavefront set $\mathrm{WF}_\hbar(u_\hbar)$ of solutions to (2) when $0 \le \epsilon_\hbar \le \hbar^{1+\delta}$ with $\delta > 0$. By proving some propagation of second microlocal wavefront sets, they showed that $\mathrm{WF}_\hbar(u_\hbar)$ cannot be reduced to a single geodesic and has to fill a Lagrangian torus — see for instance [Wunsch 2008, Theorem B; 2012, Theorem 3]. Note that, as in [Anantharaman et al. 2015], the results of Vasy and Wunsch hold for general classes of nondegenerate completely integrable systems. Under the assumption that $\hbar^{1-\delta} \ll \epsilon_\hbar \ll 1$, Wunsch also exhibited examples of quasimodes of order $\mathcal{O}(\hbar^\infty)$ for the operator $\widehat{P}_\epsilon(\hbar)$ which concentrate on closed geodesics. This result was reported in [Anantharaman et al. 2015, Section 5.3], and it shows that $\epsilon_\hbar = \hbar$ is the critical size for which one can expect to have singular concentration phenomena for perturbations of the free semiclassical Schrödinger operator $-\frac{1}{2}\hbar^2\Delta$. In particular, for stronger perturbation $\epsilon_\hbar \gg \hbar$, one cannot expect to have uniform bounds for L^p norms even for a small range of p. A notable feature of Wunsch's construction is that the singularity is located on critical points of the potential V restricted to certain closed geodesics. In some sense, this type of singularity is similar to the ones that may occur in the case of Zoll manifolds [Macià and Rivière 2016; 2017]. Motivated by this observation, we will combine the ideas from [Anantharaman and Macià 2014; Macià and Rivière 2016] in order to derive some properties on the regularity of solutions to (2) when $\epsilon_\hbar \gg \hbar$. In particular, we will identify precisely the concentration phenomena that may occur and also show nonconcentration properties by propagation of second microlocal data. Note that, when written in nonsemiclassical terms, the regime we are interested in corresponds to the eigenvalue problem

$$-\Delta u_\lambda(x) + f(\lambda)\, V(x)\, u_\lambda(x) = \lambda^2 u_\lambda(x), \quad x \in \mathbb{T}^d, \qquad \|u_\lambda\|_{L^2(\mathbb{T}^d)} = 1,$$

where $1 \ll f(\lambda) \ll \lambda^2$.

For the sake of simplicity, we will focus on the case of the rational torus $\mathbb{T}^2$ and assume $V \in \mathcal{C}^\infty(\mathbb{T}^2; \mathbb{R})$. However, it is most likely that our analysis could be extended to more general completely integrable systems of dimension 2 following the approach of [Anantharaman et al. 2015]. As the small perturbation regime[2] $0 \le \epsilon_\hbar \le \hbar$ was studied in great detail in all the above references, here we will focus on the strong perturbation regime and we shall assume throughout the article that

$$\lim_{\hbar \to 0^+} \epsilon_\hbar = 0 \quad \text{and} \quad \lim_{\hbar \to 0^+} \hbar \epsilon_\hbar^{-1} = 0. \tag{4}$$

In order to state our results, we need some simple geometric preliminaries. Recall that the geodesics of $\mathbb{T}^2$ are either closed or dense curves. For $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2 - \{0\}$ and $x \in \mathbb{T}^2$, the geodesic $s \mapsto x + s\xi$ is dense provided ξ_1 and ξ_2 are linearly independent over $\mathbb{Q}$; otherwise it is periodic. We denote by $\Omega_1 \subset \mathbb{R}^2 - \{0\}$ the set of ξ that generate a periodic geodesic and by Ω_2 its complement in $\mathbb{R}^2 - \{0\}$. Consider the average of V along geodesics:

$$\mathcal{I}(V)(x, \xi) := \lim_{T \to +\infty} \frac{1}{T} \int_0^T V(x + s\xi)\, ds.$$

[2]Note that, for the nonsemiclassical version, it means that $f(\lambda) \le 1$.

Clearly, $\mathcal{I}(V)$ is a zero-homogeneous function with respect to ξ. Moreover, a classical result by Kronecker implies

$$\mathcal{I}(V)(x,\xi) = \begin{cases} (1/L_\xi)\int_0^{L_\xi} V(x+s(\xi/\|\xi\|))\,ds & \text{if } \xi \in \Omega_1, \\ \int_{\mathbb{T}^2} V(y)\,dy & \text{if } \xi \in \Omega_2, \end{cases}$$

where L_ξ denotes the length of any geodesic with velocity ξ. In particular, despite the fact that $\mathcal{I}(V)$ is not continuous in general, one has $\mathcal{I}(V)(\cdot\,,\xi) \in \mathcal{C}^\infty(\mathbb{T}^2;\mathbb{R})$ for any $\xi \in \mathbb{R}^2 - \{0\}$, and $\|\mathcal{I}(V)\|_{L^\infty(\mathbb{T}^2\times\mathbb{R}^2)} \leq \|V\|_{L^\infty(\mathbb{T}^2)}$.

Then, we define the set of critical geodesics:

$$\mathcal{C}(V) := \{x_0 \in \mathbb{T}^2 : \text{there exists } \xi \in \Omega_1 \text{ such that } \partial_x \mathcal{I}(V)(x_0,\xi) = 0\}. \tag{5}$$

Note that $\mathcal{C}(V)$ is a union of closed geodesics of $\mathbb{T}^2$. For every closed geodesic γ of $\mathbb{T}^2$, we denote by δ_γ the normalized Lebesgue measure along this closed geodesic. Then, we define $\mathcal{N}(V)$ as the convex closure of the set of probability measures δ_γ, where $\gamma \subset \mathcal{C}(V)$. With these conventions in mind, we can state our main result:

Theorem 1.1. *Suppose that $d = 2$ and that (4) holds. Let $(u_\hbar)_{\hbar\to 0^+}$ be a sequence satisfying (2). Then, for any accumulation point ν of the sequence of probability measures*

$$\nu_\hbar(dx) := |u_\hbar(x)|^2\,dx,$$

and for any closed geodesic γ, one has

$$\nu(\gamma) \neq 0 \quad \Longrightarrow \quad \gamma \subset \mathcal{C}(V).$$

Moreover, ν can be decomposed as

$$\nu = f\,dx + \nu_{\mathrm{sing}},$$

where $f \in L^1(\mathbb{T}^2)$ and where $\nu_{\mathrm{sing}} \in \mathcal{N}(V)$.

Recall from the propagation properties of semiclassical measures [Gérard 1991; Zworski 2012] that any ν as in Theorem 1.1 must a priori be a convex combination of the Lebesgue measure and of the measures δ_γ, where γ runs over the set of all closed geodesics. This theorem shows that singular concentration along closed geodesics can only occur along certain closed orbits associated with critical points of the averages of V along closed geodesics. This result is sharp in the sense that Wunsch's construction in [Anantharaman et al. 2015] shows that one can find quasimodes such that $\nu(\gamma) = 1$ for a given closed geodesic. Despite these unavoidable concentration phenomena, Theorem 1.1 also shows that the accumulation points enjoy certain regularity properties. This extra regularity will come out from our analysis by making a second microlocalization of size $\epsilon_\hbar$ along rational directions, and it will be induced by certain Lagrangian tori associated to our problem. Note that these two aspects are close to the situation of Zoll manifolds treated in [Macià and Rivière 2016; 2017]. The main difference is that there exist infinitely many directions where the flow is periodic with periods tending to $+\infty$. We would like to treat these tori of periodic orbits as in these references, and this can be achieved via rescaling the variables along these rational directions; see Section 3D for more details. Finally, as we shall see in Sections 2 and 3, our analysis holds in the more general context of the time-dependent Schrödinger equation.

Organization of the article. Section 2 places our problem in the more general framework of the time-dependent Schrödinger equation associated with $\widehat{P}_\epsilon(\hbar)$: Theorem 1.1 becomes a direct consequence of the more general Theorem 2.1, which deals with the evolution problem. The proof of this result is obtained by characterizing time-dependent semiclassical measures for solutions to the Schrödinger equation. Following a strategy similar to that in [Anantharaman and Macià 2014; Macià 2010], such a characterization can be obtained by using two-microlocal techniques. In Section 3, we introduce the two-microlocal framework of our analysis that is needed to formulate our main results, Theorems 3.6 and 3.7. Section 4 presents several applications of these results. We first give the proof of Theorem 2.1; then we present a structure result for semiclassical measures of the evolution equation, Theorem 4.1, which we apply to compute the propagation of wave packet solutions (Proposition 4.3). This shows that Theorem 2.1 is sharp in some sense. The proofs of the two-microlocal statements of Section 3 are given in Section 5. Finally, the article contains two appendices. Appendix A contains the proof of a geometric result which already appeared in [Macià and Rivière 2016] and which we adapt to the context of $\mathbb{T}^2$. In Appendix B, we collect a few tools from semiclassical analysis.

In the following (except in Appendix B), we will always suppose that $d = 2$ and that (4) holds even if part of the result holds in greater generality.

2. Semiclassical measures for the time-dependent Schrödinger equation

As was already mentioned, Theorem 1.1 is a consequence of our analysis of the time-dependent semiclassical Schrödinger equation:

$$i\hbar\,\partial_t v_\hbar = \widehat{P}_\epsilon(\hbar)v_\hbar, \qquad v_\hbar|_{t=0} = u_\hbar \in L^2(\mathbb{T}^2), \qquad \|u_\hbar\|_{L^2} = 1. \tag{6}$$

For the sake of simplicity, we shall focus on sequences of initial data oscillating at the frequency $\hbar^{-1}$. Thus, we will always assume the following properties hold:

$$\limsup_{\hbar\to 0} \|\mathbf{1}_{[R,\infty)}(-\hbar^2\Delta)u_\hbar\|_{L^2(M)} \to 0 \quad \text{as } R \to \infty, \tag{7}$$

$$\limsup_{\hbar\to 0} \|\mathbf{1}_{[0,\delta]}(-\hbar^2\Delta)u_\hbar\|_{L^2(M)} \to 0 \quad \text{as } \delta \to 0^+. \tag{8}$$

Fix now a sequence of time scales $(\tau_\hbar)_{\hbar\to 0^+}$ such that

$$\lim_{\hbar\to 0^+} \tau_\hbar = +\infty.$$

We will deal with time-scaled solutions to the perturbed Schrödinger equation. More precisely, if $v_\hbar$ is a solution to (6), then we shall study the behavior of

$$t \mapsto v_\hbar(\tau_\hbar t, \cdot\,).$$

As we will see below, the scale $\tau_\hbar = \epsilon_\hbar^{-1}$ is critical for this problem, and Theorem 1.1 follows from the analysis of the time-dependent equation in the regime $\tau_\hbar \gg \epsilon_\hbar^{-1}$.

2A. ***Time-dependent semiclassical measures.*** For a given t in $\mathbb{R}$, we denote the Wigner distribution at time t by

$$\langle w_\hbar(t), a\rangle := \langle v_\hbar(t), \mathrm{Op}_\hbar^w(a)\, v_\hbar(t)\rangle, \tag{9}$$

where $\mathrm{Op}_\hbar^w(a)$ is an $\hbar$-pseudodifferential operator with principal symbol $a \in \mathcal{C}_c^\infty(T^*\mathbb{T}^2)$ — see Appendix B. Above, $v_\hbar(t)$ denotes the solution at time t of (6) with initial conditions satisfying the oscillating assumptions (7) and (8). This quantity represents the distribution of the L^2-mass of the solution to (6) in the phase space $T^*\mathbb{T}^2$. According to [Macià 2009], we can extract a subsequence $\hbar_n \to 0^+$ as $n \to +\infty$ such that, for every a in $\mathcal{C}_c^\infty(T^*\mathbb{T}^2)$ and for every θ in $L^1(\mathbb{R})$,

$$\lim_{\hbar_n \to 0^+} \int_{\mathbb{R}\times T^*\mathbb{T}^2} \theta(t)\langle w_{\hbar_n}(t\tau_{\hbar_n}), a\rangle \, dt = \int_{\mathbb{R}\times T^*\mathbb{T}^2} \theta(t)\, a(x,\xi)\, \mu(t, dx, d\xi)\, dt,$$

where, for a.e. t in $\mathbb{R}$, $\mu(t)$ is a finite positive Radon measure on $T^*\mathbb{T}^2$. Recall also that, for a.e. $t \in \mathbb{R}$, $\mu(t)$ is in fact a *probability measure* which does not put any mass on the zero section, thanks to the frequency assumption (8). In other words,

$$\mu(t)(\mathring{T}^*\mathbb{T}^2) = 1 \quad \text{for a.e. } t \in \mathbb{R}, \tag{10}$$

where

$$\mathring{T}^*\mathbb{T}^2 := \{(x,\xi) \in T^*\mathbb{T}^2 : \xi \neq 0\}.$$

Moreover, for a.e. t in $\mathbb{R}$, $\mu(t)$ is *invariant by the geodesic flow* φ^s on $T^*\mathbb{T}^2$.

For instance, $\mu(t)$ can be the normalized Lebesgue measure along a closed orbit of the geodesic flow. We will denote by $\mathcal{M}(\tau,\epsilon)$ the set of accumulation points of the sequences $(\mu_\hbar)$, where $\mu_\hbar(t,\cdot) := w_\hbar(t\tau_\hbar,\cdot)$, as the sequence of initial data $(u_\hbar)$ varies among normalized sequences satisfying (7) and (8). Similarly, one can define $\mathcal{N}(\tau,\epsilon)$ to be the set of accumulation points of the sequences $(n_\hbar)$ of time-dependent probability measures on $\mathbb{T}^2$, $n_\hbar(t,dx) := |v_\hbar(t\tau_\hbar, x)|^2\, dx$, obtained by letting the initial data vary among sequences satisfying (7), (8). Using (7), one can verify that

$$\mathcal{N}(\tau,\epsilon) = \left\{ \int_{\mathbb{R}^2} \mu(t,x,d\xi) : \mu \in \mathcal{M}(\tau,\epsilon) \right\}. \tag{11}$$

2B. ***Statement of the results.*** In order to relate the time-dependent approach to the quasimode case, we can remark that, given a sequence of quasimodes $(u_\hbar)_{\hbar\to 0^+}$ satisfying (2), we can always find a sequence of time scales $(\tau_\hbar)$ such that

$$\lim_{\hbar\to 0} \tau_\hbar \epsilon_\hbar = +\infty,$$

and, for every $t \in \mathbb{R}$,

$$\lim_{\hbar\to 0} \|v_\hbar(\tau_\hbar t,\cdot) - e^{-i\tau_\hbar t/(2\hbar)} u_\hbar\|_{L^2(\mathbb{T}^2)} = 0,$$

where $v_\hbar$ denotes the solution to (6) with initial condition $u_\hbar$. This choice of $(\tau_\hbar)$ ensures that any accumulation point ν of the sequence of probability measures $(|u_\hbar|^2\, dx)$ belongs to $\mathcal{N}(\tau,\epsilon)$ (even though it is constant in t), since it is also an accumulation point of $(|v_\hbar(\tau_\hbar t,\cdot)|^2\, dx)$. In particular, Theorem 1.1 follows from the more general statement:

Theorem 2.1. *Suppose that*

$$\lim_{\hbar\to 0} \tau_\hbar \epsilon_\hbar = +\infty.$$

Let $t \mapsto \nu(t)$ be an element of $\mathcal{N}(\tau, \epsilon)$. Then, for any closed geodesic γ not included inside $\mathcal{C}(V)$ and for a.e. t in $\mathbb{R}$, one has

$$\nu(t)(\gamma) = 0.$$

Moreover, $\nu(t)$ can be decomposed as

$$\nu(t) = f(t)\, dx + \nu_{\mathrm{sing}}(t),$$

where, for a.e. t in $\mathbb{R}$, $f(t) \in L^1(\mathbb{T}^2)$ and $\nu_{\mathrm{sing}}(t) \in \mathcal{N}(V)$.

The first step in the proof of this result is the partition of $\mathbb{R}^2 - \{0\}$ into φ^s-invariant subsets that was used in [Macià 2010; Anantharaman and Macià 2014]. Recall that $\Lambda \subset \mathbb{Z}^2$ is a *primitive lattice of rank 1* provided that $\dim\langle\Lambda\rangle = 1$ and that $\langle\Lambda\rangle \cap \mathbb{Z}^2 = \Lambda$, where $\langle\Lambda\rangle$ is the linear subspace of $\mathbb{R}^2$ spanned by Λ. We introduce the invariant set of rational covectors

$$\Omega_1 = \bigsqcup_{\Lambda \text{ rank-1 primitive}} \Lambda^\perp - \{0\},$$

and its complement Ω_2 inside $\mathbb{R}^2 - \{0\}$, which is still invariant. Observe that this is consistent with the conventions of the Introduction. Because of (10), we can decompose the measure as follows:

$$\mu(t) = \mu(t)\rceil_{\mathbb{T}^2\times\Omega_2} + \sum_{\Lambda \text{ rank-1 primitive}} \mu(t)\rceil_{\mathbb{T}^2\times\Lambda^\perp - \{0\}}. \tag{12}$$

As a consequence of the invariance by the geodesic flow, it can be verified that $\mu(t)\rceil_{\mathbb{T}^2\times\Omega_2}$ is in fact independent of the x-variable. Hence, in order to prove Theorem 2.1, one only has to study the regularity of $\mu(t)\rceil_{\mathbb{T}^2\times\Lambda^\perp - \{0\}}$ for every rank-1 primitive sublattice Λ. This will be achieved using two-microlocal tools adapted to this problem. The end of the proof of Theorem 2.1 is presented in Section 4A. For time scales $\tau_\hbar = \mathcal{O}(\epsilon_\hbar^{-1})$, we obtain a more precise result, in the sense that each component of the time-dependent semiclassical measure $\mu(t)$ according to the partition (12) can be completely determined from the initial data that were used to generate it. Again, the relation with the sequence of initial data is elucidated using the class of two-microlocal semiclassical measures that will be introduced in the next section. A precise statement is given in Theorem 4.1, Section 4B.

