ANALYSIS & PDE

Volume 11 No. 6 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. 6, 2018
dx.doi.org/10.2140/apde.2018.11.1343

msp

PROPAGATION OF CHAOS, WASSERSTEIN GRADIENT FLOWS AND TORIC KÄHLER–EINSTEIN METRICS

ROBERT J. BERMAN AND MAGNUS ÖNNHEIM

Motivated by a probabilistic approach to Kähler–Einstein metrics we consider a general nonequilibrium statistical mechanics model in Euclidean space consisting of the stochastic gradient flow of a given (possibly singular) quasiconvex N-particle interaction energy. We show that a deterministic "macroscopic" evolution equation emerges in the large N-limit of many particles. This is a strengthening of previous results which required a uniform two-sided bound on the Hessian of the interaction energy. The proof uses the theory of weak gradient flows on the Wasserstein space. Applied to the setting of permanental point processes at "negative temperature", the corresponding limiting evolution equation yields a drift-diffusion equation, coupled to the Monge–Ampère operator, whose static solutions correspond to toric Kähler–Einstein metrics. This drift-diffusion equation is the gradient flow on the Wasserstein space of probability measures of the K-energy functional in Kähler geometry and it can be seen as a fully nonlinear version of various extensively studied dissipative evolution equations and conservation laws, including the Keller–Segel equation and Burger's equation. In a companion paper, applications to singular pair interactions in one dimension are given.

1. Introduction

The present work is motivated by the probabilistic approach to the construction of canonical metrics, or more precisely Kähler–Einstein metrics, on complex algebraic varieties introduced in [Berman 2013a; 2017], formulated in terms of certain β-deformations of determinantal (fermionic) point processes. The approach in those papers uses ideas from equilibrium statistical mechanics (Boltzmann–Gibbs measures) and the main challenge concerns the existence problem for Kähler–Einstein metrics on a complex manifold X with *positive* Ricci curvature, which is closely related to the seminal Yau–Tian–Donaldson conjecture in complex geometry. In this paper, which is one in a series, we will be concerned with a

MSC2010: 00A05.
Keywords: statistical mechanics, Kähler–Einstein metrics, propagation of chaos, Langvin equation.

dynamic version of the probabilistic approach in [Berman 2013a; 2017]. In other words, we are in the realm of nonequilibrium statistical mechanics, where the relaxation to equilibrium is studied. As the general complex geometric setting appears to be extremely challenging, due to the severe singularities and nonlinearity of the corresponding interaction energies, we will here focus on the real analog of the complex setting introduced in [Berman 2013b], taking place in $\mathbb{R}^n$ and which corresponds to the case when X is a *toric* complex algebraic variety. As explained in that paper in this real setting the determinantal (fermionic) processes are replaced by *permanental* (bosonic) processes and convexity plays the role of positive Ricci curvature/plurisubharmonicity (see Section 5C for some geometric background).

Our main result (Theorem 1.1) shows that a deterministic evolution equation on the space of all probability measures on $\mathbb{R}^n$ emerges from the underlying stochastic dynamics, which as explained below, can be seen as a new "propagation of chaos" result. The evolution equation in question is a drift-diffusion equation coupled to the fully nonlinear real Monge–Ampère operator. It turns out that in the case of the real line (i.e., $n = 1$) this equation is closely related to various extensively studied evolution equations, notably the Keller–Segel equation in chemotaxis [Keller and Segel 1970], Burger's equation [Hopf 1950; Frisch and Bec 2001] in the theory of nonlinear waves and scalar conservation laws and the deterministic version of the Kardar–Parisi–Zhang (KPZ) equation describing surface growth [Kardar et al. 1986]. In the higher-dimensional real case, the equation can be viewed as a dissipative viscous version of the semigeostrophic equation appearing in dynamic meteorology; see [Loeper 2006; Ambrosio et al. 2014]. Moreover, closely related evolution equations appear in cosmology and in particular in Brenier's approach to the Zeldovich model used in the early universe reconstruction problem [Shandarin and Zel'dovich 1989; Frisch et al. 2002; Brenier 2011; 2016].

As we were not able to deduce the type of propagation of chaos result we needed from previous general results and approaches, the main body of the paper establishes the appropriate propagation of chaos result, which, to the best of our knowledge, is new and hopefully the result, as well as the method of proof, is of independent interest. As will be clear below, our approach heavily relies on the theory of weak gradient flows on the Wasserstein L^2-space $\mathcal{P}_2(\mathbb{R}^n)$ of probability measure on $\mathbb{R}^n$ developed in the seminal work of Ambrosio, Gigli and Savaré [Ambrosio et al. 2005], which provides a rigorous framework for the Otto calculus [2001]. In particular, just as in [Ambrosio et al. 2005], convexity (or more generally λ-convexity) plays a prominent role. Our limiting evolution equation will appear as the gradient flow on $\mathcal{P}_2(\mathbb{R}^n)$ of a certain free-energy-type functional F. Interestingly, as observed in [Berman 2013c; Berman and Berndtsson 2013] the functional F may be identified with Mabuchi's K-energy functional on the space of Kähler metrics, which plays a key role in Kähler geometry. The gradient flows of F with respect to other metrics (the Mabuchi–Donaldson–Semmes metric and Calabi's gradient metric) are the renowned Calabi flow and Kähler–Ricci flow respectively [Chen and Zheng 2013]. The regularity and large-time properties of the evolution equation appearing here will be studied elsewhere [Berman and Lu $\geq$ 2018; Berman $\geq$ 2018].

In the remaining part of the introduction we will state our main results: first a general propagation of chaos result, assuming a uniform Lipschitz and convexity assumption on the interaction energy, and then the application to permanental point processes and toric Kähler–Einstein metrics. In the companion paper

[Berman and Önnheim 2016] we give a more general formulation of the propagation of chaos result, by relaxing some of the assumptions (in particular, this yields sharp convergence results for strongly singular repulsive pair interactions when $n = 1$).

1A. ***Propagation of chaos and Wasserstein gradient flows.*** Consider a system of N identical particles diffusing on the n-dimensional Euclidean space

$$X := \mathbb{R}^n$$

and interacting by a symmetric energy function $E^{(N)}(x_1, x_2, \dots, x_N)$, at a fixed inverse temperature β. According to nonequilibrium statistical mechanics, the distribution of particles at time t is described by the following system of stochastic differential equations (SDEs), under suitable regularity assumptions on $E^{(N)}$:

$$dx_i(t) = -\frac{\partial}{\partial x_i} E^{(N)}(x_1, x_2, \dots, x_N)\, dt + \sqrt{\frac{2}{\beta}}\, dB_i(t), \tag{1-1}$$

where B_i denotes N independent Brownian motions on $\mathbb{R}^n$; the equation is called the (overdamped) *Langevin equation* in the physics literature [Schwabl 2002, Section 8.1.2]. In other words, this is the Itô diffusion on $\mathbb{R}^n$ describing the (downward) gradient flow of the function $E^{(N)}$ on the configuration space X^N perturbed by a noise term. A classical problem in mathematical physics going back to Boltzmann and made precise by Kac [1956] is to show that, in the many-particle limit where $N \to \infty$, a *deterministic* macroscopic evolution emerges from the stochastic microscopic dynamics described by (1-1). More precisely, denoting by δ_N the empirical measures

$$\delta_N := \frac{1}{N} \sum \delta_{x_i}, \tag{1-2}$$

the SDEs (1-1) define a curve $\delta_N(t)$ of random measures on X. The problem is to show that, if at the initial time $t = 0$ the random variables x_i are independent with identical distribution μ_0, then the empirical measure $\delta_N(t)$ converges in law to a curve μ_t of measures on $\mathbb{R}^n$,

$$\lim_{N \to \infty} \delta_N(t) = \mu_t \tag{1-3}$$

at any time $t > 0$. In the terminology of [Kac 1956], see also [Sznitman 1991], this means that *propagation of chaos* holds. It should be stressed that the previous statement admits a pure PDE formulation, not involving any stochastic calculus (see Section 2C) and it is this analytic point of view that we will adopt here.[1]

Of course, if propagation of chaos is to hold then some consistency assumptions have to be made on the sequence $E^{(N)}$ of energy functions as N tends to infinity. The standard assumption in the literature ensuring that propagation of chaos does hold is that $E^{(N)}(x_1, x_2, \dots, x_N)$ can be as written as

$$E^{(N)}(x_1, x_2, \dots, x_N) = N E(\delta_N) \tag{1-4}$$

for a fixed functional E on the space of $\mathcal{P}(X)$ of all probability measures on X, where E is assumed to have appropriate regularity properties (to be detailed below). This is sometimes called a mean field model.

[1] From a differential geometric point of view the SDEs (1-1) correspond, under the transformation $\mu \mapsto e^{E/2}\mu$, to the heat flow on X^N of the Witten Laplacian of the "Morse function" E, but we will not elaborate on this point here.

By the results in [Braun and Hepp 1977; Sznitman 1991; Dawson and Gärtner 1987; Mischler et al. 2015], it then follows that the limit $\mu_t(=\rho_t\,dx)$ with initial data $\mu_0(=\rho_0\,dx)$ is uniquely determined and satisfies an explicit nonlinear evolution equation on $\mathcal{P}(X)$ of the form

$$\frac{d\rho_t}{dt}=\frac{1}{\beta}\Delta\rho_t-\nabla\cdot(\rho_t b[\rho_t]), \tag{1-5}$$

where we have identified $\mu(=\rho\,dx)$ with its density ρ and $b[\mu]$ is a function on $\mathcal{P}(X)$ taking values in the space of vector fields on X:

$$b[\mu]=-\nabla(dE_{|\mu}), \tag{1-6}$$

where the differential $dE_{|\mu}$ at μ is identified with a function on X, by standard duality (the alternative suggestive notation $b[\rho]=-\nabla(\partial E(\rho)/\partial\rho)$ is often used in the literature). In the kinetic theory literature drift-diffusion equations of the form (1-6) are usually called *McKean–Vlasov equation*s [McKean 1966; 1967]. More generally, the results referred to above hold in the more general setting where the gradient vector field $-(\partial/\partial x_i)E^{(N)}(x_1,x_2,\dots,x_N)$ on X appearing in (1-1) is replaced by $b[\delta_N]$ for a given function $b[\mu]$ on $\mathcal{P}(X)$, taking values in the space of vector fields and satisfying appropriate continuity properties.

One of the main aims of the present work is to introduce a new approach to the propagation of chaos result (1-3) for the stochastic dynamics (1-1) which exploits the gradient structure of the equations in question and which applies under weaker assumptions than the previous results, referred to above. As indicated above, our main motivation for weakening the assumptions comes from the applications to toric Kähler–Einstein metrics described below. In that case $E^{(N)}$ satisfies the following assumptions, which will be referred to as the *main assumptions*: $E^{(N)}$ is uniformly Lipschitz continuous in each variable separately, i.e., there is a constant C such that

(MA1) $$|\nabla_{x_i}E^{(N)}|\leq C, \tag{1-7}$$

and there exists a (finite) functional $E(\mu)$ on the Wasserstein space $\mathcal{P}(\mathbb{R}^n)$ such that

(MA2) $$\frac{1}{N}E^{(N)}(x_1,x_2,\dots,x_N)=E(\delta_N)+o(1), \tag{1-8}$$

where $o(1)$ denotes a sequence of functionals on $\mathcal{P}(\mathbb{R}^n)$ converging pointwise to zero on $\mathcal{P}(\mathbb{R}^n)$ as $N\to\infty$. Moreover, $E^{(N)}$ is λ-*convex* on X^N for some real number λ, which means that the (distributional) Hessians are uniformly bounded from below on $\mathbb{R}^{nN}$,

(MA3) $$(\nabla^2E^{(N)})\geq\lambda I, \tag{1-9}$$

where I denotes the identity matrix on $\mathbb{R}^{nN}$. This implies, in particular, that there exists a unique solution to the evolution equation (1-5) in the sense of weak gradient flows on the space $\mathcal{P}_2(\mathbb{R}^n)$ of all probability measures with finite second moments equipped with the Wasserstein L^2-metric [Ambrosio et al. 2005]:

$$\frac{d\mu_t}{dt}=-\nabla F_\beta(\mu_t),$$

where F_β is the free-energy-type functional corresponding to the macroscopic energy $E(\mu)$ at inverse temperature β,

$$F_\beta(\mu) = E(\mu) + \frac{1}{\beta} H(\mu),$$

and where $H(\mu)$ is the Boltzmann entropy of μ (see Section 2A for notation).

Theorem 1.1. *Let $E^{(N)}$ be a sequence of symmetric functions on $(\mathbb{R}^n)^N$ satisfying the main assumptions (1-8), (1-7) and (1-9) and consider the corresponding system of SDEs (1-1). If the initial data $x_i(0)$ consists of independent and identically distributed random vectors with law $\mu_0 \in \mathcal{P}_2(\mathbb{R}^n)$, then, at any fixed positive time t, the corresponding empirical measures converge in law, as $N \to \infty$, to the measure $\mu_t \in \mathcal{P}_2(\mathbb{R}^n)$, where the curve $t \mapsto \mu_t$ is the gradient flow on the Wasserstein space $\mathcal{P}_2(\mathbb{R}^n)$ of the free energy functional F_β, emanating from μ_0.*

In fact, the convergence of the laws will be shown to hold in the L^2-Wasserstein topology. This leads to a strong form of propagation of chaos in the present setting (implying that the correlations between the random vectors $x_i(t)$ and $x_j(t)$ tend to zero as $N \to \infty$, if $i \neq j$; see Section 3B).

It should be stressed that the key point of our approach is that we do not need to assume that the drift $b[\mu](x)$ defined by (1-6) has any continuity properties with respect to μ or x, in contrast to previous work [Braun and Hepp 1977; Sznitman 1991; Dawson and Gärtner 1987; Mischler et al. 2015]. This will be crucial in the applications to toric Kähler–Einstein metrics, where the interaction energies $E^{(N)}$ are smooth and convex, but the norms of corresponding Hessians are not uniformly bounded in N (which is reflected in the fact that the corresponding function $x \mapsto b(\mu)(x)$ is not continuous for a general μ).

We recall that if the drift is assumed to have suitable continuity properties, then the existence of a solution to the drift-diffusion equation (1-5) can be established using fixed-point-type arguments [Sznitman 1991]. However, in our case we have, in general, to resort to the weak gradient flow solutions provided by the general theory in [Ambrosio et al. 2005], where the solution ρ_t can be characterized uniquely by a differential inequality called the *evolutionary variational inequality* (*EVI*). As shown in that work, the corresponding solution ρ_t satisfies the drift-diffusion equation (1-5) in a suitable weak sense, as follows formally from the Otto calculus [Otto 2001].

1A1. *Idea of the proof of Theorem 1.1 and comparison with previous results.* The starting point of the proof is the basic fact that the SDEs (1-1) on X^N admit a PDE formulation. Indeed, as recalled in Section 2C, they correspond to a linear evolution $\mu_N(t)$ of probability measures (or densities) on X^N, given by the corresponding forward Kolmogorov equation (also called the Fokker–Planck equation). Given this fact, our proof of Theorem 1.1 proceeds in a variational manner, building on [Jordan et al. 1998; Ambrosio et al. 2005] (and inspired by the approach introduced in [Messer and Spohn 1982] in the static setting of Gibbs measure): the rough idea is to show that any weak limit curve $\Gamma(t)$ of the laws

$$\Gamma_N(t) := (\delta_N)_* \mu_N(t) \in \mathcal{P}_2(Y), \quad Y = \mathcal{P}_2(\mathbb{R}^n),$$

is of the form $\Gamma(t) := \delta_{\mu_t}$, where the curve μ_t in $\mathcal{P}_2(\mathbb{R}^n)$ is uniquely determined by a "dynamic minimizing property". To this end we first discretize time, by fixing a small time mesh $\tau := t_{j+1} - t_j$, and replace, for any fixed N, the curve $\Gamma_N(t)$ with its discretized version $\Gamma_N^\tau(t_j)$, defined by a variational Euler scheme (a "minimizing movement" in De Giorgi's terminology) as in [Jordan et al. 1998; Ambrosio et al. 2005]. We then establish a discretized version of Theorem 1.1 saying that if, at a given discrete time t_j, the convergence

$$\lim_{N\to\infty} \Gamma_{t_j}^N = \delta_{\mu_{t_j}^\tau}$$

holds in the L^2-Wasserstein metric then the convergence also holds at the next time step t_{j+1} (using a variational argument). In particular, since, by assumption, the convergence above holds at the initial time 0 it "propagates" by induction to hold at any later discrete time. Finally, we prove Theorem 1.1 by letting the mesh τ tend to zero. This last step uses that the error estimates established in [Ambrosio et al. 2005], for discretization schemes as above, only depend on a uniform lower bound λ on the convexity of the interaction energies.

Our proof appears to be rather different from the probabilistic approaches in [Sznitman 1991; Dawson and Gärtner 1987], which are based on a study of nonlinear martingales, and the recent PDE approach in [Mischler et al. 2015]. The latter approaches require a two-sided uniform bound on the Hessian of the interaction energy $E^{(N)}$, while we only require a uniform lower bound.

It may also be illuminating to think about the convergence of $\Gamma_N(t)$ towards $\Gamma(t)$ as a kind of stability result for the sequence of weak gradient flows on $\mathcal{P}_2(Y)$, associated to the corresponding mean free energies, viewed as functionals on $\mathcal{P}_2(Y)$. This situation is somewhat similar to the stability result for gradient flows on $\mathcal{P}_2(H)$ in [Ambrosio et al. 2005; 2009], where H is a Hilbert space, but the main difference here is that the underlying space Y is not a Hilbert space, as opposed to the setting in those works, which prevents one from directly applying the error estimates in [Ambrosio et al. 2005] on the space $\mathcal{P}_2(Y)$ itself; this analog is expanded on in the companion paper [Berman and Önnheim 2016].

1B. ***Applications to permanental point processes at negative temperature and toric Kähler–Einstein metrics.*** Let P be a convex body in $\mathbb{R}^n$ containing zero in its interior and denote by $P_{\mathbb{Z}}$ the lattice points in P, i.e., the intersection of the convex body P with the integer lattice $\mathbb{Z}^n$. We fix an auxiliary ordering $p_1, \dots, p_N$ of the N elements of $P_{\mathbb{Z}}$. Given a configuration $(x_1, \dots, x_N)$ of N points on X, we denote by $\operatorname{Per}(x_1, \dots, x_N)$ the number defined as the permanent of the rank-N matrix with entries $A_{ij} := e^{x_i\cdot p_j}$:

$$\operatorname{Per}(x_1, \dots, x_N) := \operatorname{Per}(e^{x_i\cdot p_j}) = \sum_{\sigma\in S_N} e^{x_1\cdot p_{\sigma(1)}+\cdots+x_N\cdot p_{\sigma(N)}}, \tag{1-10}$$

where S_N denotes the symmetric group on N letters. This defines a symmetric function on $\mathbb{R}^{nN}$ which is canonically attached to P (i.e., it is independent of the choice of ordering of $P_{\mathbb{Z}}$).[2] We will consider the large-N limit which appears when P is replaced by the sequence kP of scaled convex bodies for any positive integer k. In particular, N depends on k as

$$N_k = \frac{k^n V(P)}{n!} + o(k^n),$$

[2]In many body quantum mechanics $\operatorname{Per}(x_1, \dots, x_N)$ appears as the N-particle wave function for a bosonic system of N particles represented by the N wave functions $e^{x\cdot p_j}$, i.e., N planar waves with imaginary momenta proportional to p_j.

where $V(P)$ denotes the Euclidean volume of P. In this setting the interaction energy is defined by

$$E^{(N_k)}(x_1, \dots, x_{N_k}) = \frac{1}{k} \log \operatorname{Per}(x_1, \dots, x_{N_k}). \tag{1-11}$$

To simplify the notation we will often drop the explicit dependence of N on k.

By the results in [Berman 2013b], the assumptions in Theorem 1.1 hold with

$$E(\mu) := -C(\mu),$$

where $C(\mu)$ is the Monge–Kantorovich optimal cost for transporting μ to the uniform probability measure ν_P on the convex body P, with respect to the standard symmetric quadratic cost function $c(x, p) = -x \cdot p$. Hence, the corresponding free-energy functional may be written as

$$F_\beta(\mu) = -C(\mu) + \frac{1}{\beta} H(\mu). \tag{1-12}$$

Theorem 1.2. *Assume that $\beta > 0$ and consider the system of SDEs (1-1), defined by the interaction energy (1-11). If the initial data $x_i(0)$ consists of independent and identically distributed random vectors with law $\mu_0 \in \mathcal{P}_2(\mathbb{R}^n)$, then at any fixed positive time t, the corresponding empirical measures converge in law, as $N \to \infty$, to the measure $\mu_t \in \mathcal{P}_2(\mathbb{R}^n)$, where the curve $t \mapsto \mu_t$ is the gradient flow on the Wasserstein space $\mathcal{P}_2(\mathbb{R}^n)$ of the free energy functional F_β (1-12) emanating from μ_0. The corresponding densities ρ_t on $\mathbb{R}^n$ satisfy the following evolution PDE in the distributional sense:*

$$\frac{\partial \rho_t}{\partial t} = \frac{1}{\beta} \Delta \rho_t + \nabla \cdot (\rho_t \nabla \phi_t), \tag{1-13}$$

where $\phi_t(x)$ is the "convex potential" of ρ_t, i.e., the convex function on $\mathbb{R}^n$ solving the Monge–Ampère equation

$$\frac{1}{V(P)} \det(\partial^2 \phi_t) = \rho_t \tag{1-14}$$

(*in the weak sense of Alexandrov*) *normalized so that $\phi(0) = 0$ and satisfying the growth condition*

$$\phi(x) \le \sup_{p \in P} p \cdot x \tag{1-15}$$

(*equivalently, the closure of the gradient image of ϕ is equal to P*).