Finally, in Section 4C, we provide explicit computations of semiclassical measures associated to wave-packets (Proposition 4.3) that yield:

(1) If $\tau_\hbar \epsilon_\hbar \to 0$, then
$$\{\delta_\gamma : \gamma \text{ periodic geodesic of } \mathbb{T}^2\} \subset \mathcal{N}(\tau, \epsilon).$$

(2) If $\tau_\hbar = \epsilon_\hbar^{-1}$, then
$$\{\delta_\gamma : \gamma \in \mathcal{C}(V)\} \subset \mathcal{N}(\tau, \epsilon).$$

3. Invariance and propagation of two-microlocal distributions

We now present our main result on the two-microlocal structure of solutions to the time-dependent Schrödinger equation along covectors in Ω_1. In particular, we show how solutions of (6) can concentrate along rational covectors.

Before stating the result, we need some additional notation. For every primitive rank-1 lattice Λ of $\mathbb{Z}^2$, we set $\mathfrak{e}_\Lambda$ to be an element in Λ such that $\mathbb{Z}\mathfrak{e}_\Lambda = \Lambda$, and $\mathfrak{e}_\Lambda^\perp$ to be the vector of same length which is directly orthogonal to $\mathfrak{e}_\Lambda$. We define

$$L_\Lambda := \|\mathfrak{e}_\Lambda\|.$$

We define two Hamiltonian maps associated to Λ as follows:

$$H_\Lambda(\xi) := \frac{1}{L_\Lambda}\langle \xi, \mathfrak{e}_\Lambda\rangle \quad \text{and} \quad H_\Lambda^\perp(\xi) := \frac{1}{L_\Lambda}\langle \xi, \mathfrak{e}_\Lambda^\perp\rangle.$$

Note that $(H_\Lambda, H_\Lambda^\perp)$ defines a (nondegenerate) completely integrable system and that

$$\|\xi\|^2 = H_\Lambda(\xi)^2 + H_\Lambda^\perp(\xi)^2.$$

3A. ***Two-microlocal distributions.*** We aim at studying the concentration of solutions to (6) over $\mathbb{T}^2 \times \Lambda^\perp$, where $\Lambda \subset \mathbb{Z}^2$ is a primitive rank-1 sublattice and where $\Lambda^\perp$ denotes the set of covectors ξ such that $H_\Lambda(\xi) = 0$. For that purpose, we consider a two-microlocal scale $\alpha_\hbar \to 0^+$ satisfying $\hbar\alpha_\hbar^{-1} \to 0$ and we define the following two-microlocal Wigner distribution:

$$w_{\Lambda,\hbar}(t) : a \in \mathcal{C}_c^\infty(T^*\mathbb{T}^2 \times \widehat{\mathbb{R}}) \mapsto \left\langle v_\hbar(t), \mathrm{Op}_\hbar^w\left(a\left(x, \xi, \frac{H_\Lambda(\xi)}{\alpha_\hbar}\right)\right) v_\hbar(t)\right\rangle.$$

Above, $\widehat{\mathbb{R}}$ is the compactified space $\mathbb{R} \cup \{\pm\infty\}$, $v_\hbar(t)$ is the solution of (6) at time t, and $\mathrm{Op}_\hbar^w(a)$ is a $\hbar$-pseudodifferential operator — see Appendix B.

Remark 3.1. Recall from (28) in Appendix B that the following useful relation holds:

$$\mathrm{Op}_\hbar^w\left(a\left(x, \xi, \frac{H_\Lambda(\xi)}{\alpha_\hbar}\right)\right) = \mathrm{Op}_{\hbar\alpha_\hbar^{-1}}^w(a(x, \alpha_\hbar\xi, H_\Lambda(\xi))),$$

and that we have made the assumption that $\hbar\alpha_\hbar^{-1} \to 0$. Therefore, the operators involved in the definition of $w_{\Lambda,\hbar}$ are semiclassical pseudodifferential operators whose symbolic calculus enjoys a gain of $\hbar\alpha_\hbar^{-1}$.

Remark 3.2. The distributions $w_{\Lambda,\hbar}$ were introduced in [Macià 2010; Anantharaman and Macià 2014] for the critical case $\alpha_\hbar = \hbar$ under a slightly different form. There, the two microlocal variable η varies in the two-point compactification of $\langle\Lambda\rangle$. Of course, this is completely equivalent to our formulation for the two-dimensional torus, but turns out to be relevant when dealing with the higher-dimensional case. As we will see, the fact that the two-microlocal scale is asymptotically bigger than $\hbar$ implies that the limiting objects are of a different nature than those obtained in [Macià 2010; Anantharaman and Macià 2014]. When $\hbar\alpha_\hbar^{-1} \to 0$, they are global variants on the torus of the two-scale semiclassical measures introduced in [Fermanian-Kammerer 2005] — see also [Anantharaman and Léautaud 2014] for a related construction on the torus, in a context related to that of [Anantharaman and Macià 2014].

Recall that we introduced a time scale $\tau_\hbar \to \infty$. From now on, we shall fix the two-microlocal scale as follows:

$$\alpha_\hbar := \begin{cases} 1/\tau_\hbar & \text{if } \tau_\hbar\epsilon_\hbar^{-1} \to 0, \\ \epsilon_\hbar & \text{otherwise.} \end{cases} \tag{13}$$

As we shall explain in Section 5A, we can extract a subsequence $\hbar_n \to 0^+$ such that, for any $a \in \mathcal{C}_c^\infty(T^*\mathbb{T}^2 \times \widehat{\mathbb{R}})$ and for any $\theta \in L^1(\mathbb{R})$,

$$\lim_{n\to+\infty} \int_{\mathbb{R}} \theta(t)\langle w_{\Lambda,\hbar_n}(t\tau_{\hbar_n}), a\rangle\, dt = \int_{\mathbb{R}} \theta(t)\left(\int_{T^*\mathbb{T}^2\times\widehat{\mathbb{R}}} a(x,\xi,\eta)\,\mu_\Lambda(t,dx,d\xi,d\eta)\right) dt,$$

where, for a.e. t in $\mathbb{R}$, $\mu_\Lambda(t)$ is an element of $\mathcal{B}'$ for some Banach space $\mathcal{B}$ that we will define in Section 5A. We denote by $\mathcal{M}_\Lambda(\tau,\epsilon)$ the set of accumulation points obtained in this manner for initial data varying among subsequences verifying (7) and (8). The main new result of this article describes some invariance and propagation properties of these quantities depending on the relative sizes of $\tau_\hbar$ and $\epsilon_\hbar$.

For every primitive rank-1 sublattice, one has (see Remark 5.3)

$$\mathcal{M}(\tau,\epsilon) = \left\{ \int_{\widehat{\mathbb{R}}} \mu_\Lambda(t,x,\xi,d\eta) : \mu_\Lambda \in \mathcal{M}_\Lambda(\tau,\epsilon) \right\}. \tag{14}$$

3B. ***First properties.*** Before proving our main results, we will verify a few preliminary results.

Proposition 3.3. *Let $\mu_\Lambda(t)$ be an element of $\mathcal{M}_\Lambda(\tau,\epsilon)$. Then, for a.e. t in $\mathbb{R}$, $\mu_\Lambda(t)$ is a positive finite Radon measure concentrated on $\mathring{T}^*\mathbb{T}^2 \times \widehat{\mathbb{R}}$.*

In what follows, we write

$$\tilde{\mu}_\Lambda(t) := \mu_\Lambda(t)\rceil_{\mathring{T}^*\mathbb{T}^2\times\mathbb{R}}, \quad \tilde{\mu}^\Lambda(t) := \mu_\Lambda(t)\rceil_{\mathring{T}^*\mathbb{T}^2\times\{\pm\infty\}}.$$

Hence, we can split the two-microlocal measure as

$$\mu_\Lambda(t) = \tilde{\mu}_\Lambda(t) + \tilde{\mu}^\Lambda(t). \tag{15}$$

The measure $\tilde{\mu}_\Lambda(t)$ describes in some sense the way the solutions of (6) concentrate in an $\epsilon_\hbar$-neighborhood of the rational direction $\Lambda^\perp$. We now give some other simple properties of these functionals which are analogous to the ones satisfied by time-dependent semiclassical measures [Macià 2009]. We shall also verify:

Proposition 3.4. *Let $\mu_\Lambda(t) \in \mathcal{M}_\Lambda(\tau,\epsilon)$. Then*:

(1) *$\tilde{\mu}_\Lambda(t)$ is a (finite) positive measure on $T^*\mathbb{T}^2 \times \mathbb{R}$ whose support is contained in $\mathbb{T}^2 \times (\Lambda^\perp - \{0\}) \times \mathbb{R}$.*

(2) *For every a in $\mathcal{C}_c^\infty(T^*\mathbb{T}^2 \times \widehat{\mathbb{R}})$,*

$$\langle \tilde{\mu}_\Lambda(t), \xi.\partial_x a\rangle = \langle \tilde{\mu}^\Lambda(t), \xi.\partial_x a\rangle = 0.$$

Neither Proposition 3.3, nor part (1) of Proposition 3.4 uses that the functions used to generate $\mu_\Lambda(t)$ are solutions to (6). This fact is only used in the second part of Proposition 3.4. Note that all these properties follow from standard arguments which need to be slightly adapted in order to fit into the two-microlocal set-up — see Section 5 for details.

3C. ***Main results.*** Consider the Hamiltonian flow $\varphi_{H_\Lambda^\perp}$ associated with $H_\Lambda^\perp$. Note that, for a continuous function b on $T^*\mathbb{T}^2 \times \widehat{\mathbb{R}}$, we can define the average along this L_Λ-periodic flow as

$$\mathcal{I}_\Lambda(b)(x,\xi,\eta) := \frac{1}{L_\Lambda}\int_0^{L_\Lambda} b(\varphi^s_{H_\Lambda^\perp}(x,\xi),\eta)\, ds.$$

A direct computation gives

$$\mathcal{I}_\Lambda(b)(x,\xi,\eta) = \frac{1}{L_\Lambda}\int_0^{L_\Lambda} b\left(x + s\frac{\mathfrak{e}_\Lambda^\perp}{L_\Lambda}, \xi, \eta\right) ds = \sum_{k\in\Lambda} \hat{b}_k(\xi,\eta) e^{2i\pi k.x},$$

provided b has the Fourier expansion

$$b(x,\xi,\eta) = \sum_{k\in\mathbb{Z}^2} \hat{b}_k(\xi,\eta)\, e^{2i\pi k.x}.$$

Moreover, if $\mathcal{I}(b)$ denotes the average of b along the geodesic flow

$$\varphi^s(x,\xi) = (x + s\xi, \xi)$$

on $T^*\mathbb{T}^2$, then the following holds:

$$\mathcal{I}(b)(x,\xi,\eta) = \mathcal{I}_\Lambda(b)(x,\xi,\eta), \quad \text{provided that } \xi \in \Lambda^\perp - \{0\}. \tag{16}$$

In the case where b only depends on x, as is the case with $b = V$, it is easy to check that $\mathcal{I}_\Lambda(V)$ does not depend on ξ and therefore we can identify it with an element in $\mathcal{C}^\infty(\mathbb{T}^2;\mathbb{R})$.

Remark 3.5. Part (2) of Proposition 3.4 implies that $\mu_\Lambda(t)$ is invariant under the geodesic flow φ^s. For b in $\mathcal{C}_c^\infty(T^*\mathbb{T}^2\times\mathbb{R})$, this observation combined with part (1) in Proposition 3.4 and identity (16) implies that, for a.e. t in $\mathbb{R}$,

$$\langle \mu_\Lambda(t), b\rangle = \langle \mu_\Lambda(t), \mathcal{I}_\Lambda(b)\rangle.$$

We shall use this property several times in our proof of Theorem 3.6 below.

We need to define an auxiliary Hamiltonian function on $\mathbb{T}^2\times\Lambda^\perp\times\mathbb{R}$

$$p_\Lambda^V\left(x,\sigma\frac{\mathfrak{e}_\Lambda^\perp}{L_\Lambda},\eta\right) := \tfrac{1}{2}\eta^2 + \mathcal{I}_\Lambda(V)(x). \tag{17}$$

Denote by $\varphi_{p_\Lambda^V}^t$ the flow of the vector field on $\mathbb{T}^2\times\Lambda^\perp\times\mathbb{R}$:

$$\eta\frac{\mathfrak{e}_\Lambda}{L_\Lambda}.\partial_x - \frac{\mathfrak{e}_\Lambda}{L_\Lambda}.\partial_x\mathcal{I}_\Lambda(V)\partial_\eta.$$

This is the Hamiltonian vector field associated to p_Λ^V with respect to the symplectic form obtained by taking the push-forward of the canonical symplectic form on $T^*\mathbb{T}^2$ via the diffeomorphism

$$T^*\mathbb{T}^2 \ni (x,\xi) \mapsto \left(x, H_\Lambda^\perp(x,\xi)\frac{\mathfrak{e}_\Lambda^\perp}{L_\Lambda}, H_\Lambda(x,\xi)\right) \in \mathbb{T}^2\times\Lambda^\perp\times\mathbb{R}. \tag{18}$$

The flow $\varphi_{p_\Lambda^V}^t$ commutes with $\varphi_{H_\Lambda^\perp}^s$ when acting on $\mathbb{T}^2\times\Lambda^\perp\times\mathbb{R}$.

We are now ready to state the main results of this article. The first one concerns the "compact" part of these two-microlocal distributions. Their possible behaviors are classified according to the limit of $\tau_\hbar\epsilon_\hbar$.

Theorem 3.6 (invariance and propagation near Λ). *Let Λ be a primitive rank-1 sublattice and let μ_Λ be an element of $\mathcal{M}_\Lambda(\tau,\epsilon)$ obtained as the limit of $(w_{\Lambda,\hbar}(t\tau_\hbar))$. Denote by μ_Λ^0 the limit of $(w_{\Lambda,\hbar}(0))$. The following results hold*:

(1) *If* $\tau_\hbar \epsilon_\hbar \to 0$ *as* $\hbar \to 0^+$, *then* $t \mapsto \tilde{\mu}_\Lambda(t)$ *is continuous, and one has, for every* a *in* $\mathcal{C}^0_c(\mathbb{T}^2 \times \Lambda^\perp \times \mathbb{R})$,

$$\tilde{\mu}_\Lambda(t)(a) = \tilde{\mu}^0_\Lambda(\mathcal{I}_\Lambda(a) \circ \varphi^t_{p^0_\Lambda}).$$

(2) *If* $\tau_\hbar \epsilon_\hbar \to c > 0$ *as* $\hbar \to 0^+$, *then* $t \mapsto \tilde{\mu}_\Lambda(t)$ *is continuous, and one has, for every* a *in* $\mathcal{C}^0_c(\mathbb{T}^2 \times \Lambda^\perp \times \mathbb{R})$,

$$\tilde{\mu}_\Lambda(t)(a) = \tilde{\mu}^0_\Lambda(\mathcal{I}_\Lambda(a) \circ \varphi^{ct}_{p^V_\Lambda}).$$

(3) *If* $\tau_\hbar \epsilon_\hbar \to +\infty$ *as* $\hbar \to 0^+$, *then one has, for a.e.* t *in* $\mathbb{R}$ *and, for every* a *in* $\mathcal{C}^0_c(\mathbb{T}^2 \times \Lambda^\perp \times \mathbb{R})$,

$$\text{for all } s \in \mathbb{R}, \quad \tilde{\mu}_\Lambda(t)(a) = \tilde{\mu}_\Lambda(t)(a \circ \varphi^s_{p^V_\Lambda}).$$

Equivalently, this theorem says that, besides invariance by the geodesic flow, the solutions of (6) satisfy some extra invariance properties in a shrinking neighborhood of the rational direction at least for times $\tau_\hbar \gg \epsilon_\hbar^{-1}$. For shorter times, the concentration in this shrinking neighborhood is completely determined by the initial data. The proof of this theorem is given in Section 5. Note that, when $\tau_\hbar \epsilon_\hbar \to 0$, the conclusion of part (1) holds even if $\epsilon_\hbar = \hbar$; this will be clear from the proof. Section 5.1 in [Anantharaman et al. 2015] provides explicit computations of two-microlocal semiclassical measures in that regime.

It is interesting to compare part (2) of Theorem 3.6 with its counterpart in [Anantharaman and Macià 2014], where the regime $\epsilon_\hbar = \hbar$ is studied in detail in any dimension (not only in the two-dimensional case analyzed here). First, the nature of the limiting object $\tilde{\mu}_\Lambda$ is rather different in that setting. It is no longer a positive measure, but rather a measure taking values in the set of Wigner transforms of positive Hermitian trace-class operators on the space $L^2(\mathbb{T}_\Lambda)$.[3] As a result, time-dependent semiclassical measures are absolutely continuous with respect to the Lebesgue measures in the x-variable. In that setting, the role of the flow $\varphi^s_{p^V_\Lambda}$ is played by the quantum flow $e^{-is(D^2_\Lambda + \mathcal{I}_\Lambda(V))}$ — see Corollary 25 in [Anantharaman and Macià 2014] for a precise statement.

The part at infinity satisfies an additional regularity property. Indeed, if we define

$$\mathcal{I}_0(a)(\xi, \eta) := \int_{\mathbb{T}^2} a(y, \xi, \eta)\, dy,$$

then the following holds:

Theorem 3.7 (regularity at infinity). *Let* Λ *be a primitive rank-*1 *sublattice and let* $\mu_\Lambda(t)$ *be an element of* $\mathcal{M}_\Lambda(\tau, \epsilon)$. *Then, one has, for every* a *in* $\mathcal{C}^\infty_c(\mathbb{T}^2 \times \mathbb{R}^2 \times \widehat{\mathbb{R}})$ *and for a.e.* t *in* $\mathbb{R}$,

$$\langle \tilde{\mu}^\Lambda(t), \mathcal{I}_\Lambda(a) - \mathcal{I}_0(a) \rangle = 0.$$

In particular, the measure $\tilde{\mu}^\Lambda(t)\rceil_{\mathbb{T}^2 \times \Lambda^\perp \times \widehat{\mathbb{R}}}$ *is constant in* x.

In other words, the part at infinity has no (nonzero) Fourier coefficients in the Λ-direction. As for Theorem 3.6, this result depends highly on the choice of two-microlocal scale we have fixed from the beginning, and other scalings would yield other properties. The first conclusion of this theorem is proved in Section 5. The last assertion follows from the invariance[4] of $\tilde{\mu}^\Lambda(t)$ under the geodesic flow, which

[3] This space consists of those functions in $L^2(\mathbb{T}^2)$ that are invariant by translations in the direction $\Lambda^\perp$.

[4] Recall also that $\mu_\Lambda(t)$ is supported on $\mathring{T}^*\mathbb{T}^2 \times \widehat{\mathbb{R}}$.

implies that for every $a \in \mathcal{C}_c^0(T^*\mathbb{T}^2 \times \widehat{\mathbb{R}})$

$$\langle \tilde{\mu}^\Lambda(t)\rceil_{\mathbb{T}^2\times\Lambda^\perp\times\widehat{\mathbb{R}}}, a\rangle = \langle \tilde{\mu}^\Lambda(t)\rceil_{\mathbb{T}^2\times\Lambda^\perp\times\widehat{\mathbb{R}}}, \mathcal{I}_\Lambda(a)\rangle = \langle \tilde{\mu}^\Lambda(t)\rceil_{\mathbb{T}^2\times\Lambda^\perp\times\widehat{\mathbb{R}}}, \mathcal{I}_0(a)\rangle.$$

Note also that the conclusion of Theorem 3.7 holds in the regime $\epsilon_\hbar = \hbar$ (in any dimension); see part (ii) of Theorem 12 in [Anantharaman and Macià 2014].

3D. ***Comparison with Zoll manifolds.*** Theorem 3.6 shares also a lot of similarities with our main result on semiclassical measures for perturbations of Zoll Laplacians in [Macià and Rivière 2016, Section 2.2]. In that case, we were considering the semiclassical operator

$$-\tfrac{1}{2}\hbar^2\Delta_g + \epsilon_\hbar^2 V,$$

where Δ_g is the Laplace Beltrami operator associated to a certain Zoll metric (say the standard metric on the canonical sphere). In the present article, we are analyzing the semiclassical measures associated to the same Schrödinger operator $\widehat{P}_\epsilon(\hbar)$. Studying the "compact" part of elements inside $\mathcal{M}_\Lambda(\tau,\epsilon)$ is equivalent to understanding the solutions of (6) near submanifolds

$$\mathbb{T}^2 \times \Lambda^\perp := \{(x,\xi) \in T^*\mathbb{T}^2 : H_\Lambda(\xi) = 0\},$$

where the geodesic flow is periodic as in the Zoll case. In order to make the comparison clearer and to justify the rescaling of order $\epsilon_\hbar$, we can rewrite our operator in a form which is very close to what we did in the Zoll framework; i.e.,

$$\widehat{P}_\epsilon(\hbar) = \frac{1}{2}\operatorname{Op}_\hbar^w(H_\Lambda^\perp)^2 + \epsilon_\hbar^2 \operatorname{Op}_\hbar^w\left(\frac{1}{2}\left(\frac{H_\Lambda}{\epsilon_\hbar}\right)^2 + V\right).$$

Thus, as in the Zoll case, we perturb in some sense a semiclassical operator $\operatorname{Op}_\hbar^w(H_\Lambda^\perp)^2$ associated to a "periodic" Hamiltonian flow and we obtain limit quantities which are invariant by the periodic flow and the Hamiltonian perturbation.

The main difference with the Zoll setting is that the perturbation depends on rescaled variables

$$\left(x, H_\Lambda^\perp(\xi), \frac{H_\Lambda(\xi)}{\epsilon_\hbar}\right) \in \mathbb{T}^2 \times \mathbb{R}^2 \simeq T^*\mathbb{T}^2.$$

For that reason, it is natural to test our Wigner distributions against symbols depending on these rescaled variables. Another notable difference with [Macià and Rivière 2016] is that, in the Zoll case, the critical time scale is of order $\epsilon_\hbar^{-2}$, while here, due to the use of rescaled variables, it is much shorter, i.e., of order $\epsilon_\hbar^{-1}$. Finally, in the Zoll case, a natural question was to discuss the case where the Radon transform of the perturbation identically vanishes [Macià and Riviere 2017]. Here, we emphasize that the $H_\Lambda^\perp$-average of the perturbation, namely $\frac{1}{2}(H_\Lambda/\epsilon_\hbar)^2 + \mathcal{I}_\Lambda(V)$ *cannot be equal to a constant* for this choice of two-microlocal rescaling.

4. Applications of the two-microlocal results

We present some applications of the results of the preceding section.

4A. ***Proof of Theorem 2.1.*** Recall that only the structure of the terms $\mu(t)\rceil_{\mathbb{T}^2\times\Lambda^\perp-\{0\}}$ in the decomposition (12) needs to be clarified. Thanks to (14) and to Proposition 3.4, we deduce

$$\mu(t)\rceil_{\mathbb{T}^2\times\Lambda^\perp-\{0\}} = \mu(t)\rceil_{\mathbb{T}^2\times\Lambda^\perp} = \int_{\mathbb{R}} \tilde{\mu}_\Lambda(t,\cdot,d\eta)\rceil_{\mathbb{T}^2\times\Lambda^\perp} + \int_{\{\pm\infty\}} \tilde{\mu}^\Lambda(t,\cdot,d\eta)\rceil_{\mathbb{T}^2\times\Lambda^\perp}.$$

According to Theorem 3.7, the contribution from the part at infinity is independent of x. Hence, we are left with studying the regularity of the measures on $\mathbb{T}^2$:

$$\int_{\Lambda^\perp\times\mathbb{R}} \tilde{\mu}_\Lambda(t,\cdot,d\xi,d\eta).$$

The measure $\tilde{\mu}_\Lambda$ is invariant under the Hamiltonian flow $\varphi^t_{H^\perp_\Lambda}$ (see Remark 3.5) and, by part (3) of Theorem 3.6, it is also invariant under the Hamiltonian flow $\varphi^t_{p^V_\Lambda}$, which commutes with $\varphi^t_{H^\perp_\Lambda}$. Using Appendix A, which describes the regularity of bi-invariant measures, we can conclude the proof of Theorem 2.1. More specifically, part (1) follows from Proposition A.1 and part (2) from Corollary A.3.

4B. ***Semiclassical measures up the critical time scale $\tau_\hbar = \epsilon_\hbar^{-1}$.*** At the time scales up to the critical scale $\epsilon_\hbar^{-1}$, we can completely determine μ_t in terms of the initial data:

Theorem 4.1. *Let $\mu \in \mathcal{M}(\tau,\epsilon)$. Suppose that it is generated by some sequence of initial data $(u_\hbar)_{\hbar\to 0^+}$. For every rank-1 primitive lattice Λ, let $\tilde{\mu}^0_\Lambda$ be the restriction to $\mathbb{T}^2\times\Lambda^\perp\times\mathbb{R}$ of the two-microlocal measure associated with $(u_\hbar)_{\hbar\to 0^+}$, and denote by μ^0 the semiclassical measure of $(u_\hbar)_{\hbar\to 0^+}$:*

(1) *If $\tau_\hbar = \epsilon_\hbar^{-1}$, then, for every $a \in \mathcal{C}^0_c(\mathbb{T}^2\times\mathbb{R}^2)$, the following holds*:

$$\int_{\mathbb{T}^2\times\mathbb{R}^2} a(x,\xi)\,\mu(t,dx,d\xi) = \int_{\mathbb{T}^2\times\mathbb{R}^2} \mathcal{I}_0(a)(\xi)\,\mu^0(dx,d\xi) + \sum_{\Lambda \text{ rank-1 primitive}} \int_{\mathbb{T}^2\times\Lambda^\perp\times\mathbb{R}} (\mathcal{I}_\Lambda(a)-\mathcal{I}_0(a))(\varphi^t_{p^V_\Lambda}(x,\xi,\eta))\,\tilde{\mu}^0_\Lambda(dx,d\xi,d\eta).$$

(2) *If $\tau_\hbar\epsilon_\hbar \to 0$, then the same result holds, provided we replace $\varphi^t_{p^V_\Lambda}$ by $\varphi^t_{p^0_\Lambda}$ in the formula above.*

The proof is as follows. Let $\mu \in \mathcal{M}(\tau,\epsilon)$, and decompose it as in (12). Using the lift property (14), we can further decompose μ as follows:

$$\mu(t) = \mu(t)\rceil_{\mathbb{T}^2\times\Omega_2} + \sum_{\Lambda \text{ rank-1 primitive}} \int_{\{\pm\infty\}} \tilde{\mu}^\Lambda(t,d\eta)\rceil_{\mathbb{T}^2\times\Lambda^\perp} + \sum_{\Lambda \text{ rank-1 primitive}} \int_{\mathbb{R}} \tilde{\mu}_\Lambda(t,\cdot,d\eta)\rceil_{\mathbb{T}^2\times\Lambda^\perp}.$$

Thanks to the invariance by the geodesic flow and to Theorem 3.7, we can conclude one more time that the first two terms on the right-hand side of the equality are independent of x. Thanks to the second part of Theorem 3.6, we can also write

$$\tilde{\mu}_\Lambda(t)\rceil_{\mathbb{T}^2\times\Lambda^\perp\times\mathbb{R}} = (\varphi^t_{p^V_\Lambda})_*(\tilde{\mu}^0_\Lambda\rceil_{\mathbb{T}^2\times\Lambda^\perp\times\mathbb{R}}) \quad \big(\text{resp. } \tilde{\mu}_\Lambda(t)\rceil_{\mathbb{T}^2\times\Lambda^\perp\times\mathbb{R}} = (\varphi^t_{p^0_\Lambda})_*(\tilde{\mu}^0_\Lambda\rceil_{\mathbb{T}^2\times\Lambda^\perp\times\mathbb{R}})\big),$$

when $\tau_\hbar = \epsilon_\hbar^{-1}$ (resp. $\tau_\hbar\epsilon_\hbar \to 0$). The result follows from the fact that the zero Fourier coefficient of $\mu(t)$ is itself equal to the zero Fourier coefficient of μ^0 thanks to the following adaptation of Proposition 29 from [Anantharaman and Macià 2014].

Lemma 4.2. *Suppose that*

$$\lim_{\hbar\to 0^+} \tau_\hbar \epsilon_\hbar^2 = 0.$$

Let μ be an element in $\mathcal{M}(\tau,\epsilon)$ and let μ^0 be the semiclassical measure of the sequence of initial data used to generate μ. Then, one has, for a.e. t in $\mathbb{R}$, and for every $b \in \mathcal{C}_c(\mathbb{R}^2)$,

$$\int_{\mathbb{T}^2\times\mathbb{R}^2} b(\xi)\,\mu(t,dx,d\xi) = \int_{\mathbb{T}^2\times\mathbb{R}^2} b(\xi)\,\mu^0(dx,d\xi).$$

4C. ***Propagation of wave packets.*** An application of Theorem 2.1 is the computation of semiclassical measures for wave-packet-type solutions to (6).

Let us first define wave-packet data on the torus. Take $\rho \in \mathcal{C}_c^\infty(\mathbb{R}^2)$ supported in a small neighborhood of the origin such that $\|\rho\|_{L^2(\mathbb{R}^2)} = 1$. Let $(x_0,\xi_0) \in \mathring{T}^*\mathbb{T}^2$ and set

$$U_\hbar^{x_0,\xi_0}(x) := \frac{1}{\sigma_\hbar}\rho\left(\frac{x-x_0}{\sigma_\hbar}\right)e^{i(\xi_0.x)/\hbar},$$

where $\sigma_\hbar \to 0^+$ and $\sigma_\hbar \gg \hbar$. Finally, write

$$u_\hbar^{x_0,\xi_0}(x) = \sum_{k\in\mathbb{Z}^2} U_\hbar^{x_0,\xi_0}(x+k). \tag{19}$$

If the support of ρ is small enough, then

$$\|u_\hbar^{x_0,\xi_0}\|_{L^2(\mathbb{T}^2)} = 1.$$

These initial data concentrate around x_0 and oscillate in the direction of ξ_0. Moreover, it is straightforward to check that $(u_\hbar^{x_0,\xi_0})$ satisfies (7) and (8). We next compute the time-dependent semiclassical measure of the sequence $(v_\hbar^{x_0,\xi_0})$ of solutions to (6) issued from the initial data $(u_\hbar^{x_0,\xi_0})$.

Proposition 4.3. *Suppose that the concentration scale $(\sigma_\hbar)$ satisfies $\hbar(\epsilon_\hbar\sigma_\hbar)^{-1} \to 0$ and that $\xi_0 \in \Omega_1$. Let $\mu^{x_0,\xi_0} \in \mathcal{M}(\tau,\epsilon)$ be generated by the initial data $(u_\hbar^{x_0,\xi_0})$. Let $\gamma(x,\xi_0)$ denote the geodesic in $\mathbb{T}^2$ issued from (x,ξ_0) and $\delta_{\gamma(x,\xi_0)}$ the uniform probability measure on that geodesic. The following hold*:

(1) *If $\tau_\hbar\epsilon_\hbar \to 0$, then*

$$\mu^{x_0,\xi_0}(t,dx,d\xi) = \delta_{\gamma(x_0,\xi_0)}(dx)\,\delta_{\xi_0}(d\xi).$$

(2) *If $\tau_\hbar = \epsilon_\hbar^{-1}$, then*

$$\mu^{x_0,\xi_0}(t,dx,d\xi) = \delta_{\gamma(x(t),\xi_0)}(dx)\,\delta_{\xi_0}(d\xi),$$

where $x(t)$ is the projection on $\mathbb{T}^2$ of $\varphi^t_{p^V_{\Lambda_{\xi_0}}}(x_0,\xi_0,0)$ with $\Lambda_{\xi_0} = \{\xi_0\}^\perp \cap \mathbb{Z}^2$. If x_0 is a critical point of $\mathcal{I}_{\Lambda_{\xi_0}}(V)$ then $x(t) = x_0$ for all $t \in \mathbb{R}$. In that case, μ^{x_0,ξ_0} is also constant in time.

Proof. Lemma 4.2 ensures that $\mu(t)$ is supported on $\mathbb{T}^2 \times \langle\xi_0\rangle$ for a.e. $t \in \mathbb{R}$. Therefore, by virtue of (14),

$$\mu(t) = \int_{\widehat{\mathbb{R}}} \mu_{\Lambda_{\xi_0}}(t,\cdot,d\eta)\rceil_{\mathbb{T}^2\times\langle\xi_0\rangle},$$

where $\mu_{\Lambda_{\xi_0}} \in \mathcal{M}_{\Lambda_{\xi_0}}(\tau, \epsilon)$ is generated by $(u_\hbar^{x_0,\xi_0})$. Let $\mu^0_{\Lambda_{\xi_0}}$ be an accumulation point of $(w_{\hbar,\Lambda_{\xi_0}}(0))$. Since $\hbar\sigma_\hbar^{-1} \ll \epsilon_\hbar \leq \tau_\hbar^{-1}$, one can verify that, in every regime,

$$\mu^0_{\Lambda_{\xi_0}}(dx, d\xi, d\eta) = \delta_{x_0}(dx)\,\delta_{\xi_0}(d\xi)\,\delta_0(d\eta);$$

e.g., see the proof of Proposition 5.2 in [Anantharaman et al. 2015]. The result then follows from Theorem 2.1. □

5. Proof of the two-microlocal statements

From this point on, we fix a primitive sublattice Λ of $\mathbb{Z}^2$ of rank 1 and we will proceed to the proofs of the results on two-microlocal distributions. Namely, we will first recall how to extract converging subsequences from the sequences $(w_{\Lambda,\hbar}(t\tau_\hbar))_{\hbar\to 0^+}$. Then, we will briefly recall how to adapt the proofs from [Anantharaman and Macià 2014] in order to prove Propositions 3.3 and 3.4. Finally, we will give the proofs of Theorems 3.6 and 3.7.

5A. ***Extracting subsequences.*** Recall that, following [Macià 2010; Anantharaman and Macià 2014; Anantharaman et al. 2015], we have introduced an auxiliary linear form whose invariance properties will be analyzed precisely. For every $a \in \mathcal{C}_c^\infty(T^*\mathbb{T}^2 \times \widehat{\mathbb{R}})$, we have set

$$\langle w_{\Lambda,\hbar}(t\tau_\hbar), a\rangle := \left\langle v_\hbar(t\tau_\hbar), \mathrm{Op}_\hbar^w\left(a\left(x, \xi, \frac{H_\Lambda(\xi)}{\alpha_\hbar}\right)\right) v_\hbar(t\tau_\hbar)\right\rangle,$$

where, recall, $\alpha_\hbar$ is given by (13). It will be useful to keep in mind Remark 3.1 throughout this section.

Remark 5.1. We emphasize that, for a in $\mathcal{C}_c^\infty(T^*\mathbb{T}^2)$, one has

$$\langle w_\hbar(t\tau_\hbar), a\rangle = \langle w_{\Lambda,\hbar}(t\tau_\hbar), a\rangle.$$

Our first step is to explain how to extract converging subsequences following more or less standard procedures [Gérard 1991; Macià 2009; Anantharaman and Macià 2014; Zworski 2012]. For the sake of completeness, we briefly recall it. For that purpose, we denote by

$$\mathcal{B} := \mathcal{C}_0^D(\mathbb{T}^2 \times \mathbb{R}^2 \times \widehat{\mathbb{R}})$$

the space of $\mathcal{C}^D$ functions on $\mathbb{T}^2 \times \mathbb{R}^2 \times \widehat{\mathbb{R}}$ all of whose derivatives tend to 0 at infinity. We choose $D > 0$ large enough so that Theorem B.2 holds for functions in $\mathcal{B}$.