Let us briefly explain how Theorem 1.2 provides a stochastic dynamic approach for constructing Kähler–Einstein metrics on toric varieties; details will appear in a separate publication [Berman ≥ 2018]. We first recall that the Kähler potential of such a metric can be identified with a convex function ϕ satisfying the Monge–Ampère equation on $\mathbb{R}^n$

$$\det(\partial^2 \phi) = e^{-\phi}, \tag{1-16}$$

subject to the growth condition (1-15), with P being the moment polytope of the toric variety; see [Wang and Zhu 2004; Berman and Berndtsson 2013]. Now, a direct computation reveals that the corresponding density $\rho := \det(\partial^2 \phi)$ (which may be identified with the volume form of the Kähler–Einstein metric) is a stationary solution of the evolution equation appearing in Theorem 1.2. This is consistent, as it must be, with the fact that the free-energy functional F_β (1-12) may be identified with Mabuchi's K-energy functional on the space of Kähler metrics (when $\beta = 1$), whose minimizers are precisely the Kähler–Einstein metrics [Berman 2013c; Berman and Berndtsson 2013]. As shown in [Berman $\geq$ 2018], this fact can be used to show that the solution ρ_t of the evolution equation appearing in Theorem 1.2 converges, when $t \to \infty$, to the volume form of a Kähler–Einstein metric on X_P, if such a metric exists, which, in turn, is equivalent to the vanishing of the barycenter of the polytope P [Wang and Zhu 2004; Berman and Berndtsson 2013]. As discussed in Section 5A this can be viewed as a generalization of well-known stability properties for scalar conservation laws. The upshot of all this is that letting first N and then t tend to infinity in the SDEs (1-1), corresponding to the interaction energy (1-11), produces a toric Kähler–Einstein metric, when such a metric exists.

As we point out in Sections 4D, 5C1 our results also apply to the *tropical* analog of the permanental setting above, which can be viewed as the tropicalization of the complex geometric setting on the corresponding toric variety. In the corresponding deterministic setting (i.e., $\beta_N = \infty$) the particles then perform zigzag paths in $\mathbb{R}^n$ generalizing the extensively studied sticky particle system on $\mathbb{R}$ [Weinan et al. 1996; Brenier and Grenier 1998; Natile and Savaré 2009]. This is closely related to the Zeldovich model for the formation of large-scale structures in cosmology; see [Frisch et al. 2002; Brenier 2011; 2016] (compare to the discussion in Section 5B).

There is also a static analog of Theorem 1.2 (formulated in terms of Gibbs measures), which yields a probabilistic tropical analog of the Yau–Tian–Donaldson conjecture on toric Fano varieties linking the existence problem for toric Kähler–Einstein metrics to a notion of stability. This result first appeared in a previous preprint version of the present paper on the arXiv, but it has been deferred to a separate publication to shorten the present paper.

1C. *Generalizations of Theorem 1.1.* Let us conclude this introduction by pointing out that Theorem 1.1 admits various generalizations, obtained by weakening the assumptions, which are developed in the companion paper [Berman and Önnheim 2016]:

- By rescaling $E^{(N)}$ we may as well allow the "inverse temperature" β appearing in the SDEs (1-1) to depend on N as long as
$$\beta_N \to \beta \in [0, \infty],$$
as $N \to \infty$. In particular, Theorem 1.1 also applies to $\beta = \infty$ where the evolution equation (1-5) becomes a pure transport equation (i.e., with no diffusion). However, the precise relation to weak solutions becomes much more subtle and is closely related to the notions of entropy solutions and viscosity solutions studied in the PDE literature [Lax 1973], as detailed in [Berman and Önnheim 2016]. In fact, one may even allow that $\beta_N = \infty$, where the corresponding convergence results yields a deterministic mean field particle approximation.

- The assumptions (MA1) and (MA2), formulas (1-7), (1-8), may be replaced by a uniform coercivity assumption on $E^{(N)}$ together with the assumption that the *mean energies* corresponding to $E^{(N)}$ converge as functionals on $\mathcal{P}_2(\mathcal{P}_2(\mathbb{R}))$, in a suitable sense (which is closely related to the notion of Γ-convergence). This ensures that a weaker form of Proposition 3.6 holds.

• The convexity assumption (MA3) on $E^{(N)}$ may be replaced by a generalized convexity property of the corresponding mean energy functional on $\mathcal{P}_2(\mathbb{R}^{nN})^{S_N}$.

1D. *Organization.* In Section 2 we start by recalling the general setup that we will need from probability, the theory of Wasserstein spaces and weak gradient flows and then turn to the proof of Theorem 1.1 in Section 3B (starting with the discretized situation). In Section 4 we go on to apply the previous general results to the permanental setting and its tropical analog. In the final section we provide on outlook on some relations to conservation laws, sticky-particle-type systems and complex geometry. The Appendix recalls the basics of the formal Otto calculus and is included to serve as a motivation for the material on Wasserstein gradient flows. The rather lengthy setup and preparatory material in Section 2 are due to our effort to make the paper readable to a rather general audience.

2. General setup and preliminaries

2A. *Notation.* Given a topological (Polish) space Y we will denote the integration pairing between measures μ on Y (always assumed to be Borel measures) and bounded continuous functions f by

$$\langle f, \mu\rangle := \int f\mu$$

(we will avoid the use of the symbol $d\mu$ since d will usually refer to a distance function on Y). In the case $Y = \mathbb{R}^D$ we will say that a measure μ *has a density,* denoted by ρ, if μ is absolutely continuous with respect to Lebesgue measure dx and $\mu = \rho\, dx$. We will denote by $\mathcal{P}(\mathbb{R}^D)$ the space of all probability measures and by $\mathcal{P}_{\mathrm{ac}}(\mathbb{R}^D)$ the subspace containing those with a density. The *Boltzmann entropy* $H(\rho)$ and *Fisher information* $I(\rho)$ (taking values in $]-\infty, \infty]$) are defined by

$$H(\rho) := \int_{\mathbb{R}^D} (\log\rho)\rho\, dx, \qquad I(\rho) = \int_{\mathbb{R}^D} \frac{|\nabla\rho|^2}{\rho}\, dx \tag{2-1}$$

(assuming that $\nabla\rho \in L^1(dx)$ and $\rho^{-1}\nabla\rho \in L^2(\rho dx)$). More generally, given a reference measure μ_0 on Y the entropy of a measure μ relative to μ_0 is defined by

$$H_{\mu_0}(\mu) = \int_{X^N} \left(\log\frac{\mu}{\mu_0}\right)\mu \tag{2-2}$$

if the probability measure μ on X is absolutely continuous with respect to μ_0 and otherwise $H(\mu) := \infty$. The relative Fisher information is defined similarly, by replacing ρ with the density μ/μ_0 in formula (2-1).

Given a lower semicontinuous (*lsc*, for short) function V on Y and $\beta \in]0, \infty]$ (the "inverse temperature") we will denote by F_β^V the corresponding (*Gibbs*) *free-energy functional with potential* V:

$$F_\beta^V(\mu) := \int_X V\mu + \frac{1}{\beta}H_{\mu_0}(\mu), \tag{2-3}$$

which coincides with $1/\beta$ times the entropy of μ relative to $e^{-V}\mu_0$.

2B. *Wasserstein spaces and metrics.* We start with the following very general setup. Let (X, d) be a given metric space, which is Polish, i.e., separable and complete, and denote by $\mathcal{P}(X)$ the space of all probability measures on X endowed with the *weak topology,* i.e., $\mu_j \to \mu$ weakly in $\mathcal{P}(X)$ if and only if $\int_X \mu_j f \to \int_X \mu f$ for any bounded continuous function f on X (this is also called the *narrow topology* in the probability literature). The metric d on X induces l^p-type metrics on the N-fold product X^N for any given $p \in [1, \infty[$:

$$d_p(x_1, \dots, x_N; y_1, \dots, y_N) := \left(\sum_{i=1}^N d(x_i, y_i)^p\right)^{1/p}.$$

The permutation group S_N on N letters has a standard action on X^N, defined by $(\sigma, (x_1, \dots, x_N)) \mapsto (x_{\sigma(1)}, \dots, x_{\sigma(N)})$ and we will denote by $X^{(N)}$ and π the corresponding quotient and quotient projection, respectively:

$$X^{(N)} := X^N/S^N, \quad \pi : X^N \to X^{(N)}. \tag{2-4}$$

The quotient $X^{(N)}$ may be naturally identified with the space of all configurations of N points on X. We will denote by $d_{(p)}$ the induced distance function on $X^{(N)}$, suitably normalized:

$$d_{(p)}(x_1, \dots, x_N; y_1, \dots, y_N) := \inf_{\sigma \in S_N} \left(\frac{1}{N}\sum_{i=1}^N d(x_i, y_{\sigma(i)})^p\right)^{1/p}.$$

The normalization factor $1/N^{1/p}$ ensures that the standard embedding of $X^{(N)}$ into the space $\mathcal{P}(X)$ of all probability measures on X,

$$X^{(N)} \hookrightarrow \mathcal{P}(X), \quad (x_1, \dots, x_N) \mapsto \delta_N := \frac{1}{N}\sum \delta_{x_i} \tag{2-5}$$

(where we will call δ_N the *empirical measure*), is an isometry when $\mathcal{P}(X)$ is equipped with the *L^p-Wasserstein metric d_{W^p}* induced by d (for simplicity we will also write $d_{W_p} = d_p$),

$$d^p_{W_p}(\mu, \nu) := \inf_\gamma \int_{X \times X} d(x, y)^p \gamma, \tag{2-6}$$

where γ ranges over all couplings between μ and ν; i.e., γ is a probability measure on $X \times X$ whose first and second marginals are equal to μ and ν respectively (see Lemma 2.3 below). We will denote by $W^p(X, d)$ the corresponding *L^p-Wasserstein space*, i.e., the subspace of $\mathcal{P}(X)$ consisting of all μ with finite p-th moments: for some (and hence any) $x_0 \in X$

$$\int_X d(x, x_0)^p \mu < \infty.$$

We will also write $W^p(X, d) = \mathcal{P}_p(X)$ when it is clear from the context which distance d on X is used.

Remark 2.1. In the terms of the Monge–Kantorovich theory of optimal transport [Villani 2003], $d^p_{W_p}(\mu, \nu)$ is the optimal cost for transporting μ to ν with respect to the cost functional $c(x, p) := d(x, y)^p$. Accordingly a coupling γ as above is often called a *transport plan* between μ and ν and it is said to be a *transport map T* if $\gamma = (I \times T)_*\mu$, where $T_*\mu = \nu$. In particular, if $X = \mathbb{R}^n$, $p = 2$ and μ and ν are absolutely continuous with respect to Lebesgue measure, then, by Brenier's theorem [1991], the optimal transport plan γ is always defined by a transport map $T(:= T_\mu^\nu)$ of the form $T_\mu^\nu = \nabla\phi$, where ϕ is a convex function on $\mathbb{R}^n$ (optimizing the dual Kantorovich functional).

We recall the following standard proposition:

Proposition 2.2. *A sequence μ_j converges to μ in the distance topology in $W^p(X, d)$ if and only if μ_j converges to μ weakly in $\mathcal{P}(X)$ and the p-th moments converge. As a consequence, if μ_j converges to μ weakly in $\mathcal{P}(X)$ and the p-th moments are uniformly bounded, i.e., for some $x_0 \in X$ there exists a constant C_0 such that*

$$\int_X d(x, x_0)^p \mu_j \le C_0,$$

then μ_j converges to μ in the distance topology in $W^{p'}(X, d)$ for any $p' < p$.

Proof. For the first statement see for example [Villani 2003, Theorem 7.12]. The second statement is certainly also well known, but for completeness we include a simple proof. Fix $x_0 \in X$ and take the decomposition

$$\int_X d(x, x_0)^{p'} \mu_j = \int_{\{d(x,x_0)\le R\}} d(x, x_0)^{p'} \mu_j + \int_{\{d(x,x_0)>R\}} d(x, x_0)^{p'} \mu_j .$$

Since $d(x, x_0)^{p'} \le d(x, x_0)^p / R^{(p-p')}$ when $d(x, x_0) \ge R$, the second integral above is bounded from above by $C_0/R^{(p-p')}$. Moreover, by the assumption of weak convergence, the first term above converges to $\int_{\{d(x,x_0)\le R\}} d(x, x_0)^{p'} \mu$, as $j \to \infty$. Finally, letting R tend to infinity concludes the proof. □

Since $Y_p := (W_p(X), d_{W_p})(:= \mathcal{P}_p(X))$ is also a Polish space we can iterate the previous construction and consider the Wasserstein space $W_q(Y) \subset \mathcal{P}(\mathcal{P}(X))$ that we will write as $W_q(\mathcal{P}_p(X))$, which is thus the space of all probability measures Γ on $\mathcal{P}(X)$ such that, for some $\mu_0 \in W_p(X)$,

$$\int_{\mathcal{P}(X)} d_p(\mu, \mu_0)^q \, \Gamma < \infty.$$

Lemma 2.3 (three isometries).

- *The empirical measure δ_N defines an isometric embedding $(X^{(N)}, d_{(p)}) \to \mathcal{P}_p(X)$.*
- *The corresponding push-forward map $(\delta_N)_*$ from $\mathcal{P}(X^{(N)})$ to $\mathcal{P}(\mathcal{P}(X))$ induces an isometric embedding between the corresponding Wasserstein spaces $W_q(X^{(N)}, d_{(p)})$ and $W_q(\mathcal{P}_p(X))$.*
- *The push-forward π_* of the quotient projection $\pi : X^N \to X^{(N)}$ induces an isometry between the subspace of symmetric measures in $W_q(X^N, (1/N^{1/p})d_p)$ and the space $W_q(X^{(N)}, d_{(p)})$.*

Proof. The first statement is a well-known consequence of the Birkhoff–Von Neumann theorem which gives that for any symmetric function $c(x, y)$ on $X \times X$ we have that if $\mu = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$ and $\nu = \frac{1}{N}\sum_{i=1}^N \delta_{y_i}$ for given $(x_1, \dots, x_N), (y_1, \dots, y_N) \in X^N$, then

$$\inf_{\Gamma(\mu,\nu)} \int c(x, y)\, d\Gamma = \inf_{\Gamma_N(\mu,\nu)} \int c(x, y)\, d\Gamma,$$

where $\Gamma_N(\mu,\nu)\subset\Gamma(\mu,\nu)$ consists of couplings of the form $\Gamma_\sigma := \frac{1}{N}\sum \delta_{x_i}\otimes\delta_{y_{\sigma(i)}}$ for $\sigma\in S_N$, where S_N is the symmetric group on N letters. The second statement then follows from the following general fact: if $f:(Y_1,d_1)\to(Y_2,d_2)$ is an isometry between two metric spaces, then f_* gives an isometry between $W_q(Y_1,d_1)$ and $W_q(Y_2,d_2)$. This follows immediately from the definitions once one observes that one may assume that the coupling γ_2 between $f_*\mu$ and $f_*\nu$ is of the form $f_*\gamma_1$ for some coupling γ_1 between μ and ν. The point is that γ can be taken to be concentrated on $f(Y_1)\times f(Y_2)$ (since this set contains the product of the supports of μ and ν) and hence one can take $\gamma_1 := (f^{-1}\otimes f^{-1})_*\gamma_2$, where $(f^{-1}\otimes f^{-1})(f(y),f(y')) := (y,y')$ is well-defined, since f induces a bijection between Y_1 and $f(Y_1)$. Finally, the last statement follows immediately from the following general claim applied to $Y=X^N$ with $d=(1/N^{1/p})d_{X^N,l^p}$ and $G=S_N$. Let G be a compact group acting by isometries on a metric space (Y,d) and consider the natural projection $\pi:Y\to Y/G$. We denote by d_G the induced quotient metric on Y/G. The push-forward π_* gives a bijection between the space $\mathcal{P}(X)^G$ or all G-invariant probability measures on X and $\mathcal{P}(X/G)$. The claim is that π_* induces an isometry between the corresponding Wasserstein spaces $\mathcal{P}_q(X)^G$ and $\mathcal{P}_q(X/G)$; i.e., $d_{W_q}(\mu,\nu)=d_{W_q}(\pi_*\mu,\pi_*\nu)$ if μ and ν are G-invariant; see [Lott and Villani 2009, Lemma 5.36]. □

Let us also recall the following classical result

Lemma 2.4. *Let μ_0 be a probability measure on X. Then $(\delta_N)_*\mu_0^{\otimes N}\to\delta_{\mu_0}$ in $\mathcal{P}(\mathcal{P}(X))$ weakly as $N\to\infty$.*

In fact, according to Sanov's classical theorem the previous convergence result even holds in the sense of large deviations at speed N with rate functional given by the relative entropy functional $H_{\mu_0}(\cdot)$ [Dembo and Zeitouni 1993, Theorem 6.2.10].

2B1. *The present setting.* We will apply the previous setup to $X=\mathbb{R}^n$ endowed with the Euclidean metric d. Moreover, we will mainly use the case $p=2$. Then the corresponding metric d_2 on X^N is the Euclidean metric on $X^N=\mathbb{R}^{nN}$. Identifying a symmetric (i.e., S_N-invariant) probability measure μ_N on X^N with a probability measures on the quotient $X^{(N)}$ (as in Lemma 2.3) the second and third points in Lemma 2.3 may (with $q=2$) be summarized by the following chain of equalities that will be used repeatedly below:

$$\frac{1}{N}d_2(\mu_N,\mu_N')^2 = d_{(2)}(\mu_N,\mu_N')^2 = d_{W_2(\mathcal{P}_2(\mathbb{R}^n))}(\Gamma_N,\Gamma_N')^2, \tag{2-7}$$

where Γ_N and Γ_N' denote the push-forwards under δ_N of μ_N and μ_N' respectively. To simplify the notation we will often simply write

$$d := d_{W_2(\mathcal{P}_2(\mathbb{R}^n))}$$

for the metric on $W_2(\mathcal{P}_2(\mathbb{R}^n))$ (or sometimes $d=d_2$).

2C. *The forward Kolmogorov equation for the SDEs and the mean free energy $F_{N,\beta}$.* Fix a positive integer N and $\beta>0$ (which may depend on N when we will later on let $N\to\infty$). Let (X,g) be a Riemannian manifold and denote by dV the volume form defined by g. In our case (X,g) will be the Euclidean space $\mathbb{R}^n$.

Consider the SDEs (1-1) on X^N with the initial condition that $x_i(0)$ are independent random variables with identical distribution $\mu_0 \in \mathcal{P}(\mathbb{R}^D)$. As is well known, under suitable regularity assumptions, this defines, for any fixed T, a probability measure η_T on the space of all continuous curves ("sample paths") in X^N, i.e., the space of continuous maps $[0, T] \to X^N$ [Stroock and Varadhan 1997, Chapter 5]. For t fixed we can thus view $x^{(N)}(t)$ as an X^N-valued random variable on the latter probability space. Then its law

$$\mu_t^{(N)} := (x^{(N)}(t))_* \eta_t$$

gives a curve of probability measures on $X^{(N)}$ of the form $\mu_t^{(N)} = \rho_t^{(N)} dV^{\otimes N}$, where the density $\rho_t^{(N)}$ satisfies the corresponding forward Kolmogorov equation

$$\frac{\partial \rho_t^{(N)}}{\partial t} = \frac{1}{\beta} \Delta \rho_t^{(N)} + \nabla \cdot (\rho_t^{(N)} \nabla E^{(N)}), \tag{2-8}$$

which thus coincides with the linear Fokker–Planck equation (A-6) on X^N with potential $V := E^{(N)}$. In this formula the initial conditions for the SDEs translates to

$$\mu^{(N)}_{|t=0} = \mu_0^{\otimes N}. \tag{2-9}$$

In particular, the law of the empirical measures $\delta_N(t)$ for the SDEs (1-1) can be written as the following probability measure on $\mathcal{P}(X)$:

$$\Gamma_N(t) := (\delta_N)_* \mu_t^{(N)},$$

where δ_N is the empirical measure defined by (2-5).

Anyway, for our purposes we may as well forget about the SDEs (1-1) and take the forward Kolmogorov equation (2-8) on X^N as our the starting point, together with the initial condition (2-9). We will exploit the well-known fact, going back to [Jordan et al. 1998] (see Theorem 2.14 below) that the latter evolution equation can be interpreted as the gradient flow on the Wasserstein space $\mathcal{P}_2(X^N)$ of the functional

$$F_\beta^{(N)}(\mu_N) = \int_{X^N} E^{(N)} \mu_N + \frac{1}{\beta} H(\mu_N),$$

where $H(\cdot)$ is the entropy relative to $\mu_0 := dV^{\otimes N}$, formula (2-2); occasionally we will omit the subscript β in the notation $F_\beta^{(N)}$.