We endow this space with its natural topology of Banach spaces. According to Theorem B.2, one knows that, for every a in $\mathcal{C}_c^\infty(\mathbb{R} \times T^*\mathbb{T}^2 \times \widehat{\mathbb{R}})$, one has

$$|\langle w_{\Lambda,\hbar}(t\tau_\hbar), a(t)\rangle| \leq C \sum_{|\alpha|\leq D} (\hbar\alpha_\hbar^{-1})^{|\alpha|/2} \|\partial^\alpha a(t)\|_\infty. \tag{20}$$

Thus, the map $t \mapsto w_{\Lambda,\hbar}(t\tau_\hbar)$ defines a bounded sequence in $L^1(\mathbb{R}, \mathcal{B})'$, and, after extracting a subsequence, one finds that there exists μ_Λ in $L^1(\mathbb{R}, \mathcal{B})'$ such that, for every a in $\mathcal{C}_c^\infty(\mathbb{R} \times T^*\mathbb{T}^2 \times \widehat{\mathbb{R}})$, one has

$$\lim_{\hbar\to 0^+} \int_{\mathbb{R}\times T^*\mathbb{T}^2\times\widehat{\mathbb{R}}} a(t,x,\xi,\eta)\, w_{\Lambda,\hbar}(t\tau_\hbar, dx, d\xi, d\eta)\, dt = \int_{\mathbb{R}\times T^*\mathbb{T}^2\times\widehat{\mathbb{R}}} a(t,x,\xi,\eta)\, \mu_\Lambda(dt, dx, d\xi, d\eta).$$

Thanks to (20) and to the fact that $\hbar\alpha_\hbar^{-1} \to 0^+$, recall that, for every θ in $\mathcal{C}_c^\infty(\mathbb{R})$ and for every a in $\mathcal{C}_c^\infty(T^*\mathbb{T}^2 \times \widehat{\mathbb{R}})$, one has

$$\left| \int_{\mathbb{R}\times T^*\mathbb{T}^2\times\widehat{\mathbb{R}}} \theta(t)\, a(x,\xi,\eta)\, \mu_\Lambda(dt,dx,d\xi,d\eta) \right| \leq C\|\theta\|_{L^1(\mathbb{R})}\|a\|_{\mathcal{C}_0^0(T^*\mathbb{T}^2\times\widehat{\mathbb{R}})}.$$

Hence, μ_Λ is absolutely continuous with respect to the t-variable; i.e., for every θ in $L^1(\mathbb{R})$ and every a in $\mathcal{C}_c^\infty(T^*\mathbb{T}^2 \times \widehat{\mathbb{R}})$, one has

$$\lim_{\hbar\to 0^+} \int_{\mathbb{R}} \theta(t)\langle w_{\Lambda,\hbar}(t\tau_\hbar), a\rangle\, dt = \int_{\mathbb{R}} \theta(t)\langle \mu_\Lambda(t), a\rangle\, dt.$$

Moreover, for a.e. t in $\mathbb{R}$, $\mu_\Lambda(t)$ is a finite Radon measure on $T^*\mathbb{T}^2 \times \widehat{\mathbb{R}}$.

5B. ***Proof of Proposition 3.3.*** We already know that the linear functionals μ_Λ are Radon measures. It remains to verify that they are positive. To see this, take $a \in \mathcal{C}_c^\infty(T^*\mathbb{T}^2 \times \widehat{\mathbb{R}})$ such that $a \geq 0$. Using the Gårding inequality (Theorem 4.32 in [Zworski 2012]), we deduce that

$$\langle w_{\Lambda,\hbar}(t\tau_\hbar), a\rangle \geq \mathcal{O}(\hbar\alpha_\hbar^{-1}) = o(1).$$

Remark 5.2. Note that the proof of the Gårding inequality in [Zworski 2012] is given in the case of $\mathbb{R}^d$. The extension to compact manifolds usually requires dealing with symbols that decay in ξ as we differentiate with respect to ξ. Yet, in the case of the torus, we can verify that this property remains true for an observable a all of whose derivatives are bounded (i.e., not necessarily decaying in ξ) as in $\mathbb{R}^d$. For that purpose, one can start from the Gårding inequality on $\mathbb{R}^d$ and apply the arguments of the proof of [Zworski 2012, Theorem 5.5], which shows L^2-boundedness of pseudodifferential of order 0 on $\mathbb{T}^d$.

After integrating against a test function θ in $L^1(\mathbb{R})$ and passing to the limit $\hbar \to 0$, one finds that, for a.e. t in $\mathbb{R}$,

$$\langle \mu_\Lambda(t), a\rangle \geq 0.$$

This concludes the proof that μ_Λ is a positive, finite Radon measure on $T^*\mathbb{T}^2 \times \widehat{\mathbb{R}}$ and one sets $\tilde{\mu}_\Lambda(t) = \mu_\Lambda(t)\rceil_{T^*\mathbb{T}^2\times\mathbb{R}}$ and $\tilde{\mu}^\Lambda(t) = \mu_\Lambda(t)\rceil_{T^*\mathbb{T}^2\times\{\pm\infty\}}$. Thanks to the frequency assumption (8), one has, for a.e. t in $\mathbb{R}$,

$$\mu_\Lambda(t)(\{\xi = 0\}) = 0. \tag{21}$$

Remark 5.3. Remark 5.1 implies that, for a.e. t in $\mathbb{R}$, the time-dependent semiclassical measure $\mu(t)$ can be obtained by

$$\mu(t) = \int_{\widehat{\mathbb{R}}} \mu_\Lambda(t,\cdot,d\eta). \tag{22}$$

5C. ***Proof of Proposition 3.4.*** Concerning the support of $\tilde{\mu}_\Lambda(t)$, we let a be an element in $\mathcal{C}_c^\infty(T^*\mathbb{T}^2 \times \mathbb{R})$ whose support does not intersect $\mathbb{T}^2 \times \Lambda^\perp \times \mathbb{R}$. Using Remark 3.1, one has

$$\mathrm{Op}_\hbar^w\left(a\left(x,\xi,\frac{H_\Lambda(\xi)}{\alpha_\hbar}\right)\right) = \mathrm{Op}^w_{\hbar\alpha_\hbar^{-1}}(a(x,\alpha_\hbar\xi,H_\Lambda(\xi))).$$

Hence, this operator is equal to 0 when $\hbar$ is small enough (thanks to our assumption on the support of a). This concludes the proof of the first part of Proposition 3.4.

Let us now discuss invariance by the geodesic flow, which is the only property that uses the particular form of $v_\hbar(t\tau_\hbar)$ so far. Again, we start with the "compact" part and we fix a to be an element in $\mathcal{C}_c^\infty(T^*\mathbb{T}^2 \times \mathbb{R})$. Using composition rules for pseudodifferential operators, we write

$$\frac{d}{dt}\langle w_{\Lambda,\hbar}(t\tau_\hbar), a\rangle = \tau_\hbar \langle w_{\Lambda,\hbar}(t\tau_\hbar), \xi.\partial_x a\rangle + \frac{i\tau_\hbar \epsilon_\hbar^2}{\hbar}\big\langle v_\hbar(t\tau_\hbar), \big[V, \mathrm{Op}^w_{\hbar\alpha_\hbar^{-1}}\big(a(x, \alpha_\hbar\xi, H_\Lambda(\xi))\big)\big] v_\hbar(t\tau_\hbar)\big\rangle.$$

Using Theorem B.3 (more specifically Remark B.4) one more time, we have

$$\big[V, \mathrm{Op}^w_{\hbar\alpha_\hbar^{-1}}\big(a(x, \alpha_\hbar\xi, H_\Lambda(\xi))\big)\big] = -\frac{\hbar}{i\alpha_\hbar}\,\mathrm{Op}^w_\hbar\left(\frac{\mathfrak{e}_\Lambda}{L_\Lambda}.\partial_x V\, \partial_\eta a\left(x, \xi, \frac{H_\Lambda(\xi)}{\alpha_\hbar}\right)\right) + \mathcal{O}(\hbar^3(\alpha_\hbar)^{-3}).$$

Combining these two identities with the facts $\hbar\alpha_\hbar^{-1} = o(1)$ and $\epsilon_\hbar\alpha_\hbar^{-1} = \mathcal{O}(1)$, we find that

$$\frac{d}{dt}\langle w_{\Lambda,\hbar}(t\tau_\hbar), a\rangle = \tau_\hbar\left(\left\langle w_{\Lambda,\hbar}(t\tau_\hbar), \xi.\partial_x a - \frac{\epsilon_\hbar^2}{\alpha_\hbar}\frac{\mathfrak{e}_\Lambda}{L_\Lambda}.\partial_x V\, \partial_\eta a\right\rangle + o(\hbar)\right).$$

Let now θ be an element in $\mathcal{C}_c^1(\mathbb{R})$. Integrating the previous equality against θ and integrating by parts, we find

$$\int_{\mathbb{R}} \theta(t)\left\langle w_{\Lambda,\hbar}(t\tau_\hbar), \xi.\partial_x a - \frac{\epsilon_\hbar^2}{\alpha_\hbar}\frac{\mathfrak{e}_\Lambda}{L_\Lambda}.\partial_x V\, \partial_\eta a\right\rangle dt = \mathcal{O}(\tau_\hbar^{-1}) + o(\hbar),$$

which implies the result for every a in $\mathcal{C}_c^\infty(T^*\mathbb{T}^2 \times \mathbb{R})$ when we let $\hbar$ go to 0. Note that we used the Calderón–Vaillancourt theorem (Theorem B.2) to bound the $\epsilon_\hbar^2\alpha_\hbar^{-1}$ term on the left-hand side of this equality.

It now remains to treat the part at infinity. Let a be an element in $\mathcal{C}_c^\infty(T^*\mathbb{T}^2 \times \widehat{\mathbb{R}})$. For every $R \geq 1$ and for every smooth cutoff function near 0, we set

$$a^R(x, \xi, \eta) := a(x, \xi, \eta)\left(1 - \chi\left(\frac{\eta}{R}\right)\right).$$

The same argument as before allows us to prove that, for every θ in $\mathcal{C}^1(\mathbb{R})$, one has

$$\int_{\mathbb{R}} \theta(t)\left\langle w_{\Lambda,\hbar}(t\tau_\hbar), (\xi.\partial_x a)^R - \frac{\epsilon_\hbar^2}{\alpha_\hbar}\frac{\mathfrak{e}_\Lambda}{L_\Lambda}.\partial_x V\, \partial_\eta a^R\right\rangle dt = o(1).$$

Thus, we can take the limit $\hbar \to 0$ and conclude the proof by letting R go to $+\infty$.

5D. ***Invariance and propagation of two-microlocal distributions.*** We now turn to the proofs of our main statements, namely Theorems 3.6 and 3.7. Analogously to [Anantharaman and Macià 2014], we define the differential operators

$$D_\Lambda := \frac{1}{i}\frac{\mathfrak{e}_\Lambda}{L_\Lambda}.\nabla \quad \text{and} \quad D_\Lambda^\perp := \frac{1}{i}\frac{\mathfrak{e}_\Lambda^\perp}{L_\Lambda}.\nabla$$

associated with the Hamiltonians H_Λ and $H_\Lambda^\perp$. One has

$$-\Delta = (D_\Lambda^\perp)^2 + D_\Lambda^2. \tag{23}$$

Recall also that, for every smooth compactly supported function b on $T^*\mathbb{T}^2$, the Egorov theorem is exact for these operators and it tells us that

$$\mathrm{Op}_\hbar^w(\mathcal{I}_\Lambda(b)) = \frac{1}{L_\Lambda}\int_0^{L_\Lambda} e^{isD_\Lambda^\perp}\,\mathrm{Op}_\hbar^w(b)e^{-isD_\Lambda^\perp}\,ds. \tag{24}$$

and that

$$[D_\Lambda^\perp, \mathrm{Op}_\hbar^w(\mathcal{I}_\Lambda(b))] = 0. \tag{25}$$

As mentioned before, this construction (which was originally presented in [Anantharaman and Macià 2014]) is reminiscent of the averaging argument of [Weinstein 1977] applied to certain one-dimensional tori that depend on Λ.

5D1. *Proof of Theorem 3.6.* Let a be an element in $\mathcal{C}_c^\infty(T^*\mathbb{T}^2\times\mathbb{R})$. We start our proof by computing the derivative of the two-microlocal Wigner distribution. One has

$$\frac{d}{dt}\langle w_{\Lambda,\hbar}(t\tau_\hbar), \mathcal{I}_\Lambda(a)\rangle = \frac{i\tau_\hbar}{\hbar}\big\langle v_\hbar(t\tau_\hbar), \big[\tfrac{1}{2}\hbar^2(D_\Lambda^\perp)^2 + \tfrac{1}{2}\hbar^2 D_\Lambda^2 + \epsilon_\hbar^2 V, \mathrm{Op}_\hbar^w(a_{\Lambda,\hbar})\big]v_\hbar(t\tau_\hbar)\big\rangle,$$

where

$$a_{\Lambda,\hbar}(x,\xi) := \mathcal{I}_\Lambda(a)\left(x,\xi,\frac{H_\Lambda(\xi)}{\alpha_\hbar}\right).$$

Using (25), we deduce that

$$\frac{d}{dt}\langle w_{\Lambda,\hbar}(t\tau_\hbar), \mathcal{I}_\Lambda(a)\rangle = \frac{i\tau_\hbar}{\hbar}\big\langle v_\hbar(t\tau_\hbar), \big[\tfrac{1}{2}\hbar^2 D_\Lambda^2 + \epsilon_\hbar^2 V, \mathrm{Op}_\hbar^w(a_{\Lambda,\hbar})\big]v_\hbar(t\tau_\hbar)\big\rangle.$$

Thanks to the commutation properties of the Weyl quantization from Remark B.4, one has

$$\begin{aligned}\frac{d}{dt}&\langle w_{\Lambda,\hbar}(t\tau_\hbar), \mathcal{I}_\Lambda(a)\rangle \\ &= \mathcal{O}(\tau_\hbar\epsilon_\hbar^2\hbar^2(\alpha_\hbar)^{-3}) \\ &\quad + \alpha_\hbar\tau_\hbar\left\langle v_\hbar(t\tau_\hbar), \mathrm{Op}_\hbar^w\left(\frac{H_\Lambda(\xi)}{\alpha_\hbar}\frac{\mathfrak{e}_\Lambda.\partial_x\mathcal{I}_\Lambda(a)(x,\xi,H_\Lambda(\xi)/\alpha_\hbar)}{L_\Lambda} - \frac{\epsilon_\hbar^2}{\alpha_\hbar^2}\partial_\eta\mathcal{I}_\Lambda(a)\frac{\mathfrak{e}_\Lambda.\partial_x V}{L_\Lambda}\right)v_\hbar(t\tau_\hbar)\right\rangle.\end{aligned} \tag{26}$$

Our assumption $\hbar \ll \epsilon_\hbar \ll \alpha_\hbar$ ensures that the remainder is in fact of order $o(\hbar\tau_\hbar)$.

We now distinguish three regimes.

First, we suppose that $\epsilon_\hbar\tau_\hbar \to 0$ as $\hbar \to 0^+$. In particular, $\alpha_\hbar = \tau_\hbar^{-1} \gg \epsilon_\hbar$. Thanks to the Calderón–Vaillancourt theorem (Theorem B.2), we can verify that the last term in the right-hand side of equality (26) is in fact $o(1)$ uniformly for t in $\mathbb{R}$. Letting $\hbar \to 0$, one finds that, for a.e. t in $\mathbb{R}$,

$$\frac{d}{dt}\langle \mu_\Lambda(t), \mathcal{I}_\Lambda(a)\rangle = \left\langle \mu_\Lambda(t), \eta\frac{\mathfrak{e}_\Lambda}{L_\Lambda}.\partial_x\mathcal{I}_\Lambda(a)\right\rangle.$$

Combining Proposition 3.4 with (21), one has then $\langle \mu_\Lambda(t), a\rangle = \langle \mu_\Lambda^0, \mathcal{I}_\Lambda(a)\circ\varphi_{p_\Lambda^0}^t\rangle$ for a.e. t in $\mathbb{R}$, which proves point (1) of the theorem.

Suppose now that $\tau_\hbar \epsilon_\hbar \to c > 0$. Letting $\hbar \to 0$, the limit measure satisfies the following transport equation for all $\theta \in \mathcal{C}^1_c(\mathbb{R})$:

$$-\int_{\mathbb{R}} \theta'(t)\langle \mu_\Lambda(t), \mathcal{I}_\Lambda(a)\rangle\, dt = c\int_{\mathbb{R}} \theta(t)\left\langle \mu_\Lambda(t), \eta\frac{\mathfrak{e}_\Lambda.\partial_x \mathcal{I}_\Lambda(a)}{L_\Lambda} - \partial_\eta \mathcal{I}_\Lambda(a)\frac{\mathfrak{e}_\Lambda.\partial_x V}{L_\Lambda}\right\rangle dt.$$

Using again Proposition 3.4 with (21), one deduces that

$$\partial_t \langle \mu_\Lambda(t), \mathcal{I}_\Lambda(a)\rangle = c\left\langle \mu_\Lambda(t), \eta\frac{\mathfrak{e}_\Lambda.\partial_x \mathcal{I}_\Lambda(a)}{L_\Lambda} - \partial_\eta \mathcal{I}_\Lambda(a)\frac{\mathfrak{e}_\Lambda.\partial_x \mathcal{I}_\Lambda(V)}{L_\Lambda}\right\rangle.$$

This proves point (2) of the theorem.

Finally, we suppose that $\tau_\hbar \epsilon_\hbar \to +\infty$. Let θ be an element in $\mathcal{C}^1_c(\mathbb{R})$. We integrate one more time equality (26) against θ, and we make an integration by parts on the left-hand side of the equality. Then, we make use of the Calderón–Vaillancourt theorem (Theorem B.2) to bound the left-hand side. After letting $\hbar$ go to 0, one finds that, for every θ in $\mathcal{C}^1_c(\mathbb{R})$,

$$\int_{\mathbb{R}} \theta(t)\left\langle \mu_\Lambda(t), \eta\frac{\mathfrak{e}_\Lambda.\partial_x \mathcal{I}_\Lambda(a)}{L_\Lambda} - \partial_\eta \mathcal{I}_\Lambda(a)\frac{\mathfrak{e}_\Lambda.\partial_x \mathcal{I}_\Lambda(V)}{L_\Lambda}\right\rangle dt = 0,$$

where we used one more time Proposition 3.4 with (21) in order to replace V by its Λ-average $\mathcal{I}_\Lambda(V)$. This implies point (3) of the theorem.

5D2. *Proof of Theorem 3.7.* Let now a be an element in $\mathcal{C}^\infty_c(\mathbb{R}^2 \times \widehat{\mathbb{R}})$ and let k be an element in $\Lambda - \{0\}$. It suffices to show that

$$\langle \tilde{\mu}^\Lambda(t), e^{-2i\pi k.x} a(\xi, \eta)\rangle = 0.$$

We fix $\chi_1(\eta) \in \mathcal{C}^\infty(\mathbb{R}, [0, 1])$ which is equal to 1 for $\eta \geq 1$ and to 0 for $\eta \leq \frac{1}{2}$. For every $R \geq 1$, we set

$$a^{R,k}_\pm(x, \xi, \eta) := e^{-2i\pi k.x} a(\xi, \eta)\chi_1\left(\pm\frac{\eta}{R}\right).$$

Remark 5.4. Let θ be an element in $\mathcal{C}^1_c(\mathbb{R})$. One has

$$\int_{\mathbb{R}} \theta(t)\,\frac{d}{dt}\left\langle w_{\Lambda,\hbar}(t\tau_\hbar), \frac{1}{\eta}a^{R,k}_\pm\right\rangle dt = -\int_{\mathbb{R}} \theta'(t)\left\langle w_{\Lambda,\hbar}(t\tau_\hbar), \frac{1}{\eta}a^{R,k}_\pm\right\rangle dt.$$

Thanks to the Calderón–Vaillancourt theorem (Theorem B.2), one knows that

$$\left\| \mathrm{Op}^w_\hbar\left(\chi\left(\frac{H_\Lambda(\xi)}{R\alpha_\hbar}\right) a\left(\xi, \frac{H_\Lambda(\xi)}{\alpha_\hbar}\right) e^{-2i\pi k.x}\frac{\alpha_\hbar}{H_\Lambda(\xi)}\right)\right\|_{L^2\to L^2} = \mathcal{O}(R^{-1}).$$

Thus, one has

$$\int_{\mathbb{R}} \theta(t)\frac{d}{dt}\left\langle w_{\Lambda,\hbar}(t\tau_\hbar), \frac{1}{\eta}a^{R,k}_\pm\right\rangle dt = \mathcal{O}(R^{-1}).$$

In order to prove the proposition, we will now compute explicitly the derivative of $\left\langle w_{\Lambda,\hbar}(t\tau_\hbar), \frac{1}{\eta}a^{R,k}_\pm\right\rangle$. For that purpose, we need to compute the following bracket:

$$\left[-\frac{\hbar^2\Delta}{2} + \epsilon^2_\hbar V, \mathrm{Op}^w_\hbar\left(a^{R,k}_\pm\left(x, \xi, \frac{H_\Lambda(\xi)}{\alpha_\hbar}\right)\frac{\alpha_\hbar}{H_\Lambda(\xi)}\right)\right].$$

Using again (25), this commutator is in fact equal to

$$\left[\frac{\hbar^2 D_\Lambda^2}{2}+\epsilon_\hbar^2 V, \mathrm{Op}_\hbar^w\left(a_\pm^{R,k}\left(x,\xi,\frac{H_\Lambda(\xi)}{\alpha_\hbar}\right)\frac{\alpha_\hbar}{H_\Lambda(\xi)}\right)\right].$$

We split this commutator in two parts. Thanks to Remark B.4, one has

$$\left[\frac{\hbar^2 D_\Lambda^2}{2}, \mathrm{Op}_\hbar^w\left(a_\pm^{R,k}\left(x,\xi,\frac{H_\Lambda(\xi)}{\alpha_\hbar}\right)\frac{\alpha_\hbar}{H_\Lambda(\xi)}\right)\right] = -2\pi\hbar\alpha_\hbar\,\mathrm{Op}_\hbar^w\left(\frac{\mathfrak{e}_\Lambda}{L_\Lambda}.k a_\pm^{R,k}\left(x,\xi,\frac{H_\Lambda(\xi)}{\alpha_\hbar}\right)\right).$$

For the other part of the commutator, we use one more time the commutation rule for pseudodifferential operators and the Calderón–Vaillancourt theorem (Theorem B.2). We find that

$$\left[V, \mathrm{Op}_\hbar^w\left(a_\pm^{R,k}\left(x,\xi,\frac{H_\Lambda(\xi)}{\alpha_\hbar}\right)\frac{\alpha_\hbar}{H_\Lambda(\xi)}\right)\right] = \mathcal{O}_{L^2\to L^2}(\hbar\alpha_\hbar^{-1}R^{-1}+\hbar^3\alpha_\hbar^{-3}).$$

As $\hbar\epsilon_\hbar^{-1}\to 0$ and $\epsilon_\hbar=\mathcal{O}(\alpha_\hbar)$, we finally get that

$$\frac{d}{dt}\left\langle w_{\Lambda,\hbar}(t\tau_\hbar), \frac{1}{\eta}a_\pm^{R,k}\right\rangle = -\frac{2\pi\tau_\hbar\alpha_\hbar\mathfrak{e}_\Lambda.k}{L_\Lambda}\langle w_{\Lambda,\hbar}(t\tau_\hbar), a_\pm^{R,k}\rangle + \mathcal{O}(\tau_\hbar\epsilon_\hbar R^{-1}) + o(\tau_\hbar\hbar).$$

Let now θ be an element in $\mathcal{C}_c^1(\mathbb{R})$. We integrate these expressions against θ. Using Remark 5.4 and making the assumption that $\limsup_{\hbar\to 0^+}\tau_\hbar\alpha_\hbar>0$, we obtain

$$\text{for all } k\in\Lambda-\{0\},\qquad \int_{\mathbb{R}}\theta(t)\langle w_{\Lambda,\hbar}(t\tau_\hbar), a_\pm^{R,k}\rangle\,dt = o(1)+\mathcal{O}(R^{-1}).$$

We now let $\hbar$ go to 0, and we get that, for every $R>0$,

$$\text{for all } k\in\Lambda-\{0\},\qquad \int_{\mathbb{R}}\theta(t)\langle \mu_\Lambda(t), a_\pm^{R,k}\rangle\,dt = \mathcal{O}(R^{-1}).$$

To get the conclusion, we let R go to $+\infty$.

Remark 5.5. From this theorem, we deduce that, for every $a(x,\xi,\eta)$ in $\mathcal{C}_c^\infty(T^*\mathbb{T}^2\times\widehat{\mathbb{R}})$ and for a.e. t in $\mathbb{R}$,

$$\tilde{\mu}^\Lambda(t)(\mathcal{I}_\Lambda(a)) = \int_{T^*\mathbb{T}^2\times\{\pm\infty\}}\hat{a}_0(\xi,\eta)\,\mu_\Lambda(t,d\xi,d\eta).$$

Appendix A. Regularity of bi-invariant measures

In this appendix, we fix Λ a primitive sublattice of $\mathbb{Z}^2$ of rank 1, and we aim at analyzing the regularity of the set of finite measures on $T^*\mathbb{T}^2$ which are invariant by the Hamiltonian flows[5] $\varphi_{H_\Lambda^\perp}^t$ and $\varphi_{p_\Lambda^V}^t$. We will now recall the results from Section 4 of [Macià and Rivière 2016] and explain how they can be adapted to the present framework. We refer the reader to this reference for the detailed proofs. We introduce the critical set in the direction of Λ,

$$\mathrm{Crit}_\Lambda(V) := \{(x,\xi)\in T^*\mathbb{T}^2 : H_\Lambda(\xi)=0 \text{ and } \partial_x\mathcal{I}_\Lambda(V)=0\}.$$

[5]By making a slight abuse of notation, we shall identify $\varphi_{p_\Lambda}^t$, a flow a priori defined on $\mathbb{T}^2\times\Lambda^\perp\times\mathbb{R}$, to a flow on $T^*\mathbb{T}^2$ via the diffeomorphism (18). Recall that $\varphi_{H_\Lambda^\perp}^t$ and $\varphi_{p_\Lambda^V}^t$ commute.

This is a closed subset of $T^*\mathbb{T}^2$ which is invariant by the Hamiltonian flows $\varphi^t_{H_\Lambda^\perp}$ and $\varphi^t_{p_\Lambda^V}$, and we introduce its complement

$$\mathcal{R}(\Lambda) := T^*\mathbb{T}^2 - \mathrm{Crit}_\Lambda(V).$$

The map

$$\phi : \mathbb{R}^2 \times \mathcal{R}(\Lambda) \ni (s,t,x,\xi) \mapsto \varphi^s_{H_\Lambda^\perp} \circ \varphi^t_{p_\Lambda^V}(x,\xi) \in \mathcal{R}(\Lambda)$$

is a group action of $\mathbb{R}^2$ on $\mathcal{R}(\Lambda)$. Moreover, for any $(x_0,\xi_0) \in \mathcal{R}(\Lambda)$, the map

$$\phi_{x_0,\xi_0} : \mathbb{R}^2 \ni (s,t) \mapsto \varphi^s_{H_\Lambda^\perp} \circ \varphi^t_{p_\Lambda^V}(x_0,\xi_0) \in \mathcal{R}(\Lambda)$$

is an immersion. Therefore, the stabilizer group G_{x_0,ξ_0} of (x_0,ξ_0) under ϕ is discrete. This proves that the orbits of the action ϕ are either diffeomorphic to the torus $\mathbb{T}^2$, to the cylinder $\mathbb{T}\times\mathbb{R}$ or to $\mathbb{R}^2$. On the other hand, the moment map,

$$\Phi : \mathcal{R}(\Lambda) \ni (x,\xi) \mapsto (H_\Lambda^\perp(\xi),\, p_\Lambda^V(x,\xi)) \in \mathbb{R}^2,$$

is a submersion, and, for every $(H,J) \in \Phi(\mathcal{R}(\Lambda))$, the level set

$$\mathcal{L}_{(H,J)} := \Phi^{-1}(H,J)$$

is a smooth submanifold of $\mathcal{R}(\Lambda)$ of dimension 2. To summarize, the pair $(H_\Lambda^\perp, p_\Lambda^V)$ forms a completely integrable system on $\mathcal{R}(\Lambda)$, and the map ϕ_{x_0,ξ_0} induces a diffeomorphism:

$$\text{for all } (x_0,\xi_0) \in \mathcal{R}(\Lambda), \quad \phi_{x_0,\xi_0} : \mathbb{R}^2/G_{x_0,\xi_0} \to \mathcal{L}^{x_0,\xi_0}_{(H_0,J_0)} \quad \text{for } (H_0,J_0) := \Phi(x_0,\xi_0).$$

Here, $\mathcal{L}^{x_0,\xi_0}_{(H_0,J_0)}$ denotes the connected component of $\mathcal{L}_{(H_0,J_0)}$ that contains (x_0,ξ_0). Therefore, if $\mathcal{L}^{x_0,\xi_0}_{(H_0,J_0)}$ is compact then it is an embedded Lagrangian torus in $T^*\mathbb{T}^2$. In that case, we shall write

$$\mathbb{T}^2_{x_0,\xi_0} := \mathbb{R}^2/G_{x_0,\xi_0}.$$

In the following, we denote by $\mathcal{R}_c(\Lambda)$ the set formed by those $(x,\xi) \in \mathcal{R}(\Lambda)$ such that $\mathcal{L}^{x,\xi}_{\Phi(x,\xi)}$ is compact. Mimicking the proof of Proposition 4.2 in [Macià and Rivière 2016], one can show that the following holds:

Proposition A.1. *Let μ be a probability measure on $\mathcal{R}(\Lambda)$ that is invariant by $\varphi^t_{H_\Lambda^\perp}$ and $\varphi^t_{p_\Lambda^V}$. Set $\bar\mu := \Phi_*\mu$. Then, for every $a \in \mathcal{C}_c(\mathcal{R}(\Lambda))$, one has*

$$\int_{\mathcal{R}(\Lambda)} a(x,\xi)\,\mu(dx,d\xi) = \int_{\Phi(\mathcal{R}(\Lambda))} \int_{\mathcal{L}_{(H,J)}} a(x,\xi)\,\lambda_{H,J}(dx,d\xi)\,\bar\mu(dH,dJ),$$

where, for $(H,J) \in \Phi(\mathcal{R}(\Lambda))$, the measure $\lambda_{H,J}$ is a convex combination of the (normalized) Haar measures on the tori $\mathcal{L}^{x_0,\xi_0}_{(H,J)}$ for $(x_0,\xi_0) \in \mathcal{L}_{(H,J)} \cap \mathcal{R}_c(\Lambda)$. In particular, for every (x,ξ) in $\mathcal{R}(\Lambda)$, one has

$$\mu\big(\{\varphi^s_{H_\Lambda^\perp}(x,\xi) : 0 \le s \le L_\Lambda\}\big) = 0.$$

An explicit formula for the restriction of the measure $\lambda_{H,J}$ to a connected component $\mathcal{L}^{x,\xi}_{(H,J)}$ with $(x,\xi) \in \mathcal{R}_c(\Lambda) \cap \mathcal{L}_{(H,J)}$ is the following:

$$\int_{\mathcal{L}^{x_0,\xi_0}_{(H,J)}} a(x,\xi)\, \lambda_{H,J}(dx, d\xi) = c \int_{\mathbb{T}^2_{x_0,\xi_0}} a(\phi_{x_0,\xi_0}(s,t))\, ds\, dt \tag{27}$$

for some constant $c \in [0,1]$.

We will now discuss the regularity of the projections of bi-invariant measures following the proof from Section 4.2 in [Macià and Rivière 2016]. We denote by $\Pi : T^*\mathbb{T}^2 \to \mathbb{T}^2$ the canonical projection. The main result from Section 4 in [Macià and Rivière 2016] is the following:

Theorem A.2. *Let μ be a probability measure on $\mathcal{R}(\Lambda)$ that is invariant by $\varphi^t_{H^\perp_\Lambda}$ and $\varphi^t_{p^V_\Lambda}$. Then, $\nu := \Pi_*\mu$ is a probability measure on $\mathbb{T}^2$ that is absolutely continuous with respect to the Lebesgue measure.*

Denote by $\mathcal{N}(\Lambda)$ the convex closure of the set of measures $\delta_{\Pi\circ\Gamma}$, where $\Gamma \subset T^*\mathbb{T}^2$ ranges over the orbits of $\varphi_{H^\perp_\Lambda}$ that are contained in $\mathrm{Crit}_\Lambda(V)$. A direct consequence of the previous theorem is the following:

Corollary A.3. *The projection $\nu := \Pi_*\mu$ of a probability measure μ on $T^*\mathbb{T}^2$ that is invariant by $\varphi^t_{H^\perp_\Lambda}$ and $\varphi^t_{p^V_\Lambda}$ can be decomposed as*

$$\nu = f\ \mathrm{vol} + \alpha \nu_{\mathrm{sing}},$$

where $f \in L^1(\mathbb{T}^2)$, $\alpha \in [0,1]$ and $\nu_{\mathrm{sing}} \in \mathcal{N}(\Lambda)$.

Note that, for a "generic" choice of V, the set of points x satisfying $\partial_x \mathcal{I}_\Lambda(V) = 0$ consists of finitely many closed geodesics of $\mathbb{T}^2$. In particular, ν_{sing} is a finite combination of measures carried by closed geodesics.

Proof. As it is simple to explain in the current framework, we briefly explain how the proof of Theorem 4.6 in [Macià and Rivière 2016] can be adapted to prove Theorem A.2 — see also Lemma 2.1 in [Bialy and Polterovich 1989]. Recall that it is sufficient to fix some (x_0,ξ_0) in $\mathcal{R}_c(\Lambda)$ and to prove that the set of points where

$$\phi_{x_0,\xi_0} : (s,t) \in \mathbb{T}^2_{x_0,\xi} \mapsto \Pi \circ \varphi^s_{H^\perp_\Lambda} \circ \varphi^t_{p^V_\Lambda}(x_0,\xi_0) \in \mathbb{T}^2$$

is not a local diffeomorphism is made of finitely many disjoint $\mathcal{C}^1$ closed curves. Such curves are called caustics. This can be proved as follows. One can verify that the points where we do not have a local diffeomorphism are defined by the points (s,t) satisfying

$$H_\Lambda(\phi_{x_0,\xi_0}(s,t)) = 0.$$

Note that, for every s in $\mathbb{R}$,

$$H_\Lambda(\varphi^t_{p^V_\Lambda}(x_0,\xi_0)) = H_\Lambda(\phi_{x_0,\xi_0}(s,t)).$$

As (x_0,ξ_0) belongs to the $\varphi^t_{p^V_\Lambda}$-invariant set $\mathcal{R}(\Lambda)$, we know that

$$\partial_x \mathcal{I}_\Lambda(V)(\varphi^t_{p^V_\Lambda}(x_0,\xi_0)) \neq 0.$$

Thus, from the Hamilton–Jacobi equations, we deduce that there exists a small open neighborhood $(t-\eta, t+\eta)$ of t such that, for every $t' \in (t-\eta, t+\eta) - \{t\}$,

$$H_\Lambda \circ \varphi^{t'}_{p^V_\Lambda}(x_0, \xi_0) \neq 0.$$

In particular, there are only finitely many values of t such that $H_\Lambda \circ \varphi^t_{p^V_\Lambda}(x_0, \xi_0) \neq 0$ and thus, there are only finitely many closed curves on $\mathbb{T}^2_{x_0,\xi_0}$ where the map ϕ_{x_0,ξ_0} is not a local diffeomorphism. □

Appendix B. Background on semiclassical analysis

In this appendix, we give a brief reminder of semiclassical analysis and we refer to [Zworski 2012] (mainly Chapters 1 to 5) for a more detailed exposition. Given $\hbar > 0$ and a in $\mathcal{S}(\mathbb{R}^{2d})$ (the Schwartz class), one can define the Weyl quantization of a as follows:

$$\text{for all } u \in \mathcal{S}(\mathbb{R}^d), \quad \mathrm{Op}^w_\hbar(a)u(x) := \frac{1}{(2\pi\hbar)^d} \iint_{\mathbb{R}^{2d}} e^{(i/\hbar)\langle x-y,\xi\rangle} a\big(\tfrac{1}{2}(x+y), \xi\big)\, u(y)\, dy\, d\xi.$$

This definition can be extended to any observable a with uniformly bounded derivatives, i.e., such that for every $\alpha \in \mathbb{N}^{2d}$, there exists $C_\alpha > 0$ such that $\sup_{x,\xi} |\partial^\alpha a(x,\xi)| \leq C_\alpha$. More generally, we will use the convention, for every $m \in \mathbb{R}$ and every $k \in \mathbb{Z}$,

$$S^{m,k} := \big\{(a_\hbar(x,\xi))_{0<\hbar\leq 1} : \text{for all } (\alpha,\beta) \in \mathbb{N}^d \times \mathbb{N}^d, \sup_{(x,\xi)\in\mathbb{R}^{2d}; 0<\hbar\leq 1} |\hbar^k \langle\xi\rangle^{-m}\, \partial_x^\alpha \partial_\xi^\beta a_\hbar(x,\xi)| < +\infty\big\},$$

where $\langle\xi\rangle := (1+\|\xi\|^2)^{1/2}$. For such symbols, $\mathrm{Op}^w_\hbar(a)$ defines a continuous operator $\mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$ which acts by duality on $\mathcal{S}'(\mathbb{R}^d)$.