Following standard terminology in statistical mechanics we will call the scaled functional $F_{N,\beta} := F_\beta^{(N)}/N$ the *mean free energy*, which is thus a sum of the *mean energy* $E_N (:= F_{N,\infty})$ and the *mean entropy* $H_N(\mu_N)$:

$$F_{N,\beta} = E_N + \frac{1}{\beta} H_N;$$

i.e.,

$$F_{N,\beta}(\mu_N) := \frac{1}{N} F_\beta^{(N)}(\mu_N) = \frac{1}{N} \int_{X^N} E^{(N)} \mu_N + \frac{1}{\beta N} H(\mu_N), \tag{2-10}$$

Note that it follows immediately from the definition that the mean entropy is additive: for any $\mu \in \mathcal{P}(X)$

$$H_N(\mu^{\otimes N}) = H(\mu).$$

In the case dV is a probability measure, it follows immediately from Jensen's inequality that $H(\mu) \geq 0$. In our Euclidean setting this is not the case but using that $\int e^{-\epsilon|x|^2}\,dx < \infty$ for any given $\epsilon > 0$ one then gets

$$H(\mu) \geq -\epsilon \int |x|^2 \mu - C_\epsilon. \tag{2-11}$$

As a consequence we have the following:

Lemma 2.5. *If the mean energy satisfies the uniform coercivity property*

$$\frac{1}{N} \int_{X^N} E^{(N)}(\mu_N) \geq -\frac{1}{2\tau_*} d_2(\mu_N, \Gamma_*)^2 - C \tag{2-12}$$

for some fixed $\tau_* > 0$ *and* $\Gamma_* \in W_2(\mathcal{P}(X))$ *and positive constant* C, *then so does* $F^{(N)}/N$.

Remark 2.6. The linear forward Kolmogorov equation (2-8) can also be viewed as the gradient flow of the *mean* free energy $\frac{1}{N}F^{(N)}$ if one instead uses the scaled metric $g_N := \frac{1}{N} g^{\otimes N}$ on X^N. Moreover, in our case, $E^{(N)}$ will be symmetric, i.e., S_N-invariant, and hence the flow defined with respect to (X^N, g_N) descends to the flow defined with respect to $X^{(N)} := X^N/S_N$ equipped with the distance function $d_{X^{(N)}}$ defined in Section 2B. Using the isometric embedding defined by the empirical measure (Lemma 2.3) we can thus view the sequence of flows on the sequence of spaces $\mathcal{P}(X^N)$ as a sequence of flows on the same (infinite-dimensional) space $W_2(\mathcal{P}(X))$ and this is the geometric motivation for the proof of Theorem 1.1.

2D. *Propagation of chaos and the L^2-Wasserstein topology.* First recall [Sznitman 1991] that a sequence $\mu^{(N)}$ of symmetric probability measures on X^N is said to be *chaotic* if there exists a probability measure μ on X such that, for any given finite number of functions $f_1, \dots, f_k$ in $C_b(X)$,

$$\lim_{N\to\infty} \int_{X^N} f_1(x_1) \cdots f_1(x_k) \mu^{(N)} = \int_X f_1 \mu \cdots \int_X f_k \mu \tag{2-13}$$

(more precisely, then $\mu^{(N)}$ is called μ-*chaotic*).

Equivalently [Sznitman 1991, Proposition 2.2], this means that the empirical measure δ_N on the probability space $(X^N, \mu^{(N)})$ converges in law towards μ, i.e., the following convergence holds with respect to the weak topology in $\mathcal{P}(\mathcal{P}(X))$:

$$\lim_{N\to\infty} (\delta_N)_* \mu^{(N)} = \delta_\mu.$$

Now consider the system of SDEs (1-1) and assume that the initial random variables $x_1(0), \dots, x_N(0)$ are independent with identical law μ_0. This means that the corresponding curve of probability measures $\mu^{(N)}(t)$ on X^N (evolving by the forward Kolmogorov equation corresponding to the SDEs) is given by $\mu_0^{\otimes N}$ when $t = 0$ (i.e., the initial condition (2-9) holds). In particular, $\mu^{(N)}(t)$ is μ_0-chaotic when $t = 0$ (by Lemma 2.4). In the terminology introduced by Kac, *propagation of chaos* is said to hold if the sequence $\mu^{(N)}(t)$ remains chaotic for any positive time t, i.e., if there exists a curve $\mu(t)$ in $\mathcal{P}(X)$ emanating from μ_0 such that the sequence $\mu^{(N)}(t)$ is $\mu(t)$-chaotic for any $t \geq 0$.

In the present setting of Theorem 1.1 we will establish propagation of chaos in a stronger sense. Namely, we will show that if $\mu_0 \in \mathcal{P}_2(\mathbb{R}^n)$, then $(\delta_N)_* \mu^{(N)}(t)$ converges to $\delta_{\mu(t)}$ in $\mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^n))$, with respect to the topology defined by the Wasserstein L^2-metric. This is stronger than propagation of chaos since it also implies that the correlations between the random variables x_1 and x_2 on $((\mathbb{R}^n)^N, \mu_N)$ tend to zero as $N \to \infty$ (by symmetry this equivalently means that the correlations between x_i and x_j tend to zero, if $i \neq j$). This is made precise by the following lemma, where $(x)_\alpha$ denotes the α-th component of a vector $x \in \mathbb{R}^n$:

Lemma 2.7. *Let $\mu^{(N)}$ be a sequence of symmetric probability measures on $(\mathbb{R}^n)^N$ such that $(\delta_N)_*\mu^{(N)}(t)$ converges to δ_μ in $\mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^n))$, with respect to the topology defined by the Wasserstein L^2-metric. Then $\mu^{(N)}$ is μ-chaotic and moreover, for any given $(\alpha_1, \alpha_2) \in \{1, \dots, n\}^2$*

$$\lim_{N\to\infty} \bigl(\mathbb{E}_N((x_1)_{\alpha_1}(x_2)_{\alpha_2}) - \mathbb{E}_N((x_1)_{\alpha_1})\mathbb{E}_N((x_2)_{\alpha_2})\bigr) = 0,$$

where $\mathbb{E}_N$ denotes the expectation with respect to $\mu^{(N)}$.

Proof. This follows readily from the definitions, but for completeness we provide a proof. By assumption the probability measures $\Gamma_N := (\delta_N)_*\mu^{(N)}$ converge to δ_μ in $\mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^n))$ in the Wasserstein L^2-metric. Using Proposition 2.2 this convergence is equivalent to having

$$\lim_{N\to\infty} \int \Gamma_N \Phi = \Phi(\mu) \tag{2-14}$$

for any continuous function Φ on $\mathcal{P}_2(\mathbb{R}^n)$ of subquadratic growth, i.e., $\Phi(\nu) \le C_0 d^2_{W_2}(\nu, \nu_0) + C_0$ for a fixed element $\nu_0 \in \mathcal{P}_2(\mathbb{R}^n)$. Taking $\nu_0 = \delta_0$, the latter growth condition means that

$$\Phi(\nu) \le C \int |x|^2 \nu + C \tag{2-15}$$

for some constant C. In particular, setting $\Phi(\nu) := \int f_1\nu \cdots \int f_k\nu$ for given bounded continuous functions $f_1, \dots, f_k$ and expanding reveals that (2-13) holds, showing that propagation of chaos holds. At this point we have only used the convergence of Γ_N towards δ_μ in the weak topology, just as in the proof of one direction of [Sznitman 1991, Proposition 2.2]. But taking $\Phi(\nu) = \int (x)_{\alpha_1}\nu \int (x)_{\alpha_2}\nu$ (which satisfies (2-15), using Hölder's inequality) gives

$$\begin{aligned}\int \Gamma_N \Phi &= N^{-2} \sum_{i,j\le N} \int (x_i)_{\alpha_1}(x_j)_{\alpha_2}\mu^{(N)} \\ &= N^{-2}(N^2 - N) \int (x_1)_{\alpha_1}(x_2)_{\alpha_2}\mu^{(N)} + N^{-1} \int N^{-1} \sum_{i\le N} (x_i)_{\alpha_1}(x_i)_{\alpha_2}\mu^{(N)}.\end{aligned}$$

Hence, letting $N \to \infty$ and using the convergence in (2-14) gives

$$\lim_{N\to\infty} \int (x_1)_{\alpha_1}(x_2)_{\alpha_2}\mu^{(N)} + 0 = \int (x)_{\alpha_1}\mu \int (x)_{\alpha_2}\mu.$$

Finally, applying the convergence in (2-14) to $\Phi(\nu) = \int (x)_\alpha \nu$ and using symmetry reveals that the right-hand side above is equal to the limit of $\mathbb{E}_N((x_1)_{\alpha_1})$ times $\mathbb{E}_N((x_2)_{\alpha_2})$ as $N \to \infty$. $\square$

2E. *Gradient flows on the L^2-Wasserstein space and variational discretizations.* In this section we will recall the fundamental results from [Ambrosio et al. 2005] that we will rely on. Let G be a lower semicontinuous function on a complete metric space (M, d). In this generality there are, as explained in that work, various notions of weak gradient flows u_t for G (or "steepest descents") emanating from an initial point u_0 in M, symbolically written as

$$\frac{du_t}{dt} = -\nabla G(u_t), \quad \lim_{t\to 0} u(t) = u_0. \tag{2-16}$$

The strongest forms of weak gradient flows on metric spaces discussed in [Ambrosio et al. 2005] concern λ-convex functionals G and are defined by the property that u_t satisfies the *evolution variational inequalities* (*EVI*)

$$\frac{1}{2}\frac{d}{dt}d^2(u_t, v) + G(u(t)) + \frac{\lambda}{2}d^2(\mu_t, \nu)^2 \le G(v) \quad \text{a.e. } t > 0, \qquad \text{for all } v \in M, \quad G(v) < \infty, \tag{2-17}$$

together with the initial condition $\lim_{t\to 0} u(t) = u_0$ in (M, d). Then u_t is uniquely determined by u_0, as shown in [Ambrosio et al. 2005, Corollary 4.3.3], and we shall say that u_t is the *EVI-gradient flow* of G emanating from u_0. We recall that λ-convexity on a metric space essentially means that the distributional second derivatives are bounded from below by λ along any geodesic segment in M (compare to below). When M has nonpositive curvature, NPC, (in the sense of Alexandrov) the existence of a solution u_t satisfying the EVI was shown by Mayer [1998] for any lower-semicontinuous λ-convex functional, by mimicking the Crandall–Liggett technique in the Hilbert-space setting.

However, in our case (M, d) will be the L^2-Wasserstein space $\mathcal{P}_2(\mathbb{R}^d)$ for the space of all probability measures μ on $\mathbb{R}^d$, which does not have nonpositive curvature (when $d > 1$). Still, as shown in [Ambrosio et al. 2005], the analog of Meyer's result does hold under the stronger assumption that G be λ-convex along any *generalized* geodesic μ_s in $\mathcal{P}_2(\mathbb{R}^d)$. For our purposes it will be enough to consider lsc λ-convex functionals G with the property that $\mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$ is weakly dense in $\{G < \infty\}$. Then the λ-convexity of G means, compare to [Ambrosio et al. 2005, Proposition 9.210], that for any generalized geodesic $\mu_s = \rho_s\, dx$ in $\mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$ the function $G(\rho_s)$ is continuous on $[0, 1]$ and the distributional second derivatives on $]0, 1[$ satisfy

$$\frac{d^2 G(\rho_s)}{d^2 s} \ge \lambda.$$

We recall that a *generalized geodesic* μ_s connecting μ_0 and μ_1 in $\mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$ is determined by specifying a "base measure" $\nu \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$. Then μ_s is defined as the following family of push-forwards:

$$\mu_s = ((1-s)T_0 + sT_1)_* \nu,$$

where T_i is the optimal transport map (defined with respect to the cost function $|x-y|^2/2$) pushing forward ν to μ_i (compare to Remark 2.1).

Remark 2.8. The bona fide Wasserstein geodesics in $\mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$ are obtained by taking $\nu = \mu_0$ (the study of convexity along such geodesics was introduced by McCann [1997], who called it displacement convexity). But as shown in [Ambrosio et al. 2005], the point of working with general base measures ν is that they can be adapted to the discrete variational scheme for constructing EVI-gradient flows by taking $\nu = \mu_{t_j}$ at the j-th time step (compare to Section 2E1).

We will be relying on the following version of Theorems 4.0.4 and 11.2.1 in [Ambrosio et al. 2005]:

Theorem 2.9. *Suppose that G is an lsc real-valued functional on $\mathcal{P}_2(\mathbb{R}^d)$ which is λ-convex along generalized geodesics and satisfies the following coercivity property: there exist constants $\tau_*, C > 0$ and $\mu_* \in \mathcal{P}_2(\mathbb{R}^d)$ such that*

$$G(\cdot) \geq -\frac{1}{\tau_*} d_2(\cdot, \mu_*)^2 - C. \tag{2-18}$$

Then there is a unique solution μ_t to the EVI-gradient flow of G, emanating from any given $\mu_0 \in \overline{\{G < \infty\}}$. The flow has the following regularizing effect: $\mu_t \in \{|\partial G| < \infty\} \subset \{G < \infty\}$. Moreover, $G(\mu_t)$ and $e^{\lambda t}|\partial G|^2(\mu_t)$ are decreasing, where $|\partial G|$ denotes the metric slope of G:

$$|\partial G|(\mu) := \limsup_{\nu \to \mu} \frac{(G(\nu) - G(\mu))^+}{d(\mu, \nu)}.$$

Remark 2.10. Many more properties of the EVI-gradient flow μ_t are established in [Ambrosio et al. 2005]. For example, μ_t defines an absolutely continuous curve $\mathbb{R} \to \mathcal{P}_2(\mathbb{R}^n)$ (in the sense of metric spaces) which is locally Lipschitz continuous on $]0, \infty[$, which is a λ-contracting semigroup. Moreover, the flows are stable under suitable approximation of the initial data and the functional G.

Under suitably regularity assumptions it shown in [Ambrosio et al. 2005] that the EVI-gradient flow $\mu_t = \rho_t \, dx$ furnished by the previous theorem satisfies Otto's evolution equation (recalled in the Appendix) in the weak sense:

Proposition 2.11. *Suppose in addition to the assumptions in the previous theorem that μ_t has a density ρ_t for $t > 0$. Then ρ_t satisfies the continuity equation (A-5) in the sense of distributions on $\mathbb{R}^d \times \mathbb{R}$ with*

$$v_t = -(\partial^0 G)(\rho_t \, dx),$$

where $\partial^0 G$ denotes the minimal subdifferential of G.

We recall that under the assumptions in the previous theorem (and assuming $\{|\partial G|^2 < \infty\} \subset \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^n)$) the many-valued *subdifferential* ∂G on the subspace $\mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^n)$ is a metric generalization of the (Fréchet) subdifferential Hilbert space theory; by definition, it satisfies a "slope inequality along geodesics":

$$(\partial G)(\mu) := \left\{ \xi \in L^2(\mu) : G(\nu) \geq G(\mu) + \langle \xi, T_\mu^\nu(x) - x \rangle_{L^2(\mu)} + \tfrac{1}{2}\lambda d_2(\nu, \mu)^2 \text{ for all } \nu \right\},$$

where T_μ^ν denotes the optimal transport map between μ and ν, as in Remark 2.1. The *minimal subdifferential* $\partial^0 G$ on $\mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^n)$ at μ is defined as the unique element in the subdifferential ∂G at μ minimizing the L^2-norm in $L^2(\mu)$; in fact, its norm coincides with the metric slope of G at μ. In [Ambrosio et al. 2005] there is also a more general notion of extended subdifferential which, however, will not be needed for our purposes.

Example 2.12. In the case when $G = H$ is the Boltzmann entropy and μ satisfies $H(\mu) < \infty$, so that μ has a density ρ, we have $(\partial^0 H)(\mu) = \rho^{-1}\nabla\rho \in L^2(\mu)$ and hence

$$|\partial H|^2(\mu) = I(\rho)$$

is the Fisher information of ρ (2-1); see [Ambrosio et al. 2005, Theorem 10.4.17].

The following result goes back to [McCann 1997]; see also [Ambrosio et al. 2005] for various elaborations:

Lemma 2.13. *The following functionals are lsc and λ-convex along any generalized geodesics in $\mathcal{P}_2(\mathbb{R}^d)$:*

- *The "potential energy" functional $\mathcal{V}(\mu) := \int V\mu$, defined by a given lsc λ-convex and lsc function V on $\mathbb{R}^d$ (and the converse also holds).*
- *The functional $\mu \mapsto \int V_N \mu^{\otimes N}$, defined by a given λ-convex function V_N on $\mathbb{R}^{dN}$.*
- *The Boltzmann entropy $H(\mu)$ (relative to dx).*

In particular, for any λ-convex function V on $\mathbb{R}^d$ the corresponding free energy functional F_β^V (2-3), is lsc and λ-convex along generalized geodesics if $\beta \in]0, \infty]$.

Combining the results above we arrive at the following

Theorem 2.14. *Assume given $\beta \in]0, \infty]$ and $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$. Let $E(\mu)$ be a lsc functional on $\mathcal{P}_2(\mathbb{R}^d)$ which is λ-convex along generalized geodesics and satisfies the coercivity condition (2-18). Denote by F_β the corresponding free energy functional, $F_\beta := E + H/\beta$. Then the EVI-gradient flow μ_t on $\mathcal{P}_2(\mathbb{R}^d)$, emanating from μ_0, of the functional F_β exists. Moreover, if $\beta < \infty$, then $\mu_t = \rho_t\, dx$, where ρ_t has finite Boltzmann entropy. In particular:*

- *If V is an lsc finite λ-convex function on $\mathbb{R}^d$, then the gradient flow of F_β^V exists, defining a weak solution of the corresponding forward Kolmogorov equation/Fokker–Planck equation (2-8) with initial condition (2-9).*
- *If moreover $E(\mu)$ is Lipschitz continuous on $\mathcal{P}_2(\mathbb{R}^d)$ and $\beta > 0$ then ρ_t has finite Boltzmann entropy and Fisher information and the following continuity equation holds in the distributional sense on $\mathbb{R}^n \times \mathbb{R}$:*
$$\frac{\partial \rho_t}{\partial t} = \frac{1}{\beta}\Delta\rho_t + \nabla(\rho_t v_t), \tag{2-19}$$
  where $v_t = \partial^0 E$ is the minimal subdifferential of E at $\mu_t = \rho_t\, dx$.

Proof. By the previous lemma, F_β is also lsc and λ-convex and by Lemma 2.5 it also satisfies the coercivity condition. Hence, the EVI-gradient flow exists according to Theorem 2.9. Moreover, by the general results in [Ambrosio et al. 2005] F_β is decreasing along the flow and, in particular, locally uniformly bounded from above on $]0, \infty[$. But, by the coercivity assumption $E > -\infty$ on $\mathcal{P}_2(\mathbb{R}^d)$ and hence it follows that $H(\mu_t) < \infty$. The second statement then follows by the previous lemma and the fact that the coercivity condition holds: by λ-convexity $f(x) := v(x) + \lambda|x|^2$ is convex and hence $f(x) \geq -C|x|$ for some constant C, proving coercivity of v. To prove the last point first observe that

$E(\mu) \geq -A - Bd(\mu,\mu_0)^2 < \infty$ on $\mathcal{P}_2(\mathbb{R}^n)$ by the Lipschitz assumption. Since $F_\beta(\mu_t) \leq C$ it follows that $H(\mu_t) < \infty$, which in particular implies that μ_t has a density ρ_t. Moreover, by Theorem 2.9 $|\partial F_\beta(\mu_t)| < \infty$ for $t > 0$. But since E is assumed Lipschitz continuous we have $|\partial F_\beta(\mu_t)| < \infty$ if and only if $|\partial H(\mu_t)| < \infty$, which means that $I(\mu_t)$ has finite Fisher information (see Example 2.12). Finally, the distributional equation follows from Proposition 2.11. □

2E1. *The variational discretization scheme* (*minimizing movements*). Recall that the proof of Theorem 2.9 in [Ambrosio et al. 2005] uses a discrete approximation scheme introduced by De Giorgi, called the *minimizing movement* scheme. It can be seen as a variational formulation of the (backward) Euler scheme. Consider the fixed time interval $[0, T]$ and fix a (small) positive number τ (the "time step"). In order to define the "discrete flow" u_j^τ corresponding to the sequence of discrete times $t_j := j\tau$, where $t_j \leq T$ with initial data u_0, one proceeds by iteration: given $u_j \in M := \mathcal{P}_2(\mathbb{R}^d)$ the next step u_{j+1} is obtained by minimizing the functional

$$u \mapsto \frac{1}{2\tau} d(u,u_j)^2 + G(u)$$

on M. The minimizer exists and is unique as long as $\tau \leq \tau_0$, where τ_0 only depends on λ and the constant τ_* appearing in the inequality (2-18). Next, one defines $u^\tau(t)$ for any $t \in [0, T]$ by setting $u^\tau(t_j) := u_j^\tau$ and demanding that $u^\tau(t)$ be constant on $]t_j, t_{j+1}[$ and right continuous; we are using a slightly different notation than the one in [Ambrosio et al. 2005, Chapter 2].

The curve u_t is then defined as the large-m limit of $u_t^{(m)}$ in (M, d); as shown in [Ambrosio et al. 2005] the limit indeed exists and satisfies the EVI (2-17) and is thus uniquely determined. More precisely, the following quantitative convergence result holds; see Theorem 4.07, formula 4.024, and Theorem 4.09 of that same work:

Theorem 2.15. *Let G be a functional on $\mathcal{P}_2(\mathbb{R}^n)$ satisfying the assumptions in Theorem 2.9 with $\lambda \geq 0$. Then*

$$d^2(u^\tau(t), u(t)) \leq \tfrac{1}{2}|\tau|^2 |\partial G|^2(u_0),$$

where $|\partial G|(u_0)$ denotes the metric slope of G at u_0. If G is only assumed to be λ-convex for some, possibly negative, λ then

$$d(u^\tau(t), u(t)) \leq C|\tau|(G(u_0) - \inf G)$$

for some constant C only depending on λ and T.

Remark 2.16. By the last paragraph on page 79 in [Ambrosio et al. 2005] even if $\lambda < 0$, one does not need a lower bound on $\inf G$ if one replaces $|\tau|$ with $|\tau|^{1/2}$, as long as u_0 is assumed to satisfy $G(u_0) < \infty$.

3. Proof of Theorem 1.1

3A. ***The main assumptions on the interaction energy $E^{(N)}$.*** Set $X = \mathbb{R}^n$ and denote by d the Euclidean distance function on X. Throughout the paper $E^{(N)}$ will denote a symmetric, i.e., S_N-invariant, sequence of functions on X^N and we will make the following *main assumptions:*

(MA1) The functional $E^{(N)}$ is Lipschitz continuous in each variable on (X, d), uniformly in N.

(MA2) There exists a finite functional $E(\mu)$ on $\mathcal{P}_2(X)$ such that

$$\left|N^{-1}E^{(N)}(x_1,\dots,x_N)-E(\delta_N)\right|\le\epsilon_N(\delta_N)$$

for a sequence of functionals ϵ_N on $\mathcal{P}_2(X)$, converging pointwise to zero.

(MA3) The sequence $E^{(N)}$ is λ-convex on (X^N,d), uniformly in N.

Lemma 3.1. *Assume that (MA1) holds. Then, under the embedding*

$$\delta_N : X^{(N)}\to\mathcal{P}_1(X)$$

the sequence $E^{(N)}/N$ admits an extension which is uniformly Lipschitz continuous on $(\mathcal{P}_1(X),d_1)$ (and hence on $(\mathcal{P}_2(X),d_2)$, by Hölder's inequality). If moreover, (MA2) holds then the extended functionals converge pointwise on $\mathcal{P}(X)$ to the functional E, which thus defines a Lipschitz continuous functional on $(\mathcal{P}_1(X),d_1)$ (and hence on $(\mathcal{P}_2(X),d_2)$).

Proof. If $E^{(N)}$ is Lipschitz continuous in each variable with Lipschitz constant L, then taking the decomposition

$$\begin{aligned}&E^{(N)}(x_1,\dots,x_N)-E^{(N)}(y_1,\dots,y_N)\\&=\big(E^{(N)}(x_1,x_2,\dots,x_N)-E^{(N)}(y_1,x_2,\dots,x_N)\big)+\cdots+\big(E^{(N)}(y_1,\dots,y_{N-1},x_N)-E^{(N)}(y_1,\dots,y_N)\big),\end{aligned}$$

where the right-hand side consists of N terms, gives

$$N^{-1}\left|E^{(N)}(x_1,\dots,x_N)-E^{(N)}(y_1,\dots,y_N)\right|\le LN^{-1}\sum_{i=1}^N d(x_i,y_i).$$

Since, $E^{(N)}$ is assumed S_N-invariant we deduce that, for any given $\sigma\in S_N$,

$$N^{-1}\left|E^{(N)}(x_1,\dots,x_N)-E^{(N)}(y_1,\dots,y_N)\right|\le LN^{-1}\sum_{i=1}^N d(x_i,y_{\sigma(i)}).$$

Hence, taking the infimum over all $\sigma\in S_N$ shows that $E^{(N)}/N$ is Lipschitz continuous on $(X^{(N)},d_{(1)})$. By the isometry property in Lemma 2.3, this means that we can identify $E^{(N)}/N$ with a Lipschitz continuous function f_N on a subset F_N of $\mathcal{P}_1(X)$. The desired extension property now follows from the general fact that any Lipschitz continuous function f defined on a subset of a metric space Y admits a Lipschitz continuous extension to all of Y. For example, the extension (that we still denote by f) can be taken as an infimal convolution [Hiriart-Urruty 1980].

To prove the last statement in the lemma we assume that (MA2) holds. Taking $\mu_0:=\delta_{x_0}$ it follows that $f_N(\mu_0)\to E(\mu_0)$. By the Arzelà–Ascoli theorem this implies that there exists a Lipschitz continuous function f on $(\mathcal{P}_1(X),d_1)$ such that, after perhaps passing to a subsequence, $f_N\to f$ uniformly on compacts of $\mathcal{P}(X)$. By the assumption (MA2) we must have $f=E$ and hence the whole sequence f_N has to convergence to E, which is thus Lipschitz continuous. As a consequence, the sequence $\epsilon_N:=f-f_N$ is also uniformly Lipschitz continuous on $\mathcal{P}(X)$. Finally, fix $\mu\in\mathcal{P}_1(X)$ and take some sequence $\boldsymbol{x}_N$ in X^N such that $\delta_N(\boldsymbol{x}_N)\to\mu$ in $\mathcal{P}_1(X)$. Then, using the triangle inequality three times together with the uniform Lipschitz continuity of $E^{(N)}$, $\epsilon^{(N)}$ and E we have

$$|E(\mu) - N^{-1}E^{(N)}(\mu)| \leq |\epsilon_N(\mu)| + 3Ld_1(\delta_N(\boldsymbol{x}_N), \mu),$$

which, by the assumption (MA2) converges to zero, as desired (we have used the same notation E^N/N for the extended functional f_N). □

The next lemma verifies that the mean free energy functional (2-10) and the free energy functional $F_\beta(:= E + H/\beta)$, corresponding to the sequence $E^{(N)}$, satisfy the assumptions in Theorem 2.9:

Lemma 3.2. *If the main assumptions hold, then the following hold for any given $\beta \in]0, \infty]$:*

- *The mean free energy functional $N^{-1}F_\beta^{(N)}$ is λ-convex along generalized geodesics in $\mathcal{P}_2(X^{(N)}, d_{(2)})$ and satisfies the following uniform coercivity property: there exist constants $\tau_*, C > 0$*

$$N^{-1}F_\beta^{(N)} \geq -\frac{1}{\tau_*}, d_{(2)}(\,\cdot\,, \delta_{(0,\ldots,0)})^2 - C. \tag{3-1}$$

- *The free energy functional F_β is λ-convex along generalized geodesics in $\mathcal{P}_2(X)$ and satisfies*

$$F_\beta \geq -\frac{1}{\tau_*} d_2(\,\cdot\,, \delta_0)^2 - C. \tag{3-2}$$

Proof. The λ-convexity of $F_\beta^{(N)}$ follows directly from the assumption (MA3) combined with first and third points in Lemma 2.13. Moreover, (MA1) together with (MA2) implies (using Lemma 3.1) that there exists a constant C such that

$$N^{-1}E^{(N)}(x_1, \ldots, x_N) \geq -LN^{-1}\sum_{i=1}^{N} |x_i| - C.$$

Using Hölder's inequality and integrating over X^N gives the uniform coercivity property (3-1) when $\beta = \infty$. The general case then follows from Lemma 2.5. Next, since E is Lipschitz continuous (by the previous lemma) the inequality (3-2) also follows in a similar manner. All that remains is thus to check that $E(\mu)$ is λ-convex along generalized geodesics in $\mathcal{P}_2(X)$. To this end we note that $E(\mu)$ is the pointwise limit on $\mathcal{P}_2(X)$ of the functionals

$$\mu \mapsto \int N^{-1}E^{(N)}\mu^{\otimes N},$$

as follows from Proposition 3.6 below (applied to $\mu_N = \mu^{\otimes N}$). For any fixed N the functional above is λ-convex along generalized geodesics (by (MA3) combined with the second point in Lemma 2.13). Letting $N \to \infty$ thus reveals that E is indeed λ-convex along generalized geodesics. As a consequence, so is F_β for any $\beta \in]0, \infty]$ (by the third point in Lemma 2.13). □

3B. ***Propagation of chaos in the time-discretized setting.*** In this section we will formulate and prove a discretized version of Theorem 1.1, assuming that the main assumptions hold. Let $\mu_0^{(N)}$ be a given sequence of symmetric elements in $\mathcal{P}_2(X^N)$ and $\mu_0 \in \mathcal{P}(X)$ be an element. Given a (small) "time step" τ we denote by $\mu_{t_j}^{(N)}$ the discretized minimizing movement of the mean free energy functional $N^{-1}F_\beta^{(N)}$ on $\mathcal{P}_2(X^{(N)}, d_{(2)})$ (2-10) emanating from $\mu_0^{(N)}$ and by μ_{t_j} the discretized minimizing movement of the free energy functional $F_\beta(:= E + H/\beta)$ on $\mathcal{P}_2(X)$ emanating from μ_0. We recall that this means (see Section 2E1) that, given $\mu_{t_j}^{(N)} \in \mathcal{P}_2(X^N)$, the next measure $\mu_{t_{j+1}}^{(N)}$ is defined as the minimizer of the following functional on $\mathcal{P}(X^N)$:

$$\frac{1}{N} J_{j+1}^{(N)}(\cdot) := \frac{1}{2\tau}\frac{1}{N} d(\cdot, \mu_{t_j}^{(N)})^2 + \frac{1}{N} F_\beta^{(N)}(\cdot). \tag{3-3}$$

Similarly, given $\mu_{t_j} \in \mathcal{P}(X)$, the next measure $\mu_{t_{j+1}}$ is defined as the minimizer of the following functional on $\mathcal{P}(X)$:

$$J_{j+1}(\cdot) = \frac{1}{2\tau}\frac{1}{N} d(\cdot, \mu_{t_j})^2 + \frac{1}{N} F_\beta(\cdot)$$

The sequences $\mu_{t_j}^{(N)}$ and μ_{t_j} are well-defined according to Lemma 3.2 and Theorem 2.9 (or rather its proof using minimizing movements, recalled in Section 2E1). We note that the sequence $\mu_{t_j}^{(N)}$ may (by the third isometry property in Lemma 2.3) be identified with the minimizing movement of the mean free energy functional $F^{(N)}/N$ on $\mathcal{P}_2(X^{(N)}, d_{(2)})$, which in turn embeds isometrically to give a discrete flow $\Gamma_{t_j}^{(N)}$ in $W_2(\mathcal{P}_2(X), d_2)$.

Theorem 3.3. *Assume that at time t_j*

$$\lim_{N\to\infty} (\delta_N)_* \mu_{t_j}^{(N)} = \delta_{\mu_{t_j}}$$

in $W_2(\mathcal{P}_2(X), d_2)$. Then, at the next time step t_{j+1}

$$\lim_{N\to\infty} (\delta_N)_* \mu_{t_{j+1}}^{(N)} = \delta_{\mu_{t_{j+1}}}$$

in $W_2(\mathcal{P}_2(X), d_2)$. As a consequence, if $\mu_{t_j}^{(N)}$ is of the form $\mu_{t_j}^{(N)} = \mu_0^{\otimes N}$ when $t_j = 0$, then $(\delta_N)_*\mu_{t_j}^{(N)}$ converges to $\delta_{\mu_{t_j}}$ in $W_2(\mathcal{P}_2(X), d_2)$ for any t_j.

The last statement follows directly from induction using the first statement and the following basic observation:

$$\mu_0 \in \mathcal{P}_2(X) \quad \Longrightarrow \quad (\delta_N)_*\mu_0^{\otimes N} \to \delta_{\mu_0} \text{ in } W_2(\mathcal{P}_2(X), d_2). \tag{3-4}$$

Indeed, by Lemma 2.4 the convergence holds in $\mathcal{P}(\mathcal{P}(X))$. Moreover, setting $\Gamma_0 := \delta_{\delta_0}$ gives

$$\int_{\mathcal{P}(X)} d^2(\Gamma, \Gamma_0)(\delta_N)_*\mu_0^{\otimes N} = \int_X |x|^2 \mu_0 = \int d^2(\Gamma_{\delta_{\mu_0}}, \Gamma_0)(\delta_N)_*\mu_0^{\otimes N}$$

and hence (3-4) follows from Proposition 2.2. Thus it will be enough to prove the first statement in the previous theorem.

3C. *Proof of Theorem 3.3.*

We start with the following direct consequence of Proposition 2.2 combined with Lemma 2.3:

Lemma 3.4. *Let μ_N be a sequence of symmetric probability measures on X^N and denote by $\Gamma_N := (\delta_N)_*\mu_N$ the corresponding probability measures on $\mathcal{P}(X)$. Assume that the d_2-distance of Γ_N to a fixed element in the Wasserstein space $W_q(\mathcal{P}_2(X))$ is uniformly bounded from above. Then, after perhaps passing to a subsequence, there is a probability measure Γ in $W_2(\mathcal{P}_2(X))$ such that*

$$\lim_{N\to\infty} (\delta_N)_*\mu_N = \Gamma$$

in $W_1(\mathcal{P}_2(X))$.

We next recall the following well-known result about the asymptotics of the mean entropy, proved in [Robinson and Ruelle 1967]; see also Theorem 5.5 in [Hauray and Mischler 2014] for generalizations. The proof is based on the subadditivity properties of the entropy.

Proposition 3.5. *Let μ_N be a sequence of probability measures on X^N such that $(\delta_N)_*\mu_N$ converges weakly to $\Gamma \in \mathcal{P}(\mathcal{P}(X))$. Then*

$$\liminf_{N\to\infty} H^{(N)}(\mu_N) \geq \int_{\mathcal{P}(X)} H(\mu)\Gamma.$$

We will also use the following result, which generalizes a result in [Messer and Spohn 1982] concerning the case when E_N is quadratic:

Proposition 3.6. *Let $\mu^{(N)}$ be a sequence of probability measures on X^N such that $\Gamma_N := (\delta_N)_*\mu_N$ converges to Γ in $W_1(\mathcal{P}_2(X))$. If $E^{(N)}$ satisfies the assumptions (MA1) and (MA2) in the main assumptions (Section 3A), then*

$$\lim_{N\to\infty} \frac{1}{N}\int_{X^N} E^{(N)}\mu^{(N)} = \int_{\mathcal{P}(X)} E(\mu)\Gamma.$$

Proof. Recall that the L^1-Wasserstein distance d_{W_1} on $W_1(Y, d)$ admits the following dual representation (the Kantorovich–Rubinstein theorem [Villani 2003, page 34]):

$$d_{W_1}(\mu, \nu) = \sup_{u\in \mathrm{Lip}_1} \int u(\mu - \nu),$$

where u ranges over all Lipschitz continuous functions on Y with Lipschitz constant 1. In the present setting we take (Y, d) as $\mathcal{P}_2(X)$ endowed with the d_2-distance. By assumption

$$d_{W_1}(\Gamma_N, \Gamma) \to 0. \tag{3-5}$$

Using the empirical measure δ_N we can identify $N^{-1}E^{(N)}$ with a uniformly Lipschitz continuous sequence of functions on $(\mathcal{P}_2(X), d_2)$, which by the main assumptions converges pointwise to the Lipschitz continuous functional $E(\mu)$ (using Lemma 3.1 applied to $p = 2$). Since $N^{-1}E^{(N)}$ is uniformly Lipschitz continuous we have

$$\lim_{N\to\infty} \int_{\mathcal{P}_2(X)} N^{-1}E^{(N)}(\Gamma_N - \Gamma) = 0$$

using the (3-5) combined with the dual representation of the L^1-Wasserstein distance. Hence, all that remains is to verify that

$$\lim_{N\to\infty} \int_{\mathcal{P}_2(X)} N^{-1}E^{(N)}\Gamma = \int_{\mathcal{P}_2(X)} E(\mu)\Gamma.$$

But this follows from the dominated convergence theorem. Indeed, $N^{-1}E^{(N)}$ converges pointwise to E on $\mathcal{P}_2(X)$ and (by the uniform Lipschitz property) is uniformly dominated by the function $A + Bd_2$, which is in $L^1(\Gamma)$, since $\Gamma \in W_1(\mathcal{P}_2(X))$. □

Next we turn to the asymptotics of the distances, establishing the following key property:

Proposition 3.7. *Assume that a sequence ν_N of symmetric probability measures on X^N satisfies*

$$\lim_{N\to\infty} (\delta_N)_* \nu_N = \delta_\nu$$

in the distance topology in $W_2(\mathcal{P}_2(X))$. Then any sequence μ_N such that $(\delta_N)_*\mu_N$ converges weakly to $\Gamma \in \mathcal{P}(\mathcal{P}(X))$ satisfies

$$\liminf_{N\to\infty} \frac{1}{N} d(\mu_N, \nu_N)^2 \geq \int_{\mathcal{P}(X)} d(\mu, \nu)^2 \Gamma(\mu)$$

and equality holds if and only if $(\delta_N)_*\mu_N$ converges to Γ in the distance topology in $W_2(\mathcal{P}_2(X))$.

Proof. Consider the isometry

$$\delta_N : (X^{(N)}, d_{X^{(N)}}) \hookrightarrow (\mathcal{P}(X), d_W), \quad (x_1, \dots, x_N) \mapsto \delta_N := \frac{1}{N} \sum \delta_{x_i},$$

defined in terms of the L^2-distances. We equip the space $\mathcal{P}(\mathcal{P}(X))$ with the L^2-Wasserstein (pre)metric d induced from distance d_W on $\mathcal{P}(X)$; i.e., we consider the subspace $W_2(\mathcal{P}(X))$. By Lemma 2.3

$$\frac{1}{N} d(\mu_N, \nu_N)^2 = d((\delta_N)_*\mu_N, (\delta_N)_*\nu_N)^2.$$

We now first assume that $(\delta_N)_*\mu_N$ converges to Γ in the d-distance topology in $W_2(\mathcal{P}_2(X))$. Then the "triangle inequality" for d immediately gives

$$\lim_{N\to\infty} d((\delta_N)_*\mu_N, (\delta_N)_*\nu_N)^2 = d(\Gamma, \delta_\nu)^2.$$

Next we will use the following simple general fact for the Wasserstein distance on $\mathcal{P}(Y, d)$:

$$d(\mu, \delta_{y_0})^2 = \int d(y, y_0)^2 \mu(y),$$

which follows from the fact that the only coupling between μ and δ_{y_0} is the product $\mu \otimes \delta_{y_0}$. Applied to $Y = \mathcal{P}(X)$ this gives

$$d((\delta_N)_*\mu_N, \delta_\nu)^2 = \int_{\mathcal{P}(X)} d(\mu, \nu)^2 \, \Gamma(\mu)$$

which concludes the proof using that $d(\delta_\mu, \delta_\nu) = d(\mu, \nu)$ by the general fact above. More generally, if $(\delta_N)_*\mu_N$ is only assumed to converge to Γ weakly in $\mathcal{P}(\mathcal{P}(X))$, then the lower semicontinuity of the Wasserstein distance function with respect to the weak topology instead gives

$$\liminf_{N\to\infty} \frac{1}{N} d(\mu_N, \nu_N)^2 \geq \int_{\mathcal{P}(X)} d(\mu, \nu)^2 \, \Gamma(\mu).$$

Finally, if equality holds above, then, by the previous arguments,

$$\lim_{N\to\infty} \int_{\mu\in\mathcal{P}(X)} d(\mu, \nu)^2 (\delta_N)_*\mu_N = \int_{\mathcal{P}(X)} d(\mu, \nu)^2 \, \Gamma(\mu)$$

(i.e., the "second moments of $(\delta_N)_*\mu_N$ converge to the second moments of Γ) and then it follows from Proposition 2.2 that $(\delta_N)_*\mu_N$ converges to Γ in the distance topology in $W_2(\mathcal{P}(X))$. ☐

3C1. *Conclusion of the proof of Theorem 3.3.* Without loss of generality we may set $\beta = 1$ and we will thus drop the subindex β from the notations. We start by observing that for any fixed μ in $\mathcal{P}(X)$ we have, by the defining property of $\mu^{(N)}_{t_{j+1}}$, that

$$\frac{1}{N} J^{(N)}_{j+1}(\mu^{(N)}_{t_{j+1}}) \leq \frac{1}{N} J^{(N)}_{j+1}(\mu^{\otimes N}),$$

where the right-hand side converges, by the propositions above, to $J_{j+1}(\mu)$ as $N \to \infty$, where $J_{j+1}(\mu) = \frac{1}{2\tau} d(\mu, \mu_j)^2 + F(\mu)$. In particular, taking $\mu = \mu_{j+1}$ gives

$$\limsup_{N\to\infty} \frac{1}{N} J^{(N)}_{j+1}(\mu^{(N)}_{t_{j+1}}) \leq J_{j+1}(\mu_{j+1}), \tag{3-6}$$

where μ_{j+1} is the unique minimizer of J_{j+1}.

Next we consider the lower bound. By the minimizing property of $\mu^{(N)}_{t_{j+1}}$ we have a uniform control on the d_2-distance:

$$d_2\big((\delta_N)_*\mu^{(N)}_{t_{j+1}}, (\delta_N)_*\mu^{(N)}_{t_j}\big)^2 = \frac{1}{N} d_2(\mu^{(N)}_{t_{j+1}}, \mu^{(N)}_{t_j})^2 \leq C. \tag{3-7}$$

Indeed, the minimizing property together with the previous bound gives

$$\frac{1}{\tau} d_2\big((\delta_N)_*\mu^{(N)}_{t_{j+1}}, (\delta_N)_*\mu^{(N)}_{t_j}\big)^2 \leq C - \frac{1}{N} F^{(N)}(\mu^{(N)}_{t_{j+1}}).$$

Hence, the inequality (3-7) follows from the uniform coercivity property of $\frac{1}{N} F^{(N)}$, formula (3-1).