Remark B.1. We also have the following relation, which we use at different stages of our proof:

$$\text{for all } \delta > 0, \text{ for all } a \in S^{m,k}, \quad \mathrm{Op}^w_\hbar(a(x,\xi)) = \mathrm{Op}^w_{\hbar\delta^{-1}}(a(x,\delta\xi)). \tag{28}$$

Among the above symbols, we distinguish the family of $\mathbb{Z}^d$-periodic symbols, which we denote by $S^{m,k}_{\mathrm{per}}$. Note that any a in $\mathcal{C}^\infty(T^*\mathbb{T}^d)$ (with bounded derivatives) defines an element in $S^{0,0}_{\mathrm{per}}$. Similarly to the proof of Theorem 4.19 in [Zworski 2012], one can verify that, for any $a \in S^{m,k}_{\mathrm{per}}$,

$$\mathrm{Op}^w_\hbar(a)(e_k) = \sum_{q\in\mathbb{Z}^d} e_q \hat{a}_{q-k}(\pi\hbar(q+k)),$$

where $e_k(x) := e^{2i\pi k.x}$, and $\hat{a}_p(\xi) := \int_{\mathbb{T}^d} a(x,\xi) e^{-2i\pi p.x}\, dx$. In particular, for any $a \in S^{m,k}_{\mathrm{per}}$, the operator $\mathrm{Op}^w_\hbar(a)$ maps trigonometric polynomials into a smooth $\mathbb{Z}^d$-periodic function, and more generally any smooth $\mathbb{Z}^d$-periodic function into a smooth $\mathbb{Z}^d$-periodic function. Thus, for every a in $S^{m,k}_{\mathrm{per}}$, the operator $\mathrm{Op}^w_\hbar(a)$ acts by duality on the space of distributions $\mathcal{D}'(\mathbb{T}^d)$. An important feature of this quantization procedure is that it defines a bounded operator on $L^2(\mathbb{T}^d)$ [Zworski 2012, Chapter 5]:

Theorem B.2 (Calderón–Vaillancourt). *There exists a constant $C_d > 0$ and an integer $D > 0$ such that, for every a in $S^{0,0}_{\mathrm{per}}$, one has, for every $0 < \hbar \leq 1$,*

$$\|\mathrm{Op}^w_\hbar(a)\|_{L^2(\mathbb{T}^d)\to L^2(\mathbb{T}^d)} \leq C_d \sum_{|\alpha|\leq D} \hbar^{|\alpha|/2} \|\partial^\alpha a\|_\infty.$$

Another important feature of the Weyl quantization procedure is the composition formula:

Theorem B.3 (composition formula). *Let $a \in S^{m_1,k_1}$ and $b \in S^{m_2,k_2}$. Then, one has, for any $0 < \hbar \leq 1$,*

$$\mathrm{Op}_\hbar^w(a) \circ \mathrm{Op}_\hbar^w(b) = \mathrm{Op}_\hbar^w(a \sharp_\hbar b)$$

in the sense of operators from $\mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$, where $a \sharp_\hbar b$ has uniformly bounded derivatives, and, for every $N \geq 0$,

$$a \sharp_\hbar b \sim \sum_{k=0}^{N} \frac{1}{k!} \big(\tfrac{1}{2} i\hbar D\big)^k (a, b) + \mathcal{O}(\hbar^{N+1}),$$

where $D(a, b)(x, \xi) = (\partial_x \partial_v - \partial_y \partial_\xi)(a(x, \xi)b(y, v))\rceil_{y=x, v=\xi}$.

We refer to Chapter 4 of [Zworski 2012] for a detailed proof of this result. We observe that for $N = 0$, the coefficient is given by the symbol ab, and for $N = 1$, it is given by $(\hbar/(2i))\{a, b\}$, where $\{\cdot, \cdot\}$ is the Poisson bracket. As before, we can restrict this result to the case of periodic symbols, and we can check that the composition formula remains valid for operators acting on $\mathcal{C}^\infty(\mathbb{T}^d)$.

Remark B.4. We note that the formula for the composed symbols is quite symmetric, and we have in fact the following useful property; for every $N \geq 0$,

$$a \sharp_\hbar b - b \sharp_\hbar a \sim \sum_{k=0}^{N} \frac{2}{(2k+1)!} \big(\tfrac{1}{2} i\hbar D\big)^{2k+1} (a, b) + \mathcal{O}(\hbar^{2N+3}).$$

Finally, note that, if $b(\xi)$ is a polynomial in ξ of order ≤ 2, one has the exact formula

$$a \sharp_\hbar b - b \sharp_\hbar a = \frac{\hbar}{2i}\{a, b\}.$$

Acknowledgements

We warmly thank the referee for his careful reading and his useful suggestions regarding the results presented in this article.

References

[Anantharaman and Léautaud 2014] N. Anantharaman and M. Léautaud, "Sharp polynomial decay rates for the damped wave equation on the torus", *Anal. PDE* **7**:1 (2014), 159–214. MR Zbl

[Anantharaman and Macià 2014] N. Anantharaman and F. Macià, "Semiclassical measures for the Schrödinger equation on the torus", *J. Eur. Math. Soc.* (*JEMS*) **16**:6 (2014), 1253–1288. MR Zbl

[Anantharaman et al. 2015] N. Anantharaman, C. Fermanian-Kammerer, and F. Macià, "Semiclassical completely integrable systems: long-time dynamics and observability via two-microlocal Wigner measures", *Amer. J. Math.* **137**:3 (2015), 577–638. MR Zbl

[Anantharaman et al. 2016a] N. Anantharaman, M. Léautaud, and F. Macià, "Delocalization of quasimodes on the disk", *C. R. Math. Acad. Sci. Paris* **354**:3 (2016), 257–263. MR Zbl

[Anantharaman et al. 2016b] N. Anantharaman, M. Léautaud, and F. Macià, "Wigner measures and observability for the Schrödinger equation on the disk", *Invent. Math.* **206**:2 (2016), 485–599. MR Zbl

[Bialy and Polterovich 1989] M. L. Bialy and L. V. Polterovich, "Lagrangian singularities of invariant tori of Hamiltonian systems with two degrees of freedom", *Invent. Math.* **97**:2 (1989), 291–303. MR Zbl

[Bourgain 1993] J. Bourgain, "Eigenfunction bounds for the Laplacian on the n-torus", *Internat. Math. Res. Notices* **1993**:3 (1993), 61–66. MR Zbl

[Bourgain 2013] J. Bourgain, "Moment inequalities for trigonometric polynomials with spectrum in curved hypersurfaces", *Israel J. Math.* **193**:1 (2013), 441–458. MR Zbl

[Bourgain and Demeter 2015] J. Bourgain and C. Demeter, "The proof of the l^2 decoupling conjecture", *Ann. of Math.* (2) **182**:1 (2015), 351–389. MR Zbl

[Bourgain et al. 2013] J. Bourgain, N. Burq, and M. Zworski, "Control for Schrödinger operators on 2-tori: rough potentials", *J. Eur. Math. Soc.* (*JEMS*) **15**:5 (2013), 1597–1628. MR Zbl

[Burq and Zworski 2004] N. Burq and M. Zworski, "Geometric control in the presence of a black box", *J. Amer. Math. Soc.* **17**:2 (2004), 443–471. MR Zbl

[Burq and Zworski 2012] N. Burq and M. Zworski, "Control for Schrödinger operators on tori", *Math. Res. Lett.* **19**:2 (2012), 309–324. MR Zbl

[Fermanian-Kammerer 2000] C. Fermanian-Kammerer, "Mesures semi-classiques 2-microlocales", *C. R. Acad. Sci. Paris Sér. I Math.* **331**:7 (2000), 515–518. MR Zbl

[Fermanian-Kammerer 2005] C. Fermanian Kammerer, "Analyse à deux échelles d'une suite bornée de L^2 sur une sous-variété du cotangent", *C. R. Math. Acad. Sci. Paris* **340**:4 (2005), 269–274. MR Zbl

[Fermanian-Kammerer and Gérard 2002] C. Fermanian-Kammerer and P. Gérard, "Mesures semi-classiques et croisement de modes", *Bull. Soc. Math. France* **130**:1 (2002), 123–168. MR Zbl

[Gérard 1991] P. Gérard, "Mesures semi-classiques et ondes de Bloch", exposé 16 in *Séminaire sur les Équations aux Dérivées Partielles,* 1990–1991, École Polytech., Palaiseau, 1991. MR Zbl

[Jakobson 1997] D. Jakobson, "Quantum limits on flat tori", *Ann. of Math.* (2) **145**:2 (1997), 235–266. MR Zbl

[Macià 2009] F. Macià, "Semiclassical measures and the Schrödinger flow on Riemannian manifolds", *Nonlinearity* **22**:5 (2009), 1003–1020. MR Zbl

[Macià 2010] F. Macià, "High-frequency propagation for the Schrödinger equation on the torus", *J. Funct. Anal.* **258**:3 (2010), 933–955. MR Zbl

[Macià 2011] F. Macià, "The Schrödinger flow in a compact manifold: high-frequency dynamics and dispersion", pp. 275–289 in *Modern aspects of the theory of partial differential equations*, edited by M. Ruzhansky and J. Wirth, Oper. Theory Adv. Appl. **216**, Springer, 2011. MR

[Macià and Rivière 2016] F. Macià and G. Rivière, "Concentration and non-concentration for the Schrödinger evolution on Zoll manifolds", *Comm. Math. Phys.* **345**:3 (2016), 1019–1054. MR Zbl

[Macià and Riviere 2017] F. Macià and G. Riviere, "Observability and quantum limits for the Schrödinger equation on the sphere", preprint, 2017. arXiv

[Miller 1996] L. Miller, *Propagation d'ondes semi-classiques à travers une interface et mesures 2-microlocales*, Ph.D. thesis, École Polytechnique, 1996.

[Nier 1996] F. Nier, "A semi-classical picture of quantum scattering", *Ann. Sci. École Norm. Sup.* (4) **29**:2 (1996), 149–183. MR Zbl

[Vasy and Wunsch 2009] A. Vasy and J. Wunsch, "Semiclassical second microlocal propagation of regularity and integrable systems", *J. Anal. Math.* **108** (2009), 119–157. MR Zbl

[Weinstein 1977] A. Weinstein, "Asymptotics of eigenvalue clusters for the Laplacian plus a potential", *Duke Math. J.* **44**:4 (1977), 883–892. MR Zbl

[Wunsch 2008] J. Wunsch, "Spreading of Lagrangian regularity on rational invariant tori", *Comm. Math. Phys.* **279**:2 (2008), 487–496. MR Zbl

[Wunsch 2012] J. Wunsch, "Non-concentration of quasimodes for integrable systems", *Comm. Partial Differential Equations* **37**:8 (2012), 1430–1444. MR Zbl

[Zworski 2012] M. Zworski, *Semiclassical analysis*, Graduate Studies in Mathematics **138**, American Mathematical Society, Providence, RI, 2012. MR Zbl

[Zygmund 1974] A. Zygmund, "On Fourier coefficients and transforms of functions of two variables", *Studia Math.* **50** (1974), 189–201. MR Zbl

Received 29 Aug 2017. Revised 17 Jan 2018. Accepted 10 Apr 2018.

FABRICIO MACIÀ: fabricio.macia@upm.es
Universidad Politécnica de Madrid, ETSI Navales, Madrid, Spain

GABRIEL RIVIÈRE: gabriel.riviere@math.univ-lille1.fr
Laboratoire Paul Painlevé (U.M.R. CNRS 8524), U.F.R. de Mathématiques, Université Lille 1, Villeneuve d'Ascq, France

ANALYSIS AND PDE
Vol. 11, No. 8, 2018
dx.doi.org/10.2140/apde.2018.11.2137

SPECTRAL DISTRIBUTION OF THE FREE JACOBI PROCESS, REVISITED

TAREK HAMDI

We obtain a description for the spectral distribution of the free Jacobi process for any initial pair of projections. This result relies on a study of the unitary operator $RU_tSU_t^*$, where R, S are two symmetries and $(U_t)_{t\geq 0}$ is a free unitary Brownian motion, freely independent from $\{R, S\}$. In particular, for nonnull traces of R and S, we prove that the spectral measure of $RU_tSU_t^*$ possesses two atoms at ± 1 and an L^∞-density on the unit circle $\mathbb{T}$ for every $t > 0$. Next, via a Szegő-type transformation of this law, we obtain a full description of the spectral distribution of $PU_tQU_t^*$ beyond the case where $\tau(P) = \tau(Q) = \frac{1}{2}$. Finally, we give some specializations for which these measures are explicitly computed.

1. Introduction

Let P, Q be two projections in a W^*-probability space $(\mathscr{A}, \tau)$ which are free with $\{U_t, U_t^*, t \geq 0\}$. The present paper is a companion to the series of papers [Collins and Kemp 2014; Demni 2008; Demni 2016; Demni and Hamdi 2018; Demni et al. 2012; Demni and Hmidi 2014] devoted to the study of the spectral distribution, hereafter μ_t, of the self-adjoint-valued process $(X_t := PU_tQU_t^*P)_{t\geq 0}$. Viewed in the compressed algebra $(P\mathscr{A}P, \tau/\tau(P))$, X_t coincides with the so-called free Jacobi process with parameter $(\tau(P)/\tau(Q), \tau(Q))$, introduced by Demni [2008] via free stochastic calculus, as a solution to a free SDE there. Properties of its measure play important roles in free entropy and free information theory; see, e.g., [Hamdi 2017; 2018; Hiai and Ueda 2009; Izumi and Ueda 2015; Voiculescu 1999]. Furthermore, μ_t completely determines the structure of the von Neumann algebra generated by P and $U_tQU_t^*$ for any $t \geq 0$, see, e.g., [Hiai and Ueda 2009; Raeburn and Sinclair 1989], yielding a continuous interpolation from the law of PQP (when $t = 0$) to the free multiplicative convolution of the spectral measures of P and Q separately (when t tends to infinity). Indeed, the pair $(P, U_tQU_t^*)$ tends towards (P, UQU^*) as $t \to \infty$, where U is a Haar unitary free from $\{P, Q\}$. The two projections P and UQU^* are therefore free, see [Nica and Speicher 2006], and hence $\mu_{PUQU^*P} = \mu_P \boxtimes \mu_{UQU^*} = \mu_P \boxtimes \mu_Q$. The Lebesgue decomposition of the last term may be found in [Voiculescu et al. 1992, Example 3.6.7]. More generally, the operators P and $U_tQU_t^*$ are not free for finite t and the process $t \mapsto (P, U_tQU_t^*)$ is known as the free liberation of the pair (P, Q); see [Voiculescu 1999]. When both projections coincide, the series of papers [Demni 2016; Demni and Hamdi 2018; Demni et al. 2012; Demni and Hmidi 2014] aims to determine μ_t for any $t > 0$. In particular, when $P = Q$ and $\tau(P) = \frac{1}{2}$, Demni, Hmidi and the author proved in [Demni et al. 2012, Corollary 3.3] that the measure μ_t possesses a continuous density on

MSC2010: 42B37, 46L54.
Keywords: free Jacobi process, free unitary Brownian motion, multiplicative convolution, spectral distribution, Herglotz transform, Szegő transformation.

$(0, 1)$ for $t > 0$ which fits that of the random variable $(I + U_{2t} + (I + U_{2t})^*)/4$. Collins and Kemp [2014] extended this result to the case of two projections P, Q with traces $\frac{1}{2}$. Afterwards this result was partially extended in [Izumi and Ueda 2015] to arbitrary traces. In Proposition 3.1 of that paper, they proved

$$\mu_t = (1 - \min\{\tau(P), \tau(Q)\})\delta_0 + \max\{\tau(P) + \tau(Q) - 1, 0\}\delta_1 + \gamma_t,$$

where γ_t is a positive measure with no atom on $(0, 1)$ for every $t > 0$. In Proposition 3.3 of the same paper, they showed that when $\tau(P) = \tau(Q) = \frac{1}{2}$, this measure coincides with the Szegő transformation of the distribution of UU_t, where U is a unitary random variable determined by the law of PQP. Collins and Kemp [2014, Lemmas 3.2 and 3.6] studied the support of the measure γ_t, for arbitrary traces, and the way in which the edges of this support are propagated, but they were still not able to prove the continuity of γ_t.

The main result proved in this paper is a complete analysis of the spectral distribution of the unitary operator $RU_tSU_t^*$ (hereafter ν_t) for any symmetries $R, S \in \mathscr{A}$ which are free with $\{U_t, U_t^*\}$. In particular, we prove that the measure

$$\nu_t - \tfrac{1}{2}|\tau(R) - \tau(S)|\delta_\pi - \tfrac{1}{2}|\tau(R) + \tau(S)|\delta_0$$

possesses a continuous density κ_t on $\mathbb{T} = (-\pi, \pi]$. Using the relationship between μ_t and ν_t, when $\{P, Q\}$ and $\{R, S\}$ are associated, see [Hamdi 2017, Theorem 4.3], we deduce the regularity of μ_t for any initial projections. In particular, we prove that the measure γ_t possesses a continuous density on $[0, 1]$:

Theorem 1.1. *Let P, Q be orthogonal projections and U_t a free unitary Brownian motion, freely independent from P, Q. For every $t > 0$, the spectral distribution μ_t of the self adjoint operator $PU_tQU_t^*P$ is given by*

$$\mu_t = (1 - \min\{\tau(P), \tau(Q)\})\delta_0 + \max\{\tau(P) + \tau(Q) - 1, 0\}\delta_1 + \frac{\kappa_t(2\arccos(\sqrt{x}))}{2\pi\sqrt{x(1-x)}}\,\mathbf{1}_{[0,1]}(x)\,dx.$$

The paper ends with a striking observation on the spectral distribution of $RU_tSU_t^*$ at finite time t when the initial symmetries building it are centered and independent with respected to classical, free, monotone and boolean convolutions. In this respect, we notice that in the case of free independence, ν_t is stationary for all traces of the symmetries, and in the rest of cases, its given by a dilation of the law of U_t for centered symmetries. The result is as follows.