Now, it follows from the induction assumption and the triangle inequality for d that $\mu^{(N)}_{t_{j+1}}$ satisfies the assumptions of Proposition 2.2. Accordingly, we may, after passing to a subsequence, assume that $\mu_N := \mu^{(N)}_{t_{j+1}}$ converges as in Lemma 3.4, or more precisely that

$$(\delta_N)_*\mu^{(N)}_{t_{j+1}} \to \Gamma$$

in $W_1(\mathcal{P}_2(X))$ for some $\Gamma \in W_2(\mathcal{P}_2(X))$. It then follows from Propositions 3.5, 3.6 and 3.7 that

$$\liminf_{N\to\infty} \frac{1}{N} J^{(N)}_{j+1}(\mu^{(N)}_{t_{j+1}}) \geq \int d\Gamma(\mu) J_{j+1}(\mu). \tag{3-8}$$

Combining the previous lower bound with the upper bound (3-6) and using that μ_{j+1} is the unique minimizer of J_{j+1} then forces $\Gamma = \delta_{\mu_{j+1}}$ and

$$\lim_{N\to\infty} \frac{1}{N} J^{(N)}_{j+1}(\mu^{(N)}_{t_{j+1}}) = J_{j+1}(\mu). \tag{3-9}$$

But this means that

$$\lim_{N\to\infty} (\delta_N)_*\mu^{(N)}_{t_{j+1}} = \delta_{\mu_{t_{j+1}}}$$

weakly in $\mathcal{P}(X)$ and by the equality (3-9) that

$$\lim_{N\to\infty} d((\delta_N)_*\mu^{(N)}_{t_{j+1}}, \delta_{\mu_{t_{j+1}}}) = d(\delta_{\mu_{t_{j+1}}}, \delta_{\mu_j}).$$

But then it follows from Proposition 3.7 (applied to $\nu = \delta_{\mu_j}$) that $(\delta_N)_*\mu_N$ converges to Γ in the distance topology in $W_2(\mathcal{P}_2(X))$, as desired.

3D. *Convergence in the nondiscrete setting: proof of Theorem 1.1.* We recall that in the previous section we had fixed a time step τ. In this section we will emphasize the dependence on τ by setting

$$\Gamma_N^\tau(t) := (\delta_N)_*(\mu_{t_j}^{(N)}), \quad \Gamma^\tau(t) := \delta_{\mu_{t_j}}.$$

The assumptions in Theorem 1.1 imply, by the last statement in Theorem 3.3, that

$$\lim_{N\to\infty} d(\Gamma_N^\tau(t), \Gamma^\tau(t)) = 0 \tag{3-10}$$

in $W_2(\mathcal{P}(X))$. Next, set

$$\Gamma_N(t) := (\delta_N)_*(\mu_t^{(N)}), \quad \Gamma(t) := \delta_{\mu_t},$$

where $\mu_t^{(N)}$ and μ_t denote the EVI-gradient flows of $F_\beta^{(N)}$ and F_β, respectively (whose existence is a consequence of Theorem 2.9 combined with Lemma 3.2). Consider now a fixed time interval $[0, T]$. For any fixed $t \in]0, T[$ we then have, by the triangle inequality,

$$d(\Gamma_N(t), \Gamma(t)) \le d(\Gamma_N(t), \Gamma_N^\tau(t)) + d(\Gamma(t), \Gamma^\tau(t)) + d(\Gamma_N^\tau(t), \Gamma^\tau(t)).$$

First assume, for simplicity, that the assumption (MA3) (Section 3A) holds with $\lambda \ge 0$. By the convexity properties in Lemma 3.2 and the isometry property in Lemma 2.3 we have, using Theorem 2.15, that $d(\Gamma_N(t), \Gamma_N^\tau(t)) \le C\tau$ (uniformly in N) and $d(\Gamma(t), \Gamma^\tau(t)) \le \tau C$. Hence, combining the previous two inequalities with the convergence (3-10) and letting first $N \to \infty$ and then $\tau \to 0$, gives $\lim_{N\to\infty} d(\Gamma_N(t), \Gamma(t)) = 0$, which proves Theorem 1.1 when $\lambda \ge 0$. Finally, in the case when $\lambda \le 0$ the previous argument still applies, with the error $O(\tau)$ replaced by $\mathcal{O}(\tau^{1/2})$ according to Remark 2.16.

4. Permanental processes and toric Kähler–Einstein metrics

In this section we will deduce Theorem 1.2, stated in the Introduction, from Theorem 1.1, proved in the previous section.

4A. *Permanental processes: setup.* Let P be a convex body in $\mathbb{R}^n$ containing zero in its interior and denote by ν_P the corresponding uniform probability measure on P; i.e.,

$$\nu_P = \frac{1_P d\lambda}{V(P)},$$

where $d\lambda$ denotes Lebesgue measure and $V(P)$ is the Euclidean volume of P. Setting $P_k := P \cap (\mathbb{Z}/k)^n$, we let N_k be the number of points in P_k and fix an auxiliary ordering $p_1, \ldots, p_{N_k}$ of the N_k elements of P_k. Given a configuration $(x_1, \ldots, x_{N_k})$ of points on $X := \mathbb{R}^n$ we set

$$E^{(N_k)}(x_1, \ldots, x_{N_k}) := \frac{1}{k} \log \sum_{\sigma \in S_{N_k}} e^{k(x_1 \cdot p_{\sigma(1)} + \cdots + x_N \cdot p_{\sigma(N_k)})}, \tag{4-1}$$

which, as explained in Section 1B, can be written as the scaled logarithm of a permanent. To simplify the notation we will often drop the subscript k and simply write $N_k = N$, since anyway $N \to \infty$ if and only if $k \to \infty$. We will denote by $C(\mu, \nu)$ the Monge–Kantorovich optimal cost for transport between the probability measures μ and ν, with respect to the standard symmetric quadratic cost function $c(x, p) = -x \cdot p$:

$$C(\mu, \nu) := \inf_{\gamma} -\int_{X \times X} x \cdot p\gamma, \tag{4-2}$$

where the γ ranges over all couplings (transport plans) between μ and ν (see Remark 2.1).

Proposition 4.1. *The main assumptions are satisfied for $E^{(N)}$ with $\lambda = 0$ and $E(\mu) = -C(\mu, \nu_P)$. Equivalently, formulated in terms of the Wasserstein L^2-distance*

$$E(\mu) = -\frac{1}{2} d_{W_2}(\mu, \nu_P)^2 + \frac{1}{2}\int x^2\, d\mu + c_P, \quad c_P := \frac{1}{2}\int p^2 \nu_P. \tag{4-3}$$

In particular, $-C(\,\cdot\,, \nu_P)$ is convex along generalized geodesics.

Proof. This follows essentially from the results in [Berman 2013b]. But for completeness we give a direct proof here:

Step 1: (MA2) holds. First observe that

$$N^{-1}|E^{(N)} - E^{(N)}_{\text{trop}}| \le C \frac{N}{\log N}, \tag{4-4}$$

where $E^{(N)}_{\text{trop}}$ denotes the tropical analog of $E^{(N)}$ (see formula (4-8) below). Indeed, fixing $(x_1, \dots, x_N)$ and denoting by σ_0 the element in S_N maximizing $\sigma \mapsto e^{x_1 p_{\sigma(1)} + \cdots + x_1 p_{\sigma(1)}}$ we have

$$e^{kx_1 \cdot p_{\sigma_0(1)} + \cdots + x_1 \cdot p_{\sigma_0(1)}}(1 + 0 + \cdots + 0) \le E^{(N)}(x_1, \dots, x_N) \le N!\, e^{kx_1 \cdot p_{\sigma_0(1)} + \cdots + kx_1 \cdot p_{\sigma_0(1)}}.$$

Hence, taking the log and dividing by k proves the inequality (4-4), using that $k^{-1}N^{-1}\log N! \to \infty$ (by Stirling, since $N \sim k^n$). Next observe that

$$-N^{-1} E^{(N)}_{\text{trop}}(\boldsymbol{x}) = \frac{1}{2} d^2(\delta_N(\boldsymbol{x}), \delta_N(\boldsymbol{p})) - \frac{1}{2}\int_{\mathbb{R}^n} |x|^2 \delta_N(\boldsymbol{x}) - \frac{1}{2}\int_{\mathbb{R}^n} |p|^2 \delta_N(p). \tag{4-5}$$

Indeed, rewriting $x \cdot p = |x - p|^2/2 - |p|^2/2 - |p|^2/2$ reveals that $2N^{-1}E^{(N)}_{\text{trop}}(\boldsymbol{x})$ is equal to the distance between $\boldsymbol{x}$ and $\boldsymbol{p}$ in $(X^{(N)}, d_{(2)})$ minus two quadratic terms. Since $\delta_N^* d = d_{(2)}$, this proves (4-5). All in all this means that assumption (MA2) is satisfied with

$$2\epsilon_N(\mu) := \left|d^2(\mu, \delta_N(\boldsymbol{p})) - d^2(\mu, \nu_P)\right| + \left|\int_{\mathbb{R}^n} |p|^2(\nu_P - \delta_N(p))\right|.$$

Step 2: (MA1) and (MA3) hold. First recall the basic fact that if ϕ_σ is a family of smooth convex functions on $\mathbb{R}^m$ and γ is a probability measure on the parameter space S then $\phi := k^{-1}\log\int e^{k\phi_\sigma}\, d\gamma(\sigma)$ is also convex, for any given positive number k, and $\nabla\phi$ is contained in the convex hull of $\{\nabla\phi_\sigma\}$. In the present setting we take $S := S_N$ endowed with the counting measure γ and $\phi_\sigma(\boldsymbol{x}) := \boldsymbol{x} \cdot \boldsymbol{p}_\sigma$, which is clearly convex and satisfies $\nabla_{x_i}\phi_\sigma \in P$. Since P is convex and uniformly bounded, this concludes the proof of Step 2. The convexity of $-C(\,\cdot\,, \nu_P)$ then follows from Lemma 2.13. Equivalently, this means that $-\frac{1}{2} d_{W_2}(\mu, \nu_P)^2$ is -1-convex. In fact, as shown in [Ambrosio et al. 2005] using a different argument $-\frac{1}{2} d_{W_2}(\,\cdot\,, \nu)^2$ is -1-convex for any fixed $\nu \in \mathcal{P}_2(\mathbb{R}^n)$. □

Next, we recall that the *Monge–Ampère measure* $\mathrm{MA}(\phi)$ of a convex function ϕ on $\mathbb{R}^n$ (4-4) is defined by the property that, for a given Borel set E,

$$\int_E \mathrm{MA}(\phi) := \int_{(\partial\phi)(E)} d\lambda,$$

where $d\lambda$ denotes Lebesgue measure and $\partial\phi$ denotes the subgradient of ϕ (which defines a multivalued map from $\mathbb{R}^n$ to $\mathbb{R}^n$). This is also called the Hessian measure; see [Villani 2003, Section 4.1.4]. In particular, if $\phi \in C^2$, then

$$\mathrm{MA}(\phi) = \det(\partial^2\phi)\,dx,$$

where $\partial^2\phi$ denotes the Hessian matrix of ϕ. We will denote by $\mathcal{C}_P$ the space of all convex functions ϕ on $\mathbb{R}^n$ whose subgradient $\partial\phi$ satisfies

$$(\partial\phi)(\mathbb{R}^n) \subset P$$

and we will say that ϕ is *normalized* if $\phi(0) = 0$. By the convexity of ϕ the gradient condition above equivalently means that ϕ is bounded from above by the support function ϕ_P of P, where $\phi_P(x) := \sup_{p\in P} p\cdot x$.

By Brenier's theorem [1991], given $\mu = \rho\,dx$ in $\mathcal{P}_2(\mathbb{R}^n)$ there exists a unique normalized $\phi \in \mathcal{C}_P$ such that

$$\mathrm{MA}(\phi) = \rho\,dx, \tag{4-6}$$

which equivalently means that the corresponding L^∞-map $\nabla\phi$ from $\mathbb{R}^n$ to P satisfies

$$(\nabla\phi)_*\mu = \nu_P.$$

Given the previous proposition we can use the differentiability result in [Ambrosio et al. 2005] for the Wasserstein L^2-distance to get the following result (see Proposition 2.11 and the subsequent discussion for the definition of the minimal subdifferential):

Lemma 4.2. *The minimal subdifferential of $-C(\cdot\,,\nu_P)$ on the subspace $\mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^n)$ of all probability measures in $\mathcal{P}_2(\mathbb{R}^n)$ which are absolutely continuous with respect to dx, may, at a given point $\rho\,dx$, be represented by the L^∞-vector field $\nabla\phi$, where ϕ is the unique normalized solution in $\mathcal{C}_P$ to (4-6).*

Proof. Given formula (4-3) this follows immediately from Theorem 10.4.12 in [Ambrosio et al. 2005] and the fact that if $\mu \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^n)$, then Brenier's theorem gives that the optimal transport plan (coupling) from $\mathbb{R}^n$ to P realizing the infimum defining $d_{W_2}(\mu,\nu_P)^2$ is given by the L^∞-map $\nabla\phi$, where ϕ solves (4-6). Since the barycentric projection appearing in Theorem 10.4.12 in [Ambrosio et al. 2005] for the transport plan defined by a transport map gives back the transport map, see Theorem 12.4.4 of the same work, this concludes the proof. □

4B. ***Existence of the gradient flow for $F_\beta(\mu)$.*** Given $\beta \in\,]0,\infty]$ we set $F_\beta(\mu) := -C(\mu,\nu_P) + H(\mu)/\beta$.

Proposition 4.3. *The gradient flow μ_t of F_β on $\mathcal{P}_2(\mathbb{R}^n)$ emanating from a given μ_0 exists for any $\beta \in\,]0,\infty]$. Moreover, for $\beta < \infty$ we have that $\mu_t = \rho_t(x)\,dx$, where ρ_t has finite Boltzmann entropy* *and Fisher information and $\rho(x,t) := \rho_t(x)$ satisfies the following equation in the sense of distributions on $\mathbb{R}^n \times]0,\infty[$:*

$$\frac{d\rho_t}{dt} = \frac{1}{\beta}\Delta\rho_t + \nabla\cdot(\rho_t\nabla\phi_t), \tag{4-7}$$

where ϕ_t is the unique normalized solution in $\mathcal{C}_P$ to (4-6) and $\nabla\phi_t$ defines a vector field with coefficients in L^∞_{loc}.

Proof. Given the previous lemma this follows immediately from Theorem 8.3.1 and Corollary 11.1.8 in [Ambrosio et al. 2005] (the case $\beta = \infty$ has previously been considered by Brenier [2010; 2011; 2016] by lifting the problem to the space of L^2-maps from $\mathbb{R}^n$ to $\mathbb{R}^n$ where Hilbert space techniques can be applied). □

4C. ***Conclusion of the proof of Theorem 1.2.*** By Proposition 4.1 the main assumptions are satisfied. Hence, Theorem 1.1 implies that the corresponding empirical measures converge in law, as $N \to \infty$, to the measure $\mu_t \in \mathcal{P}_2(\mathbb{R}^n)$, where the curve $t \mapsto \mu_t$ is the gradient flow on the Wasserstein space $\mathcal{P}_2(\mathbb{R}^n)$ of the corresponding free energy functional F_β, emanating from μ_0. Finally, Proposition 4.3 says that the gradient flow in question satisfies the evolution equation appearing in Theorem 1.2.

4D. ***The tropical setting.*** The results above are also valid when the permanental interaction energy $E^{(N_k)}(x_1,\dots,x_{N_k})$ is replaced by its tropical analog, i.e., the convex piecewise affine convex function

$$E^{(N_k)}_{\mathrm{trop}}(x_1,\dots,x_{N_k}) := \max \sum_{\sigma\in S_{N_k}} x_1\cdot p_{\sigma(1)} + \cdots + x_N\cdot p_{\sigma(N_k)}. \tag{4-8}$$

In other words this is a *tropical permanent,* i.e., the permanent of the rank-N matrix $(x_i\cdot p_j)$ in the tropical semiring over $\mathbb{R}$, i.e., the set $\mathbb{R}\cup\{-\infty\}$ where the plus and multiplication operations are defined by $\max\{a,b\}$ and $a+b$, respectively [Itenberg and Mikhalkin 2012]. Equivalently, in terms of discrete transport theory this means that

$$E^{(N_k)}_{\mathrm{trop}}(x_1,\dots,x_{N_k}) := -C((\delta_N(x),\delta_N(p))).$$

Passing to the tropical setting has, in particular, computational advantages. Indeed, while all known methods for evaluating (general) permanents take exponential time, the tropical permanent above is, by its very definition, the optimal value of a linear assignment problem and can be computed using an algorithm of cubic-time complexity; see the discussion in [Brenier et al. 2003].

5. Outlook

In this final section we point out some relations between the limiting evolution equation appearing in Theorem 1.2 (whose static solutions correspond to toric Kähler–Einstein metrics) and other well-known evolution equations. We also indicate some relations to sticky particle systems appearing at the microscopic level (i.e., for finite N) and the complex geometric picture. These relations will be elaborated on in a sequel to the present paper [Berman $\geq$ 2018].

5A. ***Relation to other evolution equations and traveling waves.*** In the one-dimensional case when $P := [-a_-, -a_+]$, integrating the evolution equation for ρ_t in Theorem 1.2 once reveals that the bounded decreasing function $u(x,t) := -\partial_x \phi_t$ (physically playing the role of a velocity field) satisfies *Burger's equation* [Hopf 1950] with positive *viscosity* $\kappa := \beta^{-1}$:

$$\partial_t u = \kappa \partial_x^2 u - u \partial_x u$$

with the left and right space asymptotics $\lim_{x\to\pm\infty} u(x,t) = a_\pm$. We recall that Burger's equation is the prototype of a nonlinear wave equation and a scalar conservation law, which is used, among many other things, as a toy model for turbulence in the Navier–Stokes equations [Frisch and Bec 2001]. Interestingly, the barycenter b_P of the polytope P coincides, in this one-dimensional situation, with the negative of the *speed* $s := (a_+ + a_-)/2$ of the time-dependent solution u in the terminology of scalar conservation laws [Lax 1973]. Hence, the vanishing condition $b_P = 0$, which in general is tantamount to the existence of a stationary solution $\rho_t = \rho$ (as discussed in connection to Theorem 1.2) simply means, from the point of view of nonlinear wave theory, that the speed s vanishes.

Similarly, the function $\phi(x,t) := \phi_t(x)$, which in complex-geometric terms is a Kähler potential, satisfies (after the appropriate normalization) the following viscous Hamilton–Jacobi equation, known as the deterministic KPZ equation in the literature on growth of random surfaces [Kardar et al. 1986; Hairer 2013]:

$$\partial_t \phi = \kappa \partial_x^2 \phi + \tfrac{1}{2}(\partial_x \phi)^2. \tag{5-1}$$

In the general higher-dimensional case, the evolution equation (1-13) (which is different than the higher-dimensional version of Burger's equation) can be seen as a dissipative viscous/diffusive version of the semigeostrophic equation appearing in dynamic meteorology; see [Loeper 2006; Ambrosio et al. 2014; Brenier 2011] for a similar situation in cosmology. Moreover, since

$$E(\mu) = -\frac{1}{2} d^2(\mu, \nu_P) + \frac{1}{2}\int |x|^2 \mu + C,$$

where d denotes the Wasserstein L^2-distance, the evolution equation (1-13) can also be seen as a quadratic perturbation (with diffusion) of the "geodesic flow" on the Wasserstein L^2-space, compare to [Ambrosio et al. 2005, Example 11.2.10], which in the one-dimensional case appears in connection to the sticky particle system [Natile and Savaré 2009]. As will be shown in [Berman $\geq$ 2018], the large-time asymptotics of the fully nonlinear evolution equation (1-13) for the probability density ρ_t in $\mathbb{R}^n$ are governed by *traveling wave solutions* in $\mathbb{R}^n$ whose speeds coincide with the negative of the barycenter b_P of the convex body P:

$$\rho_t(x) = \rho(x - b_P t) + o(t), \quad t \to \infty,$$

where the error terms $o(t)$ tends to zero in $L^1(\mathbb{R}^n)$ (and even in relative entropy) and where the limiting profile ρ is uniquely determined from a variant of the Monge–Ampère equation (1-16) together with the condition that its barycenter coincides with the barycenter of the initial data (thus breaking the translation symmetry). In complex-geometric terms, ρ corresponds to a certain canonical Kähler–Einstein metric ω on X with conical singularities "at infinity", playing the role of Calabi's extremal metrics in this context. More generally, as will be elaborated on in [Berman $\geq$ 2018], the results above apply in a more general setting where the measure ν_P is multiplied by a density g, which amounts to replacing the Monge–Ampère equation $\mathrm{MA}(\phi)$ with $g(\nabla\phi)\,\mathrm{MA}(\phi)$ and which from the point of view of scalar conservation laws corresponds to a general concave flux function f (when $n=1$).

5B. ***The microscopic picture: sticky particles in $\mathbb{R}^n$.*** It can be shown that the attractive Newtonian interaction energy in $\mathbb{R}$ is the one-dimensional version of the tropical permanental energy $E^{(N)}_{\mathrm{trop}}(x_1,\dots,x_N)$ appearing in Section 4D. In the general higher-dimensional setting it turns out that a very concrete interpretation of the corresponding EVI gradient flow of $E^{(N)}_{\mathrm{trop}}(x_1,\dots,x_N)$ on $\mathbb{R}^{nN}$ can be given; in particular the particles perform zigzag paths with velocity vectors contained in the polytope $-P$ generalizing the sticky N-particle system on the real line.[3] Moreover, there is a static solution to the corresponding deterministic N-particle system if and only if the "discrete" barycenter of P vanishes,

$$\frac{1}{N}(p_1+\cdots+p_N)=0,$$

which is consistent with the fact that the discrete barycenter can be interpreted as the mean velocity of the particles. In general, any initial configuration of points $(x_1,\dots,x_N)(0)$ is assembled, in a finite time, into a single particle x_*, namely the barycenter of $\{x_1,\dots,x_N\}$, which moves at the mean velocity above. The results in the present paper can also be used to study the large N-limit of this deterministic system (which can be seen as a dissipative version of the Hamiltonian particle system introduced in [Cullen et al. 2007] as a discretization of the semigeostrophic equations). But the key point of our approach is that it allows noise to be added to the particle system. Then the role of the large N-limit of x_* is played by the volume form μ_* of a Kähler–Einstein metric on the toric variety determined by the polytope P (compare the discussion in Section 5A).

Interestingly, a similar particle system on $\mathbb{R}^n$ appears in Brenier's approach [2011; 2016] to the early universe reconstruction problem in cosmology [Frisch and Bec 2001] (in connection to the so-called Zeldovich approximation). In fact, our results can be used to validate the formal large N-limit of the N-particle system with noise introduced in [Brenier 2016, Section 2.3].[4]

5C. ***The complex geometric picture.*** In this final section we provide some complex-geometric motivation for the present paper; a more detailed account, including the relations to the Yau–Tian–Donaldson conjecture and tropicalization, will appear elsewhere

Let X be an n-dimensional compact complex manifold. A metric g on X is said to be Kähler–Einstein if g has constant Ricci curvature and g is Kähler; i.e., in local holomorphic coordinates g can be represented as the real part of the positive definite complex Hessian $\partial\phi(z)/(\partial z_i\partial\bar{z}_j)$ of a local function $\phi(z)$ called the Kähler potential of g. If such a metric g exists with positive Ricci curvature, then X is necessarily a projective algebraic variety which is Fano; i.e., the holomorphic (anticanonical) line bundle $L:=\det(TX)$ over X is positive.

[3] When $n=1$ the dynamics is determined by the property that total mass and momentum is conserved in collisions and that the particles stick together when they collide; see [Brenier and Grenier 1998].

[4] As pointed out in [Brenier 2016, Section 2.3], the formal argument used there, which is based on the classical Freidlin–Wentzel theory, as in [Dawson and Gärtner 1987], would require a Lipschitz bound on the drift.

As shown in [Berman 2013a], a Fano manifold comes with a sequence of canonical N-particle random point processes. The number of particles N arises as the pluriantigenera of X:

$$N = N_k := \dim H^0(X, L^{\otimes k}), \quad k = 1, 2, 3, \ldots,$$

where $H^0(X, L^{\otimes k})$ denotes the complex vector space consisting of the global holomorphic sections of the k-th tensor power of L. The Fano condition ensures that $N_k \to \infty$ as $k \to \infty$. The local density of the corresponding canonical symmetric probability measure $\mu^{(N_k)}$ on X^{N_k} is defined by

$$\rho^{(N_k)}(z_1, \ldots, z_{N_k}) := \frac{1}{Z_{N_k}} \left|\det(z_1, \ldots, z_{N_k})\right|^{-2/k}, \quad \det(z_1, \ldots, z_{N_k}) := \det(s_i(z_j)), \tag{5-2}$$

where $\det(z_1, \ldots, z_{N_k}) \in H^0(X^{N_k}, L^{\otimes k})$ is the Vandermonde-type determinant formed from a given base $s_1, \ldots, s_{N_k}$ in $H^0(X, L^{\otimes k})$ and Z_{N_k} is the corresponding normalization constant ensuring that the probability measure has unit mass (by homogeneity $\rho^{(N_k)}$ is independent of the choice of base). However, since the local density $\rho^{(N_k)}(z_1, \ldots, z_{N_k})$ has singularities (for example when two points on X merge), the normalization constant Z_{N_k} may be infinite, which means that the random point processes are only well-defined if $Z_{N_k} < \infty$. Such a Fano manifold X was called Gibbs stable in [Berman 2013a], where it was shown that the condition can be rephrased in purely algebrogeometric terms (see also [Fujita 2016] for further developments). It was conjectured in [Berman 2013a] that this condition is equivalent to X admitting a (unique) Kähler–Einstein metric (which necessarily has positive Ricci curvature) whose volume form may be recovered as the deterministic large N-limit of the empirical measures of the corresponding random point processes.[5]

The motivation for the present paper comes from a dynamic approach to the latter conjecture where one introduces the interaction energy

$$E^{(N_k)}(z_1, \ldots, z_N) := \frac{1}{k} \log |\det(z_1, \ldots, z_{N_k})|^2,$$

which is attractive, in the sense that it tends to $-\infty$ as two particles merge. Locally, this object is represented by a plurisubharmonic function, but in order to get a globally well-defined function on X^{N_k} one also has to fix a background Kähler metric g on X (representing the first Chern class of X) whose volume form dV_g then induces a metric $\|\cdot\|$ on L which is used to replace the absolute values above. The point is that, if X is Gibbs stable, the canonical probability measure $\mu^{(N_k)}$ on X^{N_k} can then be represented globally as the corresponding Gibbs measure at inverse temperature $\beta = 1$ (which is independent of the choice of metric g),

$$\mu^{(N_k)} = \frac{1}{Z_{N_k}} e^{-E^{(N_k)}} dV_g^{\otimes N_k} \left(= \frac{1}{Z_{N_k}} \|\det\|^{-2/k} dV_g^{\otimes N_k} \right),$$

i.e., as a determinantal point process on X at negative temperature. The different zero-temperature case was studied in [Berman et al. 2011].

[5] The convergence of the processes in the opposite case when the dual $\det(T^*X)$ of $\det(TX)$ is positive was settled in [Berman 2013a] (the limit is then the volume form of the unique Kähler–Einstein metric on X with negative Ricci curvature, whose existence was first established in the seminal works of Aubin and Yau).

At any rate, even if X is not Gibbs stable one can still look at the stochastic gradient flow of $E^{(N_k)}$ on the N_k-fold product of the Riemannian manifold (X, g). From this dynamic perspective Gibbs stability simply means that the corresponding stochastic process has an invariant measure, to wit, $\mu^{(N_k)}$. Accordingly, the natural dynamic generalization of the conjecture referred to above is that a (unique) Kähler–Einstein metric g_{KE} exists precisely when the stochastic gradient flow of $E^{(N_k)}$ admits a stationary measure and then its volume form $dV_{g_{\mathrm{KE}}}$ can be recovered from the joint large N- and large t-limit of the flow. More precisely, conjecturally the large N-limit of the corresponding stochastic gradient flows is described by the complex version of the evolution equation (4-7), obtained by replacing the real Monge–Ampère operator with its complex counterpart. The latter flow is, at least formally, the Wasserstein gradient flow of a free-energy-type functional $F(\mu)$ on $\mathcal{P}_2(X, g)$ and F can be identified with the K-energy functional on the space of Kähler metrics in $c_1(X)$ (using the Calabi–Yau isomorphism) [Berman 2013c]. Unfortunately, the study of the latter flows is plagued by various analytical difficulties stemming from the singularities of $E^{(N_k)}$ and the lack of convexity. For example, even in the simplest case when X is the Riemann sphere, i.e., the one-point compactification of the complex plane $\mathbb{C}$, so that $E^{(N_k)}$ is simply the attractive logarithmic pair interaction between N_k equal charges on $\mathbb{C}$, the convergence of the large N-limit, for a fixed time, is a long-standing open problem (however, see [Fournier and Jourdain 2017] for very recent partial results).

5C1. *The toric setting and its tropicalization.* The complex geometric setting which is relevant to the present paper appears when X is a toric Fano manifold, i.e., X admits a holomorphic action of the real n-torus T such that (X, T) can be realized as an equivariant compactification of the complex torus $\mathbb{C}^{*n}$ (with its standard T-action) [Donaldson 2008]. Such a compactification X is determined by a convex polytope P, which has the property that under the dense embedding of $\mathbb{C}^{*n}$ into X, the complex vector space $H^0(X, L^{\otimes k})$ may be identified with the space of all holomorphic Laurent polynomials $f(z)$ on $\mathbb{C}^{*n}$ of the form

$$f(z) = \sum_{m \in kP \cap \mathbb{Z}^n} a_m z^m$$

(using multi-index notation). In particular, introducing an ordering $m_1, \dots, m_{N_k}$ on the integer points of $kP \cap \mathbb{Z}^n$ gives a basis $s_{m_1}(z), \dots, s_{m_{N_k}}$ of multinomials in $H^0(X, L^{\otimes k})$, which can be used to represent

$$\det(z_1, \dots, z_{N_k}) = \sum_{\sigma \in S_N} (-1)^{\operatorname{sign}(\sigma)} z_1^{m_{\sigma(1)}} \cdots z_{N_k}^{m_{\sigma(N)}}. \tag{5-3}$$

Now, the real vector space $\mathbb{R}^n$ makes its appearance when introducing logarithmic coordinates on $\mathbb{C}^{*n}$, i.e., as the image of the Log map

$$\operatorname{Log} : \mathbb{C}^{*n} \to \mathbb{R}^n, \quad z \mapsto x := (\log |z_1|^2, \dots, \log |z_n|^2),$$

whose fibers are the orbits of the action of T. Using this map, T-invariant metrics on $L \to X$ with positive curvature may be identified with convex functions $\phi(x)$ on $\mathbb{R}^n$ such that $(\partial\phi)(\mathbb{R}^n) \subset P$. In this picture the permanental density $\operatorname{Per}(x_1, \dots, x_{N_k})$ arises as the push-forward to $\mathbb{R}^n$, under the Log map, of the determinant density (5-3). In other words, the smooth convex permanental energy $E^{(N)}_{\mathrm{per}}(x_1, \dots, x_N)$, formula (1-11), on $\mathbb{R}^n$ is an averaged version of the singular plurisubharmonic interaction energy $E^{(N_k)}$ on $\mathbb{C}^{*n}$:

$$E^{(N)}_{\mathrm{per}}(x_1,\dots,x_N)=\frac{1}{k}\log\int_{T^{N_k}} e^{kE^{(N_k)}}\,d\theta^{\otimes N_k}. \tag{5-4}$$

Similarly, its tropical version $E^{(N)}_{\mathrm{trop}}(x_1,\dots,x_N)$ is the piecewise affine convex function on $\mathbb{R}^{nN}$ obtained as the tropicalization of the Laurent polynomial $\det(z_1,\dots,z_{N_k})$ on $\mathbb{C}^{*nN_k}$.[6] Accordingly, Theorem 1.2 should be seen in the light of the well-known philosophy of replacing an elusive complex-geometric problem by a more tractable convex-geometric one, by the process of tropicalization; see, for example, [Itenberg and Mikhalkin 2012].

Appendix: The Otto calculus

In this appendix we briefly recall Otto's [2001] beautiful (formal) Riemannian interpretation of the Wasserstein L^2-metric d_2 on $\mathcal{P}_2(\mathbb{R}^n)$. The material is included with the nonexpert in mind as a motivation for the material on gradient flows on $\mathcal{P}^2(\mathbb{R}^n)$ recalled in Section 2E.

The Otto metric. For simplicity we will consider probability measures of the form $\mu=\rho\,dx$, where ρ is smooth positive everywhere (in order to make the arguments below rigorous one should also specify the rate of decay of ρ at ∞ in $\mathbb{R}^n$). The corresponding subspace of probability measures in $\mathcal{P}_2(\mathbb{R}^n)$ will be denoted by $\mathcal{P}$. First recall that the ordinary "affine tangent vector" of a curve ρ_t in $\mathcal{P}$ at $\rho:=\rho_0$, when ρ_t is viewed as a curve in the affine space $L^1(\mathbb{R}^n)$), is the function $\dot\rho$ on $\mathbb{R}^n$ defined by

$$\dot\rho(x):=\frac{d\rho_t(x)}{dt}_{|t=0}.$$

Next, let us show how to identify $\dot\rho$ with a vector field $v_{\dot\rho}$ in $L^2(\rho\,dx,\mathbb{R}^n)$, which, by definition, is the (nonaffine) "tangent vector" of ρ_t at ρ; i.e., $v_{\dot\rho}\in T_\rho\mathcal{P}$. First, since the total mass of ρ_t is preserved, we have $\int\dot\rho\,dx=0$ and hence there is a vector field v on $\mathbb{R}^n$ solving the continuity equation

$$\dot\rho=-\nabla\cdot(\rho v). \tag{A-1}$$

In geometric terms this means that

$$\rho_t\,dx=(F_t^V)_*(\rho_0\,dx)+o(t), \tag{A-2}$$

where F_t^V is the family of maps defined by the flow of V. Now, under suitable regularity assumptions, $v_{\dot\rho}$ may be defined as the "optimal" vector field v solving the previous equation, in the sense that it minimizes the L^2-norm in $L^2(\rho\,dx,\mathbb{R}^n)$. The Otto metric is then defined by

$$g(v_{\dot\rho},v_{\dot\rho})=\inf_v\int\rho|v|^2\,dx=\int\rho|v_{\dot\rho}|^2\,dx, \tag{A-3}$$

[6] Incidentally, tropicalization may be interpreted as a zero-temperature limit by writing the tropical sum $\max\{a,b\}$ as the limit of $T^{-1}\log(e^{(1/T)a}+e^{(1/T)b})$ as $T\to 0$; compare to the discussion in [Itenberg and Mikhalkin 2012].

which can be seen as the linearized version of the defining formula (2-6) for the Wasserstein L^2-metric. By Hodge theory, the optimal vector field $v_{\dot\rho}$ may be written as $v_{\dot\rho} = \nabla\phi$ for a unique normalized function ϕ on $\mathbb{R}^n$ (under suitable assumptions).

The microscopic point of view. Let us remark that a simple heuristic "microscopic" derivation of the Otto metric can be given using the isometry defined by the empirical measure δ_N (Lemma 2.3). Indeed, given a curve $(x_1(t), \dots, x_N(t))$ in the Riemannian product $\left(X^N, \frac{1}{N}g^{\otimes N}\right)$ with tangent vector $(dx_1(t)/dt, \dots, dx_1(t)/dt)$ at $t = 0$ we can write its squared Riemannian norm at $(x_1(0), \dots, x_N(0))$ as

$$\left\| \left(\frac{dx_1(t)}{dt}, \dots, \frac{dx_1(t)}{dt} \right) \right\|^2 = \int |v|^2 \delta_N(0), \tag{A-4}$$

where $\delta_N(t) := \frac{1}{N}\sum \delta_{x_i(t)}$ and v is any vector field on $X = \mathbb{R}^n$ such that $v(x_i) = dx_i(t)/dt_{|t=0}$. Note that setting $\rho_t := \delta_N(t)$, the vector field v satisfies the push-forward relation (A-2) (with vanishing error term). Moreover, since passing to the quotient X^N/S_N does not effect the corresponding curve ρ_t, minimizing with respect to the action of the permutation group S_N in formula (A-4) corresponds to the infimum defining the Otto metric in formula (A-3).

Relation to gradient flows and drift-diffusion equations. If G is a smooth functional on $\mathcal{P}$ then a direct computations reveals that its (formal) gradient with respect to the Otto metric at ρ corresponds to the vector field $v(x) = \nabla_x(\partial G(\rho)/\partial\rho)$. In other words, the gradient flow of $G(\rho)$ may be written as

$$\frac{\partial \rho_t(x)}{\partial t} = \nabla_x \cdot (\rho v_t(x)), \qquad v_t(x) = \nabla_x \frac{\partial G(\rho)}{\partial \rho}_{|\rho=\rho_t} \tag{A-5}$$

In particular, for the Boltzmann entropy $H(\rho)$, formula (2-1), one gets, since $\partial G(\rho)/\partial\rho = \log\rho$ (using that the mass is preserved), that the corresponding gradient flow is the heat (diffusion) equation and the gradient flow structure then implies that $H(\rho_t)$ is decreasing along the heat equation. Moreover, a direct calculation reveals that H is *convex* on $\mathcal{P}$ in sense that the Hessian of H is nonnegative and hence it also follows from general principles that the squared Riemannian norm $|\nabla H|^2(\rho_t)$ is decreasing. In fact, by definition $|\nabla H|^2(\rho)$ coincides with the Fisher information functional $I(\rho)$, formula (2-1). More generally, the gradient flow of the Gibbs free energy F_β^V is given by the diffusion equation with linear drift $\nabla_x V$,

$$\frac{\partial \rho_t}{\partial t} = \frac{1}{\beta}\Delta_x \rho_t + \nabla_x \cdot (\rho_t \nabla_x V), \tag{A-6}$$

often called the linear Fokker–Planck equation in the mathematical physics literature. The study of the previous flow using a variational discretization scheme on $\mathcal{P}^2(\mathbb{R}^n)$ was introduced in [Jordan et al. 1998] (compare to Section 5C1).

Acknowledgments

It is a pleasure to thank Eric Carlen for several stimulating discussions and whose inspiring lecture in the Kinetic Theory seminar at Chalmers concerning [Blanchet et al. 2012] drew our attention to the Otto calculus and the theory of gradient flows on the Wasserstein space. Thanks also to Luigi Ambrosio for helpful comments on the paper and to Yann Brenier, José Carrillo, Maxime Hauray and Bernt Wennberg for providing us with references and to the referee for many comments which helped to improve the exposition of the paper. This work was supported by grants from the Swedish Research Council, the Knut and Alice Wallenberg Foundation and the European Research Council.

References

[Ambrosio et al. 2005] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Birkhäuser, Basel, 2005. MR Zbl

[Ambrosio et al. 2009] L. Ambrosio, G. Savaré, and L. Zambotti, "Existence and stability for Fokker–Planck equations with log-concave reference measure", *Probab. Theory Related Fields* **145**:3-4 (2009), 517–564. MR Zbl

[Ambrosio et al. 2014] L. Ambrosio, M. Colombo, G. De Philippis, and A. Figalli, "A global existence result for the semi-geostrophic equations in three dimensional convex domains", *Discrete Contin. Dyn. Syst.* **34**:4 (2014), 1251–1268. MR Zbl

[Berman 2013a] R. J. Berman, "Kähler–Einstein metrics, canonical random point processes and birational geometry", preprint, 2013. To appear in *AMS Proceedings of the 2015 Summer Research Institute on Algebraic Geometry*. arXiv

[Berman 2013b] R. J. Berman, "Statistical mechanics of permanents, real-Monge–Ampère equations and optimal transport", preprint, 2013. arXiv

[Berman 2013c] R. J. Berman, "A thermodynamical formalism for Monge–Ampère equations, Moser–Trudinger inequalities and Kähler–Einstein metrics", *Adv. Math.* **248** (2013), 1254–1297. MR Zbl

[Berman 2017] R. J. Berman, "Large deviations for Gibbs measures with singular Hamiltonians and emergence of Kähler–Einstein metrics", *Comm. Math. Phys.* **354**:3 (2017), 1133–1172. MR Zbl

[Berman $\geq$ 2018] R. J. Berman, "Stability of some fully-nonlinear shock waves and toric Kähler–Einstein metrics", in preparation.

[Berman and Berndtsson 2013] R. J. Berman and B. Berndtsson, "Real Monge–Ampère equations and Kähler–Ricci solitons on toric log Fano varieties", *Ann. Fac. Sci. Toulouse Math.* (6) **22**:4 (2013), 649–711. MR Zbl

[Berman and Lu $\geq$ 2018] R. J. Berman and C. H. Lu, "A drift-diffusion equation in complex geometry and convergence towards Kähler–Einstein metrics", in preparation.

[Berman and Önnheim 2016] R. J. Berman and M. Önnheim, "Propagation of chaos for a class of first order models with singular mean field interactions", preprint, 2016. arXiv

[Berman et al. 2011] R. Berman, S. Boucksom, and D. Witt Nyström, "Fekete points and convergence towards equilibrium measures on complex manifolds", *Acta Math.* **207**:1 (2011), 1–27. MR Zbl

[Blanchet et al. 2012] A. Blanchet, E. A. Carlen, and J. A. Carrillo, "Functional inequalities, thick tails and asymptotics for the critical mass Patlak–Keller–Segel model", *J. Funct. Anal.* **262**:5 (2012), 2142–2230. MR Zbl

[Braun and Hepp 1977] W. Braun and K. Hepp, "The Vlasov dynamics and its fluctuations in the $1/N$ limit of interacting classical particles", *Comm. Math. Phys.* **56**:2 (1977), 101–113. MR Zbl

[Brenier 1991] Y. Brenier, "Polar factorization and monotone rearrangement of vector-valued functions", *Comm. Pure Appl. Math.* **44**:4 (1991), 375–417. MR Zbl

[Brenier 2010] Y. Brenier, "Hilbertian approaches to some non-linear conservation laws", pp. 19–35 in *Nonlinear partial differential equations and hyperbolic wave phenomena*, edited by H. Holden and K. H. Karlsen, Contemp. Math. **526**, Amer. Math. Soc., Providence, RI, 2010. MR Zbl

[Brenier 2011] Y. Brenier, "A modified least action principle allowing mass concentrations for the early universe reconstruction problem", *Confluentes Math.* **3**:3 (2011), 361–385. MR Zbl

[Brenier 2016] Y. Brenier, "A double large deviation principle for Monge–Ampère gravitation", *Bull. Inst. Math. Acad. Sin. (N.S.)* **11**:1 (2016), 23–41. MR Zbl

[Brenier and Grenier 1998] Y. Brenier and E. Grenier, "Sticky particles and scalar conservation laws", *SIAM J. Numer. Anal.* **35**:6 (1998), 2317–2328. MR Zbl

[Brenier et al. 2003] Y. Brenier, U. Frisch, M. Hénon, G. Loeper, S. Matarrese, R. Moyhayaee, and A. Sobolevskiĭ, "Reconstruction of the early universe as a convex optimization problem", *Mon. Not. R. Astron. Soc.* **346**:2 (2003), 501–524.