Theorem 1.2. *Let λ_t be the probability distribution of the free unitary Brownian motion U_t and $\mu = \frac{1}{2}(\delta_1 + \delta_{-1})$ (considered as a law on $\mathbb{T}$). We denote respectively by $\boxtimes, *, \uplus$ and $\triangleright$ the free, classical, boolean and monotone multiplicative convolutions. Then, for all $t \geq 0$:*

(1) *The measure $(\mu \boxtimes \mu) \boxtimes \lambda_t$ coincides with $\mu \boxtimes \mu$.*

(2) *The push-forward of $(\mu * \mu) \boxtimes \lambda_t$ by the map $z \mapsto z^2$ coincides with the law of U_{2t}.*

(3) *The push-forward of $(\mu \uplus \mu) \boxtimes \lambda_t$ by the map $z \mapsto z^3$ coincides with the law of U_{3t}.*

(4) *The push-forward of $(\mu \triangleright \mu) \boxtimes \lambda_t$ by the map $z \mapsto z^4$ coincides with the law of U_{4t}.*

The paper is organized as follows. For sake of completeness, we recall in the next section some preliminaries which gather useful information about the Herglotz transform of probability measures on

the unit circle, and the spectral distribution of the free unitary Brownian motion. In Section 3, we fix the basic ideas and notation for the rest of the work presented. In Section 4, we describe the spectral measure ν_t and prove our main result. In the last section, we present explicit computations of the spectral measure ν_t at finite time t when the initial operators are assumed to be centered and classically boolean or monotone independent.

2. Preliminaries

The Herglotz transform. Let $\mathscr{M}_{\mathbb{T}}$ denotes the set of probability measures on the unit circle $\mathbb{T}$. The normalized Lebesgue measure on $\mathbb{T}$ will be denoted by m. The Herglotz transform H_μ of a measure $\mu \in \mathscr{M}_{\mathbb{T}}$ is the analytic function in the unit disc $\mathbb{D}$ defined by the formula

$$H_\mu(z) = \int_{\mathbb{T}} \frac{\zeta + z}{\zeta - z}\, d\mu(\zeta).$$

This function is related to the moment-generating function of the measure μ

$$\psi_\mu(z) = \int_{\mathbb{T}} \frac{z}{\zeta - z}\, d\mu(\zeta), \quad z \in \mathbb{D},$$

by the simple formula $H_\mu(z) = 1 + 2\psi_\mu(z)$. Since any distribution on the unit circle is uniquely determined by its moments, we deduce that H_μ uniquely determines μ. One of the important applications of H is given in the following result; see, e.g., [Cima et al. 2006, Theorem 1.8.9]:

Theorem 2.1 (Herglotz). *The Herglotz transform sets up a bijection between analytic functions H on $\mathbb{D}$ with $\Re H \geq 0$ and $H(0) > 0$ and the nonzero measures $\mu \in \mathscr{M}_{\mathbb{T}}$.*

For $0 < p < \infty$, let $H^p(\mathbb{D})$ be the space of analytic functions f on $\mathbb{D}$ such that

$$\sup_{0<r<1} \int_{\mathbb{T}} |f(r\zeta)|^p\, d\zeta < \infty.$$

For $p = \infty$, let $H^\infty(\mathbb{D})$ denote the Hardy space consisting of all bounded analytic functions on $\mathbb{D}$ with the sup-norm. Let $L^p(\mathbb{T})$ denote the Lebesgue spaces on the circle $\mathbb{T}$ with respect to the normalized Lebesgue measure. The following result proves the existence of a boundary function for all $f \in H^p(\mathbb{D})$.

Theorem 2.2 [Cima et al. 2006, Theorem 1.9.4]. *Let $0 < p \leq \infty$ and $f \in H^p(\mathbb{D})$. Then the boundary function $\tilde{f}(\zeta)$ exists for m-almost all ζ in $\mathbb{T}$ and belongs to $L^p(\mathbb{T})$. Furthermore, the norms of f in $H^p(\mathbb{D})$ and of $\tilde{f}(\zeta)$ in $L^p(\mathbb{T})$ coincide.*

We know, see, e.g., [Cima et al. 2006, Lemma 2.1.11], that $H_\mu \in H^p(\mathbb{D})$ for all $0 < p < 1$; thus $\widetilde{H}_\mu(\zeta)$ exists for m-almost all ζ in $\mathbb{T}$. The density of μ can be recovered then from the boundary values of $\Re H_\mu$ by Fatou's theorem [Cima et al. 2006, Theorem 1.8.6] since $\Re \widetilde{H}_\mu = d\mu/dm$ m-a.e. Note that the atoms of $\mu \in \mathscr{M}_{\mathbb{T}}$ can also be recovered from H_μ by Lebesgue's dominated convergence theorem via

$$\lim_{r \to 1^-} (1 - r) H_\mu(r\zeta) = 2\mu\{\zeta\} \quad \text{for all } \zeta \in \mathbb{T}.$$

Spectral distribution of the free unitary Brownian motion. For $\mu \in \mathscr{M}_{\mathbb{T}}$, let ψ_μ denote its moment-generating function and χ_μ the function $\psi_\mu/(1+\psi_\mu)$. If μ has nonzero mean, we denote by χ_μ^{-1} the inverse function of χ_μ in some neighborhood of zero. In this case the Σ-transform of μ is defined by $\Sigma_\mu(z) = (1/z)\chi_\mu^{-1}(z)$. The spectral distribution λ_t of the free unitary Brownian motion was introduced by Biane [1997a] as the unique probability measure on $\mathbb{T}$ such that its Σ-transform is given by

$$\Sigma_{\lambda_t}(z) = \exp\left(\frac{t}{2}\frac{1+z}{1-z}\right).$$

It is the multiplicative analog of the semicircular distribution. Its moments are the large-size limits of observables of the free Brownian motion (of dimension d) $(U_t^{(d)})_{t\geq 0}$ on the unitary group $\mathscr{U}(d)$:

$$\lim_{d\to\infty} \frac{1}{d}\mathbb{E}(\mathrm{tr}[U_{t/d}^{(d)}]^k) = \int_{\mathbb{T}} \zeta^k \, d\lambda_t(\zeta), \quad k \geq 0.$$

This result was proved independently by Biane [1997a] and Rains [1997], who explicitly calculated these moments:

$$\tau(U_t^k) = e^{-kt/2}\sum_{j=0}^{k-1} \frac{(-t)^j}{j!}\binom{k}{j+1}k^{j-1}, \quad k \geq 0. \tag{2-1}$$

The equality (2-1) can be transformed into the PDE

$$\partial_t H + zH\,\partial_z H = 0, \tag{2-2}$$

with the initial condition $H(0,z) = (1+z)/(1-z)$ for the Herglotz transform $H_{\lambda_{2t}}(z)$; see, e.g., the proof of [Izumi and Ueda 2015, Proposition 3.3]. The measure λ_t is described in [Biane 1997b] from the boundary behavior of the inverse function of $H_{\lambda_t}(z)$ as follows.

Theorem 2.3 [Biane 1997b]. *For every $t > 0$, the measure λ_t has a continuous density ρ_t with respect to the normalized Lebesgue measure on $\mathbb{T}$. Its support is the connected arc $\{e^{i\theta} : |\theta| \leq g(t)\}$ with*

$$g(t) := \tfrac{1}{2}\sqrt{t(4-t)} + \arccos\left(1 - \tfrac{1}{2}t\right)$$

for $t \in [0,4]$, and the whole circle for $t > 4$. The density ρ_t is determined by $\Re h_t(e^{i\theta})$, where $z = h_t(e^{i\theta})$ is the unique solution (with positive real part) to

$$\frac{z-1}{z+1}e^{zt/2} = e^{i\theta}.$$

3. Notation

We use here the same symbols as in [Hamdi 2017; 2018]. To a given pair of projections P, Q in $\mathscr{A}$ that are independent of $(U_t)_{t\geq 0}$ we associate the symmetries $R = 2P - I$ and $S = 2Q - I$. Set $\alpha = \tau(R)$ and $\beta = \tau(S)$. We sometimes use the notation $a = |\alpha - \beta|/2$ and $b = |\alpha + \beta|/2$ for simplicity. Keep the symbols μ_t and ν_t above. The unit circle is identified with $(-\pi, \pi]$ by $e^{i\theta}$. According to [Hamdi 2017, Section 3], the measure ν_t is connected to μ_t by the formula

$$\nu_t = 2\hat{\mu}_t - \tfrac{1}{2}(2-\alpha-\beta)\delta_\pi - \tfrac{1}{2}(\alpha+\beta)\delta_0, \tag{3-1}$$

where

$$\hat{\mu}_t := \tfrac{1}{2}(\tilde{\mu}_t + (\tilde{\mu}_t|_{(0,\pi)}) \circ j^{-1}) \tag{3-2}$$

is the symmetrization on $(-\pi, \pi)$, with the mapping $j : \theta \in (0, \pi) \mapsto -\theta \in (-\pi, 0)$, of the positive measure $\tilde{\mu}_t(d\theta)$ on $[0, \pi]$ obtained from $\mu_t(dx)$ via the variable change $x = \cos^2(\theta/2)$. Equivalently, we obtain the following relationship between the Herglotz transforms H_{μ_t} and H_{ν_t}:

$$H_{\nu_t}(z) = \frac{z-1}{z+1} H_{\mu_t}\left(\frac{4z}{(1+z)^2}\right) - 2(\alpha+\beta)\frac{z}{z^2-1}; \tag{3-3}$$

see [Hamdi 2017, Corollary 4.2]. The function $H_{\nu_t}(z)$, which we shall denote by $H(t, z)$, is analytic in both variables $z \in \mathbb{D}$ and $t > 0$, see [Collins and Kemp 2014, Theorem 1.4], and solves the PDE

$$\partial_t H + zH\,\partial_z H = \frac{2z(\alpha z^2 + 2\beta z + \alpha)(\beta z^2 + 2\alpha z + \beta)}{(1-z^2)^3}, \tag{3-4}$$

see [Hamdi 2017, Proposition 2.3]. Let

$$K(t, z) := \sqrt{H(t, z)^2 - \left(a\frac{1-z}{1+z} + b\frac{1+z}{1-z}\right)^2}. \tag{3-5}$$

The PDE (3-4) is then transformed into

$$\partial_t K + zH(t, z)\,\partial_z K = 0.$$

Note that steady state solution $K(\infty, z)$ is the constant $\sqrt{1-(a+b)^2}$; see [Hamdi 2017, Remark 3.3]. The ordinary differential equations (ODEs for short) of the characteristic curves associated with this PDE are

$$\begin{cases} \partial_t \phi_t(z) = \phi_t(z) H(t, \phi_t(z)), \quad \phi_0(z) = z, \\ \partial_t [K(t, \phi_t(z))] = 0. \end{cases} \tag{3-6}$$

The second ODE of (3-6) implies that $K(t, \phi_t(z)) = K(0, z)$, while the first one is nothing but the radial Loewner ODE, see [Lawler 2005, Theorem 4.14], which defines a unique family of conformal transformations ϕ_t from some region $\Omega_t \subset \mathbb{D}$ onto $\mathbb{D}$ with $\phi_t(0) = 0$ and $\partial_z \phi_t(0) = e^t$. Moreover, from [Lawler 2005, Remark 4.15], ϕ_t is invertible from Ω_t onto $\mathbb{D}$ and it has a continuous extension to $\mathbb{T} \cap \overline{\Omega}_t$ by [Hamdi 2018, Proposition 2.1]. Integrating the first ODE in (3-6), we get

$$\phi_t(z) = z \exp\left(\int_0^t H(s, \phi_s(z))\, ds\right).$$

Let us define

$$h_t(r, \theta) = 1 - \int_0^t \frac{1 - |\phi_s(re^{i\theta})|^2}{-\ln r} \int_{\mathbb{T}} \frac{1}{|\xi - \phi_s(re^{i\theta})|^2}\, d\nu_s(\xi)\, ds,$$

so that

$$\ln|\phi_t(re^{i\theta})| = \ln r + \Re \int_0^t H(s, \phi_s(re^{i\theta}))\, ds = (\ln r) h_t(r, \theta). \tag{3-7}$$

Define $R_t : [-\pi, \pi] \to [0, 1]$ as

$$R_t(\theta) = \sup\{r \in (0, 1) : h_t(r, \theta) > 0\},$$

and let

$$I_t = \{\theta \in [-\pi, \pi] : h_t(\theta) < 0\},$$

where $h_t(\theta) = \lim_{r\to 1^-} h_t(r, \theta) \in \mathbb{R} \cup \{-\infty\}$; see the fact given under Lemma 3.2 in [Hamdi 2018]. The next result gives a description of Ω_t and its boundary.

Proposition 3.1 [Hamdi 2018, Proposition 3.3]. *For any $t > 0$, we have*:

(1) $\Omega_t = \{re^{i\theta} : h_t(r, e^{i\theta}) > 0\}$.

(2) $\partial\Omega_t \cap \mathbb{D} = \{re^{i\theta} : h_t(r, e^{i\theta}) = 0 \text{ and } \theta \in I_t\}$.

(3) $\partial\Omega_t \cap \mathbb{T} = \{e^{i\theta} : h_t(r, e^{i\theta}) = 0 \text{ and } \theta \in [-\pi, \pi] \setminus I_t\}$.

In closing, we recall the following result which will be of use later on; see the proof of Theorem 1.1 in [Hamdi 2018].

Lemma 3.2 [Hamdi 2018]. *For every $t > 0$, the function $K(t, \cdot)$ has a continuous extension to the unit circle $\mathbb{T}$.*

4. Analysis of spectral distributions of $RU_tSU_t^*$

In this section, we shall prove Theorem 1.1. To this end, we start by giving a description of the spectral measure ν_t of $RU_tSU_t^*$ for any $t > 0$, and deriving a formula for its density. We notice that from the asymptotic freeness of R and $U_tSU_t^*$, the measure ν_t converges weakly as $t \to \infty$, see [Hamdi 2017, Proposition 2.6], to

$$\nu_\infty = a\delta_\pi + b\delta_0 + \frac{\sqrt{-(\cos\theta - r_+)(\cos\theta - r_-)}}{2\pi|\sin\theta|}\,\mathbf{1}_{(\theta_-,\theta_+)\cup(-\theta_+,-\theta_-)}\,d\theta, \tag{4-1}$$

with $r_\pm = -\alpha\beta \pm \sqrt{(1-\alpha^2)(1-\beta^2)}$ and $\theta_\pm = \arccos r_\pm$. The following theorem asserts that an analogous result holds for finite t.

Theorem 4.1. *For every $t > 0$, the measure $\nu_t - a\delta_\pi - b\delta_0$ is absolutely continuous with respect to the normalized Lebesgue measure on $\mathbb{T} = (-\pi, \pi]$. Moreover, its density κ_t at the point $e^{i\theta}$ is equal to the real part of*

$$\sqrt{[K(t, e^{i\theta})]^2 + (a+b)^2 - 1 - \frac{(\cos\theta - r_+)(\cos\theta - r_-)}{\sin^2\theta}}.$$

Proof. Define the function

$$L(t, z) = \int_{\mathbb{T}} \frac{e^{i\theta} + z}{e^{i\theta} - z}(\nu_t - a\delta_\pi - b\delta_0)(d\theta) = H(t, z) - a\frac{1-z}{1+z} - b\frac{1+z}{1-z}.$$

The real part of this function is nothing but the Poisson integral of the measure $\nu_t - a\delta_\pi - b\delta_0$. Using (3-5) and multiplying by the conjugate, we get

$$\begin{aligned} L(t,z) &= \frac{K(t,z)^2}{\sqrt{K(t,z)^2 + \left(a\frac{1-z}{1+z} + b\frac{1+z}{1-z}\right)^2} + a\frac{1-z}{1+z} + b\frac{1+z}{1-z}} \\ &= \frac{(1-z^2)K(t,z)^2}{\sqrt{[(1-z^2)K(t,z)]^2 + [a(1-z)^2 + b(1+z)^2]^2} + a(1-z)^2 + b(1+z)^2}. \end{aligned}$$

Note that $K(t,z)$ extends continuously to $\mathbb{T}$ by Lemma 3.2. The denominator of the above expression does not vanish on the closed unit disc and

$$z \mapsto (1-z^2)^2 K(t,z)^2 + [a(1-z)^2 + b(1+z)^2]^2 = (1-z^2)H(t,z)^2$$

does not take negative values. These together imply that $L(t,z)$ has a continuous extension on the boundary $\mathbb{T}$. Hence, by uniqueness of the Herglotz representation (see Theorem 2.1), the measure $\nu_t - a\delta_\pi - b\delta_0$ is absolutely continuous with respect to the Haar measure in $\mathbb{T}$ and its density is given by

$$\begin{aligned} \Re\left[H(t,e^{i\theta}) - a\frac{1-e^{i\theta}}{1+e^{i\theta}} - b\frac{1+e^{i\theta}}{1-e^{i\theta}}\right] &= \Re\sqrt{[K(t,e^{i\theta})]^2 + \left[a\frac{1-e^{i\theta}}{1+e^{i\theta}} - b\frac{1+e^{i\theta}}{1-e^{i\theta}}\right]^2} \\ &= \Re\sqrt{[K(t,e^{i\theta})]^2 - [a\tan(\theta/2) - b\cot(\theta/2)]^2}. \end{aligned}$$

To complete the proof, we need only show that

$$[a\tan(\theta/2) - b\cot(\theta/2)]^2 = 1 - (a+b)^2 + \frac{(\cos\theta - r_+)(\cos\theta - r_-)}{\sin^2\theta}$$

or equivalently that

$$(1-a^2-b^2)\sin^2\theta - a^2\sin^2\theta\tan^2(\theta/2) - b^2\sin^2\theta\cot^2(\theta/2) = -(\cos\theta - r_+)(\cos\theta - r_-).$$

Working from the left-hand side and using the identities

$$\sin^2\theta = 1 - \cos^2\theta, \quad \sin^2\theta\tan^2(\theta/2) = (1-\cos\theta)^2, \quad \sin^2\theta\cot^2(\theta/2) = (1+\cos\theta)^2,$$

we get

$$(1-a^2-b^2)(1-\cos^2\theta) - a^2(1-\cos\theta)^2 - b^2(1+\cos\theta)^2.$$

Rearranging these terms, we obtain

$$-\cos^2\theta + 2(a^2-b^2)\cos\theta - 2(a^2+b^2) + 1.$$

So, by substituting the equalities $\alpha\beta = b^2 - a^2$ and $\alpha^2 + \beta^2 = 2(a^2+b^2)$, we obtain the required formula:

$$-\cos^2\theta - 2\alpha\beta\cos\theta + 1 - \alpha^2 - \beta^2 = -(\cos\theta - r_+)(\cos\theta - r_-). \qquad \square$$

Remark 4.2. We can prove directly that κ_t is an L^∞-density. In fact, by (3-5), we have

$$K(t,z)^2 = H(t,z)^2 - \left(a\frac{1-z}{1+z} + b\frac{1+z}{1-z}\right)^2 = L(t,z)\left(L(t,z) + 2a\frac{1-z}{1+z} + 2b\frac{1+z}{1-z}\right).$$

Then

$$(\Re L(t,z))^2 \le \Re L(t,z)\Re\left(L(t,z) + 2a\frac{1-z}{1+z} + 2b\frac{1+z}{1-z}\right) \le |K(t,z)^2|.$$

But, the function $K(t,z)$ is analytic in $\mathbb{D}$ and extends continuously to $\mathbb{T}$. It becomes then of Hardy class $H^\infty(\mathbb{D})$, and hence the density of $\nu_t - a\delta_\pi - b\delta_0$ belongs to $L^\infty(\mathbb{T})$ by [Koosis 1998, Theorem on p. 15].