[Chen and Zheng 2013] X. Chen and K. Zheng, "The pseudo-Calabi flow", *J. Reine Angew. Math.* **674** (2013), 195–251. MR Zbl

[Cullen et al. 2007] M. Cullen, W. Gangbo, and G. Pisante, "The semigeostrophic equations discretized in reference and dual variables", *Arch. Ration. Mech. Anal.* **185**:2 (2007), 341–363. MR Zbl

[Dawson and Gärtner 1987] D. A. Dawson and J. Gärtner, "Large deviations from the McKean–Vlasov limit for weakly interacting diffusions", *Stochastics* **20**:4 (1987), 247–308. MR Zbl

[Dembo and Zeitouni 1993] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Jones and Bartlett, Boston, 1993. MR Zbl

[Donaldson 2008] S. K. Donaldson, "Kähler geometry on toric manifolds, and some other manifolds with large symmetry", pp. 29–75 in *Handbook of geometric analysis, I*, edited by L. Ji et al., Adv. Lect. Math. **7**, International Press, Somerville, MA, 2008. MR Zbl

[Fournier and Jourdain 2017] N. Fournier and B. Jourdain, "Stochastic particle approximation of the Keller–Segel equation and two-dimensional generalization of Bessel processes", *Ann. Appl. Probab.* **27**:5 (2017), 2807–2861. MR Zbl

[Frisch and Bec 2001] U. Frisch and J. Bec, "Burgulence", pp. 341–383 in *New trends in turbulence* (Les Houches, 2000), edited by M. Lesieur et al., EDP Sci., Les Ulis, 2001. MR Zbl

[Frisch et al. 2002] U. Frisch, S. Matarrese, R. Mohayaee, and A. Sobolevski, "A reconstruction of the initial conditions of the universe by optimal mass transportation", *Nature* **417** (2002), 260–262.

[Fujita 2016] K. Fujita, "On Berman–Gibbs stability and K-stability of $\mathbb{Q}$-Fano varieties", *Compos. Math.* **152**:2 (2016), 288–298. MR Zbl

[Hairer 2013] M. Hairer, "Solving the KPZ equation", *Ann. of Math.* (2) **178**:2 (2013), 559–664. MR Zbl

[Hauray and Mischler 2014] M. Hauray and S. Mischler, "On Kac's chaos and related problems", *J. Funct. Anal.* **266**:10 (2014), 6055–6157. MR Zbl

[Hiriart-Urruty 1980] J.-B. Hiriart-Urruty, "Extension of Lipschitz functions", *J. Math. Anal. Appl.* **77**:2 (1980), 539–554. MR Zbl

[Hopf 1950] E. Hopf, "The partial differential equation $u_t + uu_x = \mu u_{xx}$", *Comm. Pure Appl. Math.* **3** (1950), 201–230. MR Zbl

[Itenberg and Mikhalkin 2012] I. Itenberg and G. Mikhalkin, "Geometry in the tropical limit", *Math. Semesterber.* **59**:1 (2012), 57–73. MR Zbl

[Jordan et al. 1998] R. Jordan, D. Kinderlehrer, and F. Otto, "The variational formulation of the Fokker–Planck equation", *SIAM J. Math. Anal.* **29**:1 (1998), 1–17. MR Zbl

[Kac 1956] M. Kac, "Foundations of kinetic theory", pp. 171–197 in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, III*, University of California Press, Berkeley, 1956. MR Zbl

[Kardar et al. 1986] M. Kardar, G. Parisi, and Y.-C. Zhang, "Dynamic scaling of growing interfaces", *Phys. Rev. Lett.* **56**:9 (1986), 889–892. Zbl

[Keller and Segel 1970] E. F. Keller and L. A. Segel, "Initiation of slime mold aggregation viewed as instability", *J. Theo. Biol.* **26**:3 (1970), 399–415. Zbl

[Lax 1973] P. D. Lax, *Hyperbolic systems of conservation laws and the mathematical theory of shock waves*, CBMS-NSF Regional Conference Series in Applied Mathematics **11**, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1973. MR Zbl

[Loeper 2006] G. Loeper, "A fully nonlinear version of the incompressible Euler equations: the semigeostrophic system", *SIAM J. Math. Anal.* **38**:3 (2006), 795–823. MR Zbl

[Lott and Villani 2009] J. Lott and C. Villani, "Ricci curvature for metric-measure spaces via optimal transport", *Ann. of Math.* (2) **169**:3 (2009), 903–991. MR Zbl

[Mayer 1998] U. F. Mayer, "Gradient flows on nonpositively curved metric spaces and harmonic maps", *Comm. Anal. Geom.* **6**:2 (1998), 199–253. MR Zbl

[McCann 1997] R. J. McCann, "A convexity principle for interacting gases", *Adv. Math.* **128**:1 (1997), 153–179. MR Zbl

[McKean 1966] H. P. McKean, Jr., "A class of Markov processes associated with nonlinear parabolic equations", *Proc. Nat. Acad. Sci. U.S.A.* **56** (1966), 1907–1911. MR Zbl

[McKean 1967] H. P. McKean, Jr., "Propagation of chaos for a class of non-linear parabolic equations", pp. 41–57 in *Stochastic differential equations* (*Lecture Series in Differential Equations, Session 7, Catholic Univ., 1967*), Air Force Office Sci. Res., Arlington, VA, 1967. MR

[Messer and Spohn 1982] J. Messer and H. Spohn, "Statistical mechanics of the isothermal Lane–Emden equation", *J. Statist. Phys.* **29**:3 (1982), 561–578. MR

[Mischler et al. 2015] S. Mischler, C. Mouhot, and B. Wennberg, "A new approach to quantitative propagation of chaos for drift, diffusion and jump processes", *Probab. Theory Related Fields* **161**:1-2 (2015), 1–59. MR Zbl

[Natile and Savaré 2009] L. Natile and G. Savaré, "A Wasserstein approach to the one-dimensional sticky particle system", *SIAM J. Math. Anal.* **41**:4 (2009), 1340–1365. MR Zbl

[Otto 2001] F. Otto, "The geometry of dissipative evolution equations: the porous medium equation", *Comm. Partial Differential Equations* **26**:1-2 (2001), 101–174. MR Zbl

[Robinson and Ruelle 1967] D. W. Robinson and D. Ruelle, "Mean entropy of states in classical statistical mechanics", *Comm. Math. Phys.* **5** (1967), 288–300. MR

[Schwabl 2002] F. Schwabl, *Statistical mechanics*, Springer, 2002. MR Zbl

[Shandarin and Zel'dovich 1989] S. F. Shandarin and Y. B. Zel'dovich, "The large-scale structure of the universe: turbulence, intermittency, structures in a self-gravitating medium", *Rev. Modern Phys.* **61**:2 (1989), 185–220. MR

[Stroock and Varadhan 1997] D. W. Stroock and S. R. S. Varadhan, *Multidimensional diffusion processes*, Grundlehren der mathematischen Wissenschaften **233**, Springer, 1997. Zbl

[Sznitman 1991] A.-S. Sznitman, "Topics in propagation of chaos", pp. 165–251 in *École d'Été de Probabilités de Saint-Flour XIX—1989*, edited by P. L. Hennequin, Lecture Notes in Math. **1464**, Springer, 1991. MR Zbl

[Villani 2003] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics **58**, Amer. Math. Soc., Providence, RI, 2003. MR Zbl

[Wang and Zhu 2004] X.-J. Wang and X. Zhu, "Kähler–Ricci solitons on toric manifolds with positive first Chern class", *Adv. Math.* **188**:1 (2004), 87–103. MR Zbl

[Weinan et al. 1996] E. Weinan, Y. G. Rykov, and Y. G. Sinai, "Generalized variational principles, global weak solutions and behavior with random initial data for systems of conservation laws arising in adhesion particle dynamics", *Comm. Math. Phys.* **177**:2 (1996), 349–380. MR Zbl

Received 28 Jun 2016. Revised 9 Oct 2017. Accepted 12 Jan 2018.

ROBERT J. BERMAN: robertb@chalmers.se
Department of Mathematical Sciences, Chalmers University of Technology, Göteborg, Sweden

MAGNUS ÖNNHEIM: onnheimm@chalmers.se
Department of Mathematical Sciences, Chalmers University of Technology and University of Götenburg, Göteborg, Sweden

ANALYSIS AND PDE
Vol. 11, No. 6, 2018
dx.doi.org/10.2140/apde.2018.11.1381

THE INVERSE PROBLEM FOR THE DIRICHLET-TO-NEUMANN MAP ON LORENTZIAN MANIFOLDS

PLAMEN STEFANOV AND YANG YANG

We consider the Dirichlet-to-Neumann map Λ on a cylinder-like Lorentzian manifold related to the wave equation related to the metric g, the magnetic field A and the potential q. We show that we can recover the jet of g, A, q on the boundary from Λ up to a gauge transformation in a stable way. We also show that Λ recovers the following three invariants in a stable way: the lens relation of g, and the light ray transforms of A and q. Moreover, Λ is an FIO away from the diagonal with a canonical relation given by the lens relation. We present applications for recovery of A and q in a logarithmically stable way in the Minkowski case, and uniqueness with partial data.

1. Introduction and main results

Let (M, g) be a Lorentzian manifold of dimension $1+n$, $n \geq 2$; i.e., g is a metric with signature $(-1, 1, \dots, 1)$. Suppose a part of ∂M is timelike. An example of M is a cylinder-like domain representing a moving and shape-changing compact manifold in the x-space (if we have fixed time and space variables) with the requirement that the normal speed of the boundary is less than 1; see Section 5.

Denote the wave operator by $\square_g$; in local coordinates $x = (x^0, \dots, x^n)$ it takes the form

$$\square_g := \frac{1}{\sqrt{|\det g|}} \partial_j (\sqrt{|\det g|} g^{jk} \partial_k).$$

Consider the following operator $P = P_{g,A,q}$, which is a first-order perturbation of $\square_g$:

$$P = P_{g,A,q} := \frac{1}{\sqrt{|\det g|}} (\partial_j - iA_j) \sqrt{|\det g|} g^{jk} (\partial_k - iA_k) + q. \tag{1}$$

Here $i = \sqrt{-1}$, A is a smooth 1-form on M, and q is a smooth function on M.

The goal of this work is to study the inverse problem of recovery of g, A and q, up to a data-preserving gauge transformation, from the outgoing Dirichlet-to-Neumann (DN) Λ map on a timelike boundary associated with the wave equation

$$Pu = 0 \quad \text{in } M. \tag{2}$$

We are motivated by applications in relativity but also in applications to classical wave-propagation problems with media moving and/or changing at a speed not negligible compared to the wave speed. We

Stefanov was partly supported by NSF Grants DMS-1301646 and DMS-1600327. Yang was partly supported by NSF Grant DMS-1715178, an AMS Simons travel grant, and startup fund from Michigan State University.
MSC2010: primary 35R30; secondary 35A27, 53B30.
Keywords: Lorentz, DN map, inverse problem, light ray transform, microlocal.

are interested in possible stability results even though some steps in the recovery are inherently unstable. This problem remains widely open. The results we prove are the following. First, we show that one can recover the jet of g, A, q at the boundary (up to a gauge transform) in a Hölder-stable way. Next, we show that one can extract the natural geometric invariants of g, A, q from Λ in a Hölder-stable way. More precisely, Λ recovers the lens relation $\mathcal{L}$ related to g, in stable way. If we know g, the light ray transform $L_1 A$ of A is recovered stably. If g and A are known, the light ray transform $L_0 q$ of q is recovered stably. The lens relation $\mathcal{L}$ is the canonical relation of the Fourier integral operator (FIO) Λ away from the diagonal, and the light ray transforms $L_1 A$ and $L_0 q$ are in fact encoded in the principal and the subprincipal symbols of it. In fact, $\mathcal{L}$ is directly measurable from Λ.

Since the results we prove are local or semilocal (near a fixed lightlike geodesic), and the proofs are microlocal, we do not formulate a global mixed problem for the wave equation at the beginning but we do consider one in Section 5. In fact, existence of solutions of such problems depends on global properties of (M, g), one of which is global hyperbolicity, which are not needed for our weaker formulation and for the proofs. Instead, we define the DN map up to smoothing operators only. In the case when one can prove the existence of a global solution, the true DN map would coincide with ours up to a smoothing error, see Section 5, and our results are not affected by adding smoothing operators.

This problem has a long history in the stationary Riemannian setting, i.e., when $M = [0, T] \times M_0$, where (M_0, g) is a compact Riemannian manifold with boundary, and the metric is $-\mathrm{d}t^2 + g_{ij}(x)\mathrm{d}x^i \mathrm{d}x^j$. The boundary control method [Belishev 1987] and Tataru's uniqueness continuation theorem [1995; 1999] give uniqueness provided that T is greater than a certain sharp critical value T, as shown by Belishev and Kurylev [1992]; see also the survey [Belishev 2007]. Stability however does not follow from such arguments. Stability results for recovering of the metric and lower-order terms appeared in [Stefanov and Uhlmann 1998; 2005b; Montalto 2014; Bellassoued and Dos Santos Ferreira 2011; Bao and Zhang 2014], with [Montalto 2014] covering the general case. A main assumption in those works is that the metric is simple, i.e., that there are no conjugate points and the boundary is strictly convex (a not so essential assumption) and the main technical tool for recovery of the metric is to reduce it to stability for the boundary/lens rigidity problem; see, e.g., [Stefanov and Uhlmann 2005a]. For related results, we refer to [Isakov and Sun 1992; Sun 1990]. Recently, the progress in treating the local rigidity problem allowed results under the more general foliation condition [Stefanov et al. 2016], which allows conjugate points. In any case, some condition is believed to be necessary for stability. It is worth noticing that all inverse (hyperbolic) *scattering* problems for compactly supported perturbations are equivalent to inverse DN map problems.

Recently, there has been increased interest in this problem or in related inverse scattering problems in time-space. Recovery of lower-order time-dependent terms for the Minkowski metric has been studied in [Stefanov 1989; Ramm and Sjöstrand 1991; Ramm and Rakesh 1991; Waters 2014; Salazar 2013; Ben Aïcha 2015; Bellassoued and Ben Aïcha 2017], and for $-\mathrm{d}t^2 + g_{ij}(x)\mathrm{d}x^i \mathrm{d}x^j$ in [Kian et al. 2018]. Eskin [2017] proved that one can recover g, A, q up to a gauge transformation, assuming existence of a global time variable t and analyticity of all coefficients with respect to it. The proof is based on an adaptation of the boundary control methods and the analyticity is needed so that one can still use the unique continuation results in [Tataru 1999]. Stability does not follow from such arguments. Other inverse problems on Lorentzian manifolds are studied in [Kurylev et al. 2014a; 2014b; Lassas et al. 2016]. The inverse scattering problem of recovering a moving boundary is studied in [Cooper and Strauss 1984; Stefanov 1991; Eskin and Ralston 2010]. The first author showed in [Stefanov 1989] that in the case where g is Minkowski and $A = 0$, the problem of recovery of q reduces to the inversion of the X-ray transform in time-space over light rays, which was shown there to be injective for functions tempered in time and uniformly compactly supported in space. In [Lassas et al. 2017], it was shown that the linearized metric problem leads to the inversion of a light ray transform of tensor fields. Such light ray transforms are inherently unstable however because they are smoothing on the timelike cone. They require specialized tools for analyzing the singularities near the lightlike cone, not fully developed in the geodesic case; see [Greenleaf and Uhlmann 1989; 1990a; 1990b]. The light ray transform has been also studied in [Boman and Quinto 1993; Begmatov 2001; Stefanov 2017; Kian 2016].

We describe the main results below. Let $x_0 \in \partial M$ and assume that ∂M is timelike near x_0. Then ∂M with the induced metric is a Lorentzian manifold as well and we choose (locally) one of the two time orientations, which we call future pointing.

Let $f \in \mathcal{E}'(\partial M)$ be supported near x_0 with $\mathrm{WF}(f)$ close to a fixed timelike $(x_0, \xi^{0\prime}) \in T^*\partial M \setminus 0$. We define the *local outgoing* solution operator $f \mapsto u$, defined up to a smoothing operator, as the operator mapping f to the *outgoing* solution u of

$$Pu \in C^\infty \quad \text{in } M \text{ near } x_0, \qquad u|_{\partial M} = f \mod C^\infty. \tag{3}$$

The term “outgoing” here refers to the following. We chose that microlocal solution (parametrix) for which the singularities of the solution are required to propagate along future-pointing bicharacteristics. We refer to the subsection on page 1385 for more details. On the other hand, it is “local” because it solves (3) near x_0 only and this keeps the singularities close enough to ∂M without allowing them to hit ∂M again.

Define the associated *local outgoing Dirichlet-to-Neumann map* as

$$\Lambda^{\mathrm{loc}}_{g,A,q} f = (\partial_\nu u - i\langle A, \nu\rangle u)|_{\partial M}, \tag{4}$$

where ν denotes the unit outer normal vector field to ∂M, and the equality is modulo smoothing operators applied to f. By definition, the $\Lambda^{\mathrm{loc}}_{g,A,q}$ is defined near x_0 only, and in fact, in some conic neighborhood of the timelike $(x_0, \xi^{0\prime})$. Since the latter is arbitrary, $\Lambda^{\mathrm{loc}}_{g,A,q}$ extends naturally to the whole timelike cone on ∂M but we keep it microlocalized near $(x_0, \xi^{0\prime})$ to emphasize what we can recover given microlocal data only.

As we show in Theorem 3.1, $\Lambda^{\mathrm{loc}}_{g,A,q}$ is actually a ΨDO (pseudodifferential operator) on the timelike cone bundle near x_0. The main result about $\Lambda^{\mathrm{loc}}_{g,A,q}$ is Theorem 3.2: a stability estimate about the recovery of the boundary jets of the coefficients.

Let $f \in \mathcal{E}'(\partial M)$ have $\mathrm{WF}(f)$ as above. Let u, as in (3), be the parametrix in a neighborhood of the future-pointing null bicharacteristic issued from the unique future-pointing lightlike covector $(x_0, \xi^0) \in T^*M \setminus 0$ with orthogonal projection $(x_0, \xi^{0\prime})$. Note that the direction of (x_0, ξ^0) and that of the bicharacteristic might be the same or opposite. Assume that this bicharacteristic hits ∂M again, transversely, at point y_0 in the codirection η^0 and let $\eta^{0\prime}$ be the corresponding orthogonal tangential projection on $T^*_{y_0}\partial M$. Then $(y_0, \eta^{0\prime})$ is timelike, as well. Let $\mathcal{U}$ and $\mathcal{V}$ be two small conic timelike neighborhoods in $T^*\partial M \setminus 0$ of $(x_0, \xi^{0\prime})$ and $(y_0, \eta^{0\prime})$, respectively. If $\mathcal{U}$ is small enough, for every timelike $(x, \xi') \in \mathcal{U}$ close to $(x_0, \xi^{0\prime})$, we can define (y, η') in the same way. This defines the *lens relation*

$$\mathcal{L} : \mathcal{U} \longrightarrow \mathcal{V}, \quad \mathcal{L}(x, \xi') = (y, \eta'); \tag{5}$$

see Figure 1. By definition, $\mathcal{L}$ is an even map in the second variable; i.e., $\mathcal{L}(x, -\xi') = (y, -\eta')$. If (x, ξ') is future-pointing (i.e., if the associated vector by the metric is such), then $(x, -\xi')$ is past-pointing but we can interpret $(y, -\eta')$ as the end point of the null geodesic with initial point projecting to $(y, -\eta')$ but moving "backward" with respect to the parameter over it. This property correlates well with Theorem 4.2 since the wave equation has two wave "speeds" of opposite signs.

The map $\mathcal{L}$ is positively homogeneous of order 1 in its second variable. Now, for f as above, let u be an outgoing microlocal solution to (3) near (the projection on the base of) the bicharacteristic γ_0 issued from (x_0, ξ^0) all the way to its second contact with ∂M at y_0. In other words, u is a distribution defined near γ_0 and solving (3) there rather than just near x_0, having future-propagating singularities only. It is unique up to a function smooth near γ_0; see Proposition 4.1. At this point, we assume that $(x_0, \xi^{0\prime})$ is not a fixed point for $\mathcal{L}$, which means that the reflected bicharacteristic does not become a periodic one after the first reflection. This solution is constructed as a solution with the boundary removed, in some neighborhood of γ_0. Since f is smooth near $(y_0, \eta^{0\prime})$ that means there is no singularity of the solution u at $(y_0, \eta^{0\prime})$; therefore, the singularity reflects at y_0. We extend the solution microlocally over a small segment of the reflected ray before reaching ∂M again; see Proposition 4.1 for details. Then we define the *global outgoing DN map* $\Lambda^{\mathrm{gl}}_{g,A,q}$ by (4) again but with the right-hand side localized to V, the projection of $\mathcal{V}$ to the base. In fact, by propagation of singularities, $\Lambda^{\mathrm{gl}}_{g,A,q} f$ has a wave-front set in $\mathcal{V}$ only and we can cut smoothly outside some neighborhood of y_0. The map $\Lambda^{\mathrm{gl}}_{g,A,q}$ is actually just semiglobal because it is the DN map restricted to a solution near one geodesic segment connecting boundary points. Also, it is only defined up to an operator smoothing near (x_0, ξ^0). If a global initial boundary value problem is well defined, $\Lambda^{\mathrm{gl}}_{g,A,q}$ coincides with the associated DN operator up to a smoothing operator; see Section 5. In Theorem 4.2, we prove that $\Lambda^{\mathrm{gl}}_{g,A,q}$ is an FIO associated with the graph of $\mathcal{L}$. In Theorem 4.3, we show that $\Lambda^{\mathrm{gl}}_{g,A,q}$ recovers $\mathcal{L}$ in a stable way, which is also a general property of FIOs associated to a local canonical diffeomorphism.