Proposition 4.3. *The support of ν_t is a subset of $\{\phi_t(R_t(\theta)e^{i\theta}) : \theta \in I_t\}$.*

Proof. By (3-7), we have

$$\int_0^t \Re H(s, \phi_s(R_t(\theta)e^{i\theta}))\, ds = -\ln R_t(\theta),$$

where we used the fact that $\ln|\phi_t(R_t(\theta)e^{i\theta})| = 0$ due to the equality $|\phi_t(R_t(\theta)e^{i\theta})| = 1$. Then, by continuity of $s \mapsto \Re H(s, \phi_s(R_t(\theta)e^{i\theta}))$ on $[0,t]$, we deduce that the assertion $\Re H(t, \phi_t(R_t(\theta)e^{i\theta})) > 0$ yields $R_t(\theta) \ne 1$. Finally, by the definition of $R_t(\theta)$ and I_t, we have

$$\{\theta : R_t(\theta) \ne 1\} = \{\theta : \exists\, r_0 \in (0,1),\ h_t(r_0, e^{i\theta}) = 0\} = \{\theta : h_t(\theta) < 0\} = I_t. \qquad \square$$

We now proceed to the proof of Theorem 1.1.

Proof of Theorem 1.1. By (3-1), we have

$$\nu_t - a\delta_\pi - b\delta_0 = 2\big[\hat{\mu}_t - (1 - \min\{\tau(P), \tau(Q)\})\delta_\pi - \max\{\tau(P) + \tau(Q) - 1, 0\}\delta_0\big].$$

This measure is absolutely continuous with respect to the normalized Lebesgue measure $d\theta/(2\pi)$ on $\mathbb{T} = (-\pi, \pi]$, by Theorem 4.1, and its density is given by the function κ_t. Hence, (3-2) implies

$$\big(\tilde{\mu}_t - (1 - \min\{\tau(P), \tau(Q)\})\delta_\pi - \max\{\tau(P) + \tau(Q) - 1, 0\}\delta_0\big)(d\theta) = \kappa_t(\theta)\frac{d\theta}{2\pi}, \quad \theta \in [0, \pi],$$

and so the desired assertion holds via the variable change $\theta = 2\arccos(\sqrt{x})$. $\square$

Remark 4.4. It is worth noting that the spectral distribution ν_t is stationary for all traces of the symmetries, when the initial operators R and S are free. Actually, by Proposition 2.5 in [Hamdi 2017], we have

$$H(0,z) = \sqrt{1 + 4z\left(\frac{b^2}{(1-z)^2} - \frac{a^2}{(1+z)^2}\right)},$$

so that

$$K(0,z) = \sqrt{H(0,z)^2 - \left(a\frac{1-z}{1+z} + b\frac{1+z}{1-z}\right)^2} = \sqrt{1 - (a+b)^2}.$$

Hence, for every $z \in \mathbb{D}$ and $t \ge 0$, we have $K(t,z) = K(0, \phi_t^{-1}(z)) = \sqrt{1-(a+b)^2}$, and therefore ν_t coincides with the measure ν_∞.

The above fact can be explained directly by use of the sequence of moments

$$m_n(t) := \tau[(PU_tQU_t^*P)^n], \quad n \geq 1.$$

In fact, we can prove by induction on n that $m_n(t)$ becomes stationary when P and Q are free. Recall from [Demni et al. 2012] that $m_n(t)$ satisfy the infinite system of ODEs

$$\partial_t m_1(t) = -m_1(t) + \tau[P]\tau[Q], \tag{4-2}$$

$$\partial_t m_n(t) = -nm_n(t) + n\sum_{k=1}^{n-1} m_{n-k}(t)(m_{k-1}(t) - m_k(t)), \quad n \geq 2, \tag{4-3}$$

with $m_0(t) = \tau[P] + \tau[Q]$. When $n = 1$, (4-2) can be solved explicitly and gives $m_1(t) = \tau[P]\tau[Q] + e^{-t}(m_1(0) - \tau[P]\tau[Q])$. Since $m_1(0) = \tau[PQ] = \tau[P]\tau[Q]$ by freeness, we get $m_1(t) = m_1(0)$. For $n \geq 2$, we note that the moments

$$c_n := m_n(0) = \tau[(PQ)^n]$$

satisfy

$$c_n = \sum_{k=1}^{n-1} c_{n-k}(c_{k-1} - c_k).$$

Assume that $m_k(t) = c_k$ holds up to level $n-1$. Then, the ODE (4-3) can be written in the form

$$\partial_t m_n(t) = -nm_n(t) + nc_n,$$

with solution the constant c_n. Thus, μ_t (and therefore ν_t) is stationary.

5. Special cases

We present here some specializations for which the measure ν_t (and hence μ_t) is explicitly determined.

Centered initial operators. That is, $\tau(R) = \tau(S) = 0$ or $a = b = 0$. In this case, the PDE (3-4) can be rewritten as

$$\partial_t H + zH\,\partial_z H = 0,$$

and the measure ν_t becomes identical to the probability distribution of UU_{2t}, where U is a free unitary whose distribution is ν_0; see [Izumi and Ueda 2015, Proposition 3.3] or [Hamdi 2017, Remark 4.7]. Hence, the measure ν_t is given by the multiplicative free convolution $\nu_0 \boxtimes \lambda_{2t}$, studied in [Zhong 2015]. The density of this measure and its support are explicitly computed in Theorem 3.8 and Corollary 3.9 of that paper. In particular, when ν_0 is a Dirac mass at 1 (on the unit circle), the Herglotz transforms $H(t, z)$ of ν_t satisfy the PDE

$$\partial_t H + zH\,\partial_z H = 0, \quad H(0, z) = \frac{1+z}{1-z}.$$

Then it follows from the uniqueness of the solution of (2-2) that $H(t, z) = H_{\lambda_{2t}}(z)$, and by uniqueness of the Herglotz representation, ν_t coincides with the law λ_{2t} of U_{2t}. Hence, by Theorem 2.3 the density

of ν_t is given by the formula $\kappa_t(\omega) = \rho_{2t}(\omega)$ and the support is the full unit circle for $t > 2$ and the set $\{e^{i\theta} : |\theta| < g(2t)\}$ for $t \in [0, 2]$.

In the rest of the paper, we illustrate how the family of measures $(\nu_t)_{t\geq 0}$ provides a continuous interpolation between freeness and different type of independence.

Classically independent initial operators. In this case, the measure ν_t is considered as a t-free convolution which interpolates between classical independence and free independence; see [Benaych-Georges and Lévy 2011]. Let R, S be two independent symmetries. From the facts given above Lemma 5.4 in [Hamdi 2017], we have

$$H(0, z) = 1 + 2\sum_{n\geq 1} \tau(R^n)\tau(S^n)z^n = \frac{1 + z^2 + 2z\tau(R)\tau(S)}{1 - z^2}.$$

In particular, when $\tau(R) = \tau(S) = 0$, the function $H(t, z)$ satisfies the PDE

$$\partial_t H + zH\,\partial_z H = 0, \quad H(0, z) = \frac{1 + z^2}{1 - z^2},$$

and hence, by (2-2), it coincides with $H_{\lambda_{4t}}(z^2)$. We retrieve then the result obtained in [Benaych-Georges and Lévy 2011, Theorem 3.6]: for any $t \geq 0$, the push-forward of ν_t by the map $z \mapsto z^2$ coincides with the law of U_{4t}. In particular, the density of ν_t is given by $\kappa_t(\omega) = \rho_{4t}(\omega^2)$ for any ω in the unit circle and the support is the full unit circle for $t > 1$ and the set $\{e^{i\theta} : |\theta| < g(4t)/2\}$ for $t \in [0, 1]$.

Boolean independent initial operators. To a given probability measure μ on the unit circle, we keep the same notation ψ_μ, H_μ and χ_μ as in Section 2. Let $\mu_1, \mu_2 \in \mathscr{M}_{\mathbb{T}}$ and set $F_\mu(z) = (1/z)\chi_\mu(z)$. Then the multiplicative boolean convolution $\mu = \mu_1 \uplus \mu_2$ is uniquely determined by

$$F_\mu(z) = F_{\mu_1}(z)F_{\mu_2}(z);$$

see [Hamdi 2015; Franz 2008] for more details. Then, for boolean independent symmetries R, S with law $\mu = \frac{1}{2}(\delta_1 + \delta_{-1})$, we have

$$\psi_\mu(z) = \frac{z^2}{1 - z^2}, \quad \chi_\mu(z) = z^2, \quad F_\mu(z) = z$$

and therefore $F_{\mu\uplus\mu}(z) = F_\mu(z)^2 = z^2$. It follows that

$$\psi_{\mu\uplus\mu}(z) = \frac{z^3}{1 - z^3} \quad \text{and} \quad H_{\mu\uplus\mu}(z) = \frac{1 + z^3}{1 - z^3}.$$

Hence, by (2-2) the Herglotz transform $H(t, z)$ of ν_t and $H_{\lambda_{6t}}(z^3)$ solve the same PDE with the initial condition $H(0, z) = (1 + z^3)/(1 - z^3)$. By uniqueness, it follows that the push-forward of ν_t by the map $z \mapsto z^3$ coincides with the law of U_{6t} for any $t \geq 0$. In particular, we have $\kappa_t(\omega) = \rho_{6t}(\omega^3)$ for any ω in the unit circle and ν_t is supported in the full unit circle for $t > \frac{2}{3}$ and the set $\{e^{i\theta} : |\theta| < g(6t)/3\}$ for $t \in \left[0, \frac{2}{3}\right]$.

Monotone independent initial operators. For $\mu_1, \mu_2 \in \mathscr{M}_{\mathbb{T}}$, the multiplicative monotone convolution $\mu = \mu_1 \triangleright \mu_2$ is uniquely determined by

$$\chi_\mu(z) = \chi_{\mu_1}(\chi_{\mu_2}(z));$$

see [Hamdi 2015; Franz 2006] for more details. Here, we shall compute the measure ν_t for monotone independent symmetries R, S with law $\mu = \frac{1}{2}(\delta_1 + \delta_{-1})$. As usual, we have

$$\psi_\mu(z) = \frac{z^2}{1-z^2}, \quad \chi_\mu(z) = z^2,$$

and then $\chi_{\mu\triangleright\mu}(z) = \chi_\mu(\chi_\mu(z)) = z^4$. Hence,

$$\psi_{\mu\triangleright\mu}(z) = \frac{z^4}{1-z^4} \quad \text{and} \quad H_{\mu\triangleright\mu}(z) = \frac{1+z^4}{1-z^4}.$$

It follows that $H(t,z) = H_{\lambda_{8t}}(z^4)$ by uniqueness. Thus, the push-forward of ν_t by the map $z \mapsto z^4$ coincides with the law of U_{8t} for any $t \geq 0$. In particular, we have $\kappa_t(\omega) = \rho_{8t}(\omega^4)$ for any ω in the unit circle and ν_t is supported in the full unit circle for $t > \frac{1}{2}$ and the set $\{e^{i\theta} : |\theta| < g(8t)/4\}$ for $t \in \left[0, \frac{1}{2}\right]$.

Finally, we recall (see the first subsection above) that $\nu_t = \nu_0 \boxtimes \lambda_{2t}$ for centered initial operators R, S (i.e., $\tau(R) = \tau(S) = 0$). Hence, the discussions so far can be summarized in Theorem 1.2.

References

[Benaych-Georges and Lévy 2011] F. Benaych-Georges and T. Lévy, "A continuous semigroup of notions of independence between the classical and the free one", *Ann. Probab.* **39**:3 (2011), 904–938. MR Zbl

[Biane 1997a] P. Biane, "Free Brownian motion, free stochastic calculus and random matrices", pp. 1–19 in *Free probability theory* (Waterloo, ON, 1995), edited by D.-V. Voiculescu, Fields Inst. Commun. **12**, Amer. Math. Soc., Providence, RI, 1997. MR Zbl

[Biane 1997b] P. Biane, "Segal–Bargmann transform, functional calculus on matrix spaces and the theory of semi-circular and circular systems", *J. Funct. Anal.* **144**:1 (1997), 232–286. MR Zbl

[Cima et al. 2006] J. A. Cima, A. L. Matheson, and W. T. Ross, *The Cauchy transform*, Mathematical Surveys and Monographs **125**, Amer. Math. Soc., Providence, RI, 2006. MR Zbl

[Collins and Kemp 2014] B. Collins and T. Kemp, "Liberation of projections", *J. Funct. Anal.* **266**:4 (2014), 1988–2052. MR Zbl

[Demni 2008] N. Demni, "Free Jacobi process", *J. Theoret. Probab.* **21**:1 (2008), 118–143. MR Zbl

[Demni 2016] N. Demni, "Free Jacobi process associated with one projection: local inverse of the flow", *Complex Anal. Oper. Theory* **10**:3 (2016), 527–543. MR Zbl

[Demni and Hamdi 2018] N. Demni and T. Hamdi, "Inverse of the flow and moments of the free Jacobi process associated with one projection", *Random Matrices Theory Appl.* **7**:2 (2018), art. id. 1850001. MR

[Demni and Hmidi 2014] N. Demni and T. Hmidi, "Spectral distribution of the free Jacobi process associated with one projection", *Colloq. Math.* **137**:2 (2014), 271–296. MR Zbl

[Demni et al. 2012] N. Demni, T. Hamdi, and T. Hmidi, "Spectral distribution of the free Jacobi process", *Indiana Univ. Math. J.* **61**:3 (2012), 1351–1368. MR Zbl

[Franz 2006] U. Franz, "Multiplicative monotone convolutions", pp. 153–166 in *Quantum probability*, edited by M. Bożejko et al., Banach Center Publ. **73**, Polish Acad. Sci. Inst. Math., Warsaw, 2006. MR Zbl

[Franz 2008] U. Franz, "Boolean convolution of probability measures on the unit circle", pp. 83–94 in *Analyse et probabilités*, edited by P. Biane et al., Sémin. Congr. **16**, Soc. Math. France, Paris, 2008. MR Zbl

[Hamdi 2015] T. Hamdi, "Monotone and boolean unitary Brownian motions", *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **18**:2 (2015), art. id. 1550012. MR Zbl

[Hamdi 2017] T. Hamdi, "Liberation, free mutual information and orbital free entropy", preprint, 2017. arXiv

[Hamdi 2018] T. Hamdi, "Free mutual information for two projections", *Complex Anal. Oper. Theory* (Online publication April 2018).

[Hiai and Ueda 2009] F. Hiai and Y. Ueda, "A log-Sobolev type inequality for free entropy of two projections", *Ann. Inst. Henri Poincaré Probab. Stat.* **45**:1 (2009), 239–249. MR Zbl

[Izumi and Ueda 2015] M. Izumi and Y. Ueda, "Remarks on free mutual information and orbital free entropy", *Nagoya Math. J.* **220** (2015), 45–66. MR Zbl

[Koosis 1998] P. Koosis, *Introduction to H_p spaces*, 2nd ed., Cambridge Tracts in Mathematics **115**, Cambridge University Press, 1998. MR Zbl

[Lawler 2005] G. F. Lawler, *Conformally invariant processes in the plane*, Mathematical Surveys and Monographs **114**, Amer. Math. Soc., Providence, RI, 2005. MR Zbl

[Nica and Speicher 2006] A. Nica and R. Speicher, *Lectures on the combinatorics of free probability*, London Mathematical Society Lecture Note Series **335**, Cambridge University Press, Cambridge, 2006. MR Zbl

[Raeburn and Sinclair 1989] I. Raeburn and A. M. Sinclair, "The C^*-algebra generated by two projections", *Math. Scand.* **65**:2 (1989), 278–290. MR Zbl

[Rains 1997] E. M. Rains, "Combinatorial properties of Brownian motion on the compact classical groups", *J. Theoret. Probab.* **10**:3 (1997), 659–679. MR Zbl

[Voiculescu 1999] D. Voiculescu, "The analogues of entropy and of Fisher's information measure in free probability theory, VI: Liberation and mutual free information", *Adv. Math.* **146**:2 (1999), 101–166. MR Zbl

[Voiculescu et al. 1992] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free random variables: a noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups*, CRM Monograph Series **1**, Amer. Math. Soc., Providence, RI, 1992. MR Zbl

[Zhong 2015] P. Zhong, "On the free convolution with a free multiplicative analogue of the normal distribution", *J. Theoret. Probab.* **28**:4 (2015), 1354–1379. MR Zbl

Received 23 Nov 2017. Revised 20 Mar 2018. Accepted 19 Apr 2018.

TAREK HAMDI: tarek.hamdi@mail.com
Department of Management Information Systems, College of Business Administration, Qassim University, Buraydah, Saudi Arabia

and

Laboratoire d'Analyse Mathématiques et Applications LR11ES11, Université de Tunis El-Manar, Tunis, Tunisia

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 11 No. 8 2018