Another fundamental object is the *light ray transform* L which integrates functions, or more generally tensor fields, along lightlike geodesics. We define L on functions by

$$L_0 f(\gamma) = \int f(\gamma(s))\, \mathrm{d}s, \tag{6}$$

and on covector fields of order 1 by

$$L_1 f(\gamma) = \int \langle f(\gamma(s)), \dot{\gamma}(s) \rangle\, \mathrm{d}s, \tag{7}$$

where $\langle f(\gamma(s)), \dot{\gamma}(s) \rangle = f_j(\gamma(s))\dot{\gamma}^j(s)$ in local coordinates and γ runs over a given set of lightlike geodesics, and we always assume that $\operatorname{supp} f$ is such that the integral is taken over a finite interval. In our results below, γ's in L_0 and L_1 are the maximal geodesics through M connecting boundary points. Unlike the Riemannian case, lightlike geodesics do not have a natural speed-1 parametrization and every rescaling of the parameter along them (even if that rescaling changes from geodesic to geodesic) keeps them being lightlike. The transform L_1 is invariant under reparametrization of the geodesics and can be considered as an integral of $\langle f, \mathrm{d}\gamma\rangle$ over the geodesics. On the other hand, L_0 is not. Despite that freedom, the property $L_0 f = 0$ does not change. One way to parametrize it is to define it locally near a lightlike geodesic hitting a timelike surface at $s = 0$, in our case, ∂M. Then the orthogonal projection $\dot{\gamma}'(0)$ of each such γ on $T\partial M$ (the prime stands for projection) determines $\dot{\gamma}(0)$, and therefore γ, uniquely. To normalize the projections on $T\partial M$, we can choose a timelike covector field Z on $T\partial M$ locally and require $g(\dot{\gamma}, Z) = \mp 1$ for future-/past-pointing directions.

In Theorem 4.4, we show that given g, one can recover $L_1 A$ in a Hölder-stable way; and if we are given g, A, one can recover $L_0 q$ in a Hölder-stable way. Notice that we do not require absence of conjugate points and we do not use Gaussian beams. Instead, we use standard microlocal tools including Egorov's theorem. In Section 5, we consider some cases where L_1 and L_0 can be inverted to derive uniqueness results. As we mentioned above, those transforms are unstable. The reason is that they are microlocally smoothing in the spacelike cone; see, e.g., [Greenleaf and Uhlmann 1990b; Stefanov 2017; Lassas et al. 2017]. Therefore, stable recovery of $L_1 A$ and $L_0 q$ does not imply Hölder-stable recovery of A_1 (up to a gauge transform) and q but allows for weaker logarithmic estimates using the estimate for recovery of q from $L_0 q$ in the Minkowski case proven in [Begmatov 2001], for example. We discuss some of those possible corollaries in Section 5. Recovery of g from $\mathcal{L}$ is an open problem, with some results about the linearized problems obtained recently in [Lassas et al. 2017].

2. Preliminaries

Notation and terminology. In what follows, we denote by U and V the projections of $\mathcal{U}$ and $\mathcal{V}$ onto the base ∂M. We freely assume that $\mathcal{U}$ and $\mathcal{V}$, and therefore, U and V are small enough to satisfy the needed requirements below.

If ξ is a covector based at a point x on ∂M, we denote by ξ' its orthogonal projection to $T_x^*\partial M$. We routinely denote covectors on $T_x^*\partial M$ by placing primes, like ξ', etc., even if a priori such a covector is not a projection of a given one.

Timelike/spacelike/lightlike vectors v are the ones satisfying $g(v,v) < 0$, or $g(v,v) > 0$, or $g(v,v) = 0$, respectively. We identify vectors and covectors by the metric. We choose an orientation in U that we call future pointing (FP). More precisely, we choose some smooth timelike vector Z in U (identified with an open set in the tangent bundle) and we call *future pointing* those timelike vectors v for which $g(v, Z) > 0$. If we have a time variable t, for example, such a choice could be $Z = \partial/\partial t$. In semigeodesic coordinates $(x^0 = t, x)$ near a spacelike hypersurface, see (9) after Lemma 2.3, FP $v = (v^0, v')$ means $v^0 > 0$. Notice that for the associated covector $(\tau, \xi) = gv$, we have $\tau < 0$.

Given a timelike $(x, \xi') \in \mathcal{U}$, assume first that ξ' is FP. Let ξ be the lightlike covector pointing into M with orthogonal projection ξ', identified with the vector $v = g^{-1}\xi$. The geodesic $\gamma_{x,\xi'}(s)$ issued from (x, v), for $s \geq 0$ will be called the FP geodesic issued from (x, ξ'). In Figure 1, left, $v = v_{\mathrm{int}}$ and

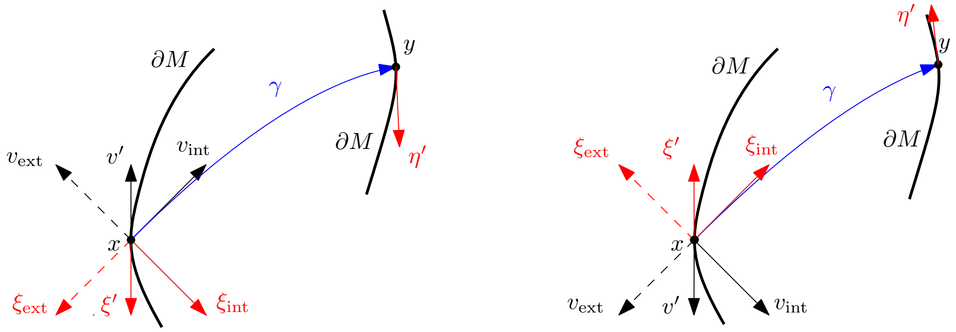

Figure 1. A tangent timelike future-pointing (FP) vector v' on the left, and a past-pointing on the right; and the two lightlike vectors v_{int} and v_{ext} with the same projection, pointing to M and outside M, respectively. The FP geodesic $\gamma = \gamma_{x,\xi'}(s)$ in both cases propagates to the future but on the right, it is determined by negative values of the parameter over it. The corresponding covectors ξ', ξ_{int} and ξ_{ext} are plotted, as well. The lens relation is $\mathcal{L}(x,\xi') = (y,\eta')$.

$\gamma_{x,\xi'}(s) = \gamma$. If (x,ξ') is past pointing, then we choose v to be the lightlike vector projecting to v' pointing to the exterior (v_{ext} in Figure 1, right) and take $\gamma_{x,\xi'}(s)$ for $s \le 0$. By propagation of singularities, a boundary singularity (x,ξ') as above would propagate either along the FP geodesics chosen above, or along the past-pointing ones (or both) that we did not choose. The choice we made reflects the requirement that singularities should propagate to the future only. We call such microlocal solutions *outgoing*. We borrow that term from scattering theory. In the case of the classical formulation of the Riemannian version of this problem, this is guaranteed by the condition $u = 0$ for $t < 0$.

Gauge invariance. There exist some gauge transformations which leave the local and the global versions of the Dirichlet-to-Neumann map $\Lambda_{g,A,q}$ invariant; thus one can only expect to recover the corresponding gauge-equivalence class. To simplify the formulations, we assume that the DN map $\Lambda_{g,A,q}$ is well defined globally on M. In our main theorems, we will apply this to the ΨDO part of $\Lambda_{g,A,q}$ first, and then Φ below needs to be the identity near a fixed point only. For the semiglobal one, we need Φ to be identity near both ends of the fixed lightlike geodesic only. Since the computations below are purely algebraic, the lemmas remain true for the localized maps with obvious modifications.

We will consider two types of gauge transformations in this part. The first one is a diffeomorphism in M which fixes ∂M.

Lemma 2.1. *Let (M,g) be a Lorentzian manifold with boundary as above, let A be a smooth 1-form and q be a smooth function on M. If $\Phi : M \to M$ is a diffeomorphism with $\Phi|_{\partial M} = \mathrm{Id}$, then*

$$\Lambda_{g,A,q} = \Lambda_{\Phi^*g,\Phi^*A,\Phi^*q}.$$

Here Id *is the identity map, and Φ^*g, Φ^*A, Φ^*q are the pullbacks of g, A, q under Φ, respectively.*

Proof. For any $f \in C^\infty(\partial M)$, let u be the solution of $\mathcal{L}_{g,A,q}u = 0$ on M with $u|_{\partial M} = f$. Define $v := \Phi^*u$ as the pull-back of u; then a simple calculation in local coordinates shows that $\mathcal{L}_{\Phi^*g,\Phi^*A,\Phi^*q}v = 0$ and

$v|_{\partial M} = f$. If we write $y = \Phi(x)$ as a local coordinate representation of Φ, then

$$\begin{aligned}\Lambda_{g,A,q} f(y) &= \nu^j(y)\frac{\partial u}{\partial y^j}(y) - i\nu^j(y)A_j(y)u(y)\Big|_{\partial M}\\ &= \nu^j(x)\frac{\partial x^l}{\partial y^j}\frac{\partial v}{\partial y^l} - i\frac{\partial x^l}{\partial y^j}\nu^j(x)\frac{\partial y^k}{\partial x^l}A_k(x)v(x)\Big|_{\partial M}\\ &= \tilde{\nu}^j(x)\frac{\partial v}{\partial x^j}(x) - i\tilde{\nu}^j(x)(\Phi^*A)_j(x)v(x)\Big|_{\partial M} = \Lambda_{\Phi^*g,\Phi^*A,\Phi^*q} f,\end{aligned}$$

where ν and $\tilde{\nu}$ are the unit normals in the y- and x-variables, respectively. The above calculation essentially verifies that $\Lambda_{g,A,q}$ is defined invariantly. Therefore, $\Lambda_{g,A,q} = \Lambda_{\Phi^*g,\Phi^*A,\Phi^*q}$. ☐

Another type of gauge invariance occurs when one makes a conformal change of the metric g. This type of gauge invariance also occurs when g is a Riemannian metric and $\Lambda_{g,A,q}$ is the corresponding Dirichlet-to-Neumann map for the magnetic Schrödinger equation; see [Dos Santos Ferreira et al. 2009, Proposition 8.2].

Lemma 2.2. *Let (M, g) be a Lorentzian manifold with boundary as above, let A be a smooth 1-form and q be a smooth function on M. If φ and ψ are smooth functions such that*

$$\varphi|_{\partial M} = \partial_\nu\varphi|_{\partial M} = 0, \quad \psi|_{\partial M} = 0,$$

then we have

$$\Lambda_{g,A,q} = \Lambda_{e^{-2\varphi}g,A-d\psi,e^{2\varphi}(q-q_\varphi)}$$

where $q_\varphi := e^{\frac{n-2}{2}\varphi}\Box_g e^{\frac{2-n}{2}\varphi}$.

Proof. A direct computation in local coordinates shows that

$$\begin{aligned}e^{\frac{n+2}{2}\varphi}P_{g,A,q}(e^{\frac{2-n}{2}\varphi}u) &= P_{e^{-2\varphi}g,A,e^{2\varphi}(q-q_\varphi)}u,\\ e^{-i\psi}P_{g,A,q}(e^{i\psi}u) &= P_{g,A-d\psi,q}u.\end{aligned}$$

For any $f \in C^\infty(\partial M)$, let u be the solution of $P_{g,A,q}u = 0$ on M with $u|_{\partial M} = f$. Setting $v := e^{\frac{n-2}{2}\varphi}e^{-i\psi}u$, we have

$$\begin{aligned}P_{e^{-2\varphi}g,A-d\psi,e^{2\varphi}(q-q_\varphi)}v &= P_{e^{-2\varphi}g,A-d\psi,e^{2\varphi}(q-q_\varphi)}(e^{\frac{n-2}{2}\varphi}e^{-i\psi}u)\\ &= e^{\frac{n+2}{2}\varphi}P_{g,A-d\psi,q}(e^{-i\psi}u) = e^{\frac{n+2}{2}\varphi}e^{-i\psi}P_{g,A,q}u = 0.\end{aligned}$$

Furthermore, notice that $\nu_{e^{-2\varphi}g} = \nu_g$ by the assumption on φ; thus

$$\begin{aligned}\Lambda_{e^{-2\varphi}g,A-d\psi,e^{2\varphi}(q-q_\varphi)}f &= \nu^j\frac{\partial v}{\partial x^j} - i\nu^j\left(A_j - \frac{\partial\psi}{\partial x^j}\right)v|_{\partial M}\\ &= \nu^j\frac{\partial(e^{\frac{n-2}{2}\varphi}e^{-i\psi}u)}{\partial x^j} - i\nu^j\left(A_j - \frac{\partial\psi}{\partial x^j}\right)(e^{\frac{n-2}{2}\varphi}e^{-i\psi}u)|_{\partial M}\\ &= \nu^j\left(-i\frac{\partial\psi}{\partial x^j}u + \frac{\partial u}{\partial x^j}\right) - i\nu^j i\nu^j\left(A_j - \frac{\partial\psi}{\partial x^j}\right)u|_{\partial M}\\ &= \nu^j\frac{\partial u}{\partial x^j} - i\nu^jA_ju|_{\partial M} = \Lambda_{g,A,q}f.\end{aligned}$$

☐

Gauge equivalent modifications of g, A, q. It is convenient to work in semigeodesic normal coordinates on a Lorentzian manifold. These coordinates are the Lorentzian counterparts of the well-known Riemannian semigeodesic coordinates for Riemannian manifolds with boundary. We formulate the existence of such coordinates in the following lemma.

Lemma 2.3. *Let S be a timelike hypersurface in M. For every $x_0 \in S$, there exist $\varepsilon > 0$, a neighborhood N of x_0 in M, and a diffeomorphism $\Psi : S \cap N \times [0, T) \to N$ such that*

(i) $\Psi(x', 0) = x'$ *for all* $x' \in S \cap N$;

(ii) $\Psi(x', x^n) = \gamma_{x'}(x^n)$, *where* $\gamma_{x'}(x^n)$ *is the unit speed geodesic issued from* x' *normal to* S.

Moreover, if $(x^0, \dots, x^{n-1})$ are local boundary coordinates on S, in the coordinate system $(x^0, \dots, x^n)$, the metric tensor g takes the form

$$g = g_{\alpha\beta} dx^\alpha \otimes dx^\beta + dx^n \otimes dx^n, \quad \alpha, \beta \le n-1. \tag{8}$$

Clearly, $g_{\alpha\beta}$ has a Lorentzian signature as well. If M has a boundary, then S can be ∂M and x^n is restricted to $[0, \varepsilon]$. A proof of the lemma can be found in [Petrov 1969] and is based on the fact that the lines $x' =$ const. and $x^n = s$ are unit speed geodesics; therefore the Christoffel symbols Γ^i_{nn} vanish for all i. We will call such coordinates the semigeodesic normal coordinates. The lemma remains true if S is spacelike with a negative sign in front of $dx^n \otimes dx^n$ in (8) (we replace the index n by 0 below), and this gives us a way to define a time function $t = x^0$ locally, and put the metric in the block form

$$g = -dt^2 + g_{ij}(t, x) dx^i \otimes dx^j, \quad 1 \le i, j \le n, \tag{9}$$

with g_{ij} Riemannian.

Now we use the gauge invariance of $\Lambda_{g,A,q}$ to alter g, A, q without changing the DN map. Three types of modifications are made in the following, labeled as (M1)–(M3) respectively.

Firstly, given two metrics g and $\tilde{g}$, one can choose diffeomorphisms as in Lemma 2.1 to obtain common semigeodesic normal coordinates. In fact, let Ψ and $\tilde{\Psi}$ be diffeomorphisms like in Lemma 2.3 with respect to g and $\tilde{g}$ respectively; then $\tilde{\Psi} \circ \Psi^{-1}$ is a diffeomorphism near ∂M which fixes ∂M. Extend $\tilde{\Psi} \circ \Psi^{-1}$ as in [Palais 1960] to be a global diffeomorphism on M. The properties of Ψ and $\tilde{\Psi}$ ensure that the two metrics g and $(\tilde{\Psi} \circ \Psi^{-1})^* \tilde{g}$ have common semigeodesic normal coordinates near ∂M. Therefore, we may assume:

(M1) If (x', x^n) are the semigeodesic normal coordinates for g, they are also the semigeodesic normal coordinates for $\tilde{g}$.

Secondly, we employ the conformal gauge invariance to replace $\tilde{g}$ with a gauge-equivalent one to obtain some identities which later will help simplify the calculations.

Lemma 2.4. *Let S be either a timelike or a spacelike hyperplane near some point $p_0 \in S$. Given smooth functions $r_2, r_3, \dots$ on S near p_0, there exists a smooth function μ near p_0 with $\mu = 0$, $\partial_\nu \mu = 0$ on S so that if $\hat{\Psi}$ is the diffeomorphism in Lemma 2.3 related to the metric $\hat{g} := e^\mu g$, then*

$$\partial_n^j \det(\hat{\Psi}^* \hat{g}) = r_j, \quad j = 2, 3, \dots,$$

on S near p_0. Here $\partial_n = \partial / \partial x^n$ with $(x^0, \dots, x^n)$ the semigeodesic normal coordinates for g.

Before giving the proof of the lemma, we remark that $(x^0, \dots, x^n)$ may not be the semigeodesic normal coordinates for $\hat{g}$.

Proof. The statement of the theorem is invariant under replacing g by $\Psi^* g$ for any local diffeomorphism Φ which preserves the boundary pointwise. Therefore, we may assume that g is replaced by $\Psi^* g$, i.e., that $x = (x', x^n)$ are semigeodesic coordinates for g.

Note first that the conformal factor does not change the property of a covector being normal to S but rescales the normal derivative and may change the higher-order ones because $\gamma_{x'}$ may change its curvature with respect to the old metric. More precisely, for the vector $e_n = (0, \dots, 0, 1)$ we have $g(e_n, e_n) = \mp 1$ but $\hat{g}(e_n, e_n) = \mp e^{\mu}$. Therefore, for the corresponding normal derivatives we have $\hat{\partial}_\nu = e^{-\frac{\mu}{2}} \partial_\nu = \partial_n$ on $x^n = 0$. Let $\hat{\gamma}_{x'}(s)$ be the normal geodesic at $x' \in S$ with $\dot{\hat{\gamma}}_{x'}$ consistent with the orientation of S, normalized by $\hat{g}(\dot{\hat{\gamma}}_{x'}(s), \dot{\hat{\gamma}}_{x'}(s)) = \mp 1$. Then for every smooth function f,

$$\partial_n^j \widehat{\Psi}^* f(x')|_{x^n=0} = \partial_n^j|_{x^n=0} f(\hat{\gamma}_{x'}(x^n)).$$

For $j = 0, 1$, the results are not affected by the conformal factor and we get

$$\widehat{\Psi}^* f(x')|_{x^n=0} = f(x', 0), \quad \partial_n \widehat{\Psi}^* f(x')|_{x^n=0} = f_n(x', 0).$$

To compute the higher-order normal derivatives, we write

$$\partial_n^2 \widehat{\Psi}^* f(x') = f_{ij} \dot{\hat{\gamma}}_{x'}^i \dot{\hat{\gamma}}_{x'}^j + f_i \ddot{\hat{\gamma}}_{x'}^i \quad \text{on } x^n = 0. \tag{10}$$

Under the conformal change of the metric, the Christoffel symbols are transformed by the law

$$\widehat{\Gamma}_{jk}^k = \Gamma_{ij}^k + \tfrac{1}{2}\delta_i^k \partial_j \mu + \tfrac{1}{2}\delta_j^k \partial_i \mu - g_{ij} \nabla^k \mu.$$

In particular,

$$\widehat{\Gamma}_{nn}^k = \Gamma_{nn}^k + \tfrac{1}{2}\delta_n^k \partial_n \mu + \tfrac{1}{2}\delta_n^k \partial_n \mu - g_{nn} \nabla^k \mu = \delta_n^k \partial_n \mu - \tfrac{1}{2} g^{kl} \partial_l \mu. \tag{11}$$

Therefore, $\widehat{\Gamma}_{nn}^k = 0$ on $x^n = 0$ and (10) reduces to

$$\partial_n^2 \widehat{\Psi}^* f(x') = f_{nn} \quad \text{on } x^n = 0. \tag{12}$$

In a similar way, we may compute $\partial_n^j \widehat{\Psi}^* f(x')$ on $x^n = 0$. The result is $\partial_n^j f$ plus normal derivatives of f of order $j-1$ and less, with coefficients depending on the normal derivatives of μ up to order $j-1$. For our purposes, the exact expression does not matter.

The metric $\hat{g}$ has the form

$$(\widehat{\Psi}^* \hat{g})_{kl} = (\hat{g}_{ij} \circ \widehat{\Psi}) \frac{\partial \widehat{\Psi}^i}{\partial x^k} \frac{\partial \widehat{\Psi}^j}{\partial x^l} = (\hat{g}_{\alpha\beta} \circ \widehat{\Psi}) \frac{\partial \widehat{\Psi}^\alpha}{\partial x^k} \frac{\partial \widehat{\Psi}^\beta}{\partial x^l} + \frac{\partial \widehat{\Psi}^n}{\partial x^k} \frac{\partial \widehat{\Psi}^n}{\partial x^l},$$

where the Greek indices range from 0 to $n-1$ (but not n). In particular,

$$\det \widehat{\Psi}^* \hat{g} = (\det \mathrm{d}\widehat{\Psi})^2 \det(\hat{g} \circ \widehat{\Psi}). \tag{13}$$

We need to understand the structure of $\partial_n^k(\det \mathrm{d}\widehat{\Psi})|_{x^n=0}$ now. For $k=0$, we have $\mathrm{d}\widehat{\Psi}|_{x^n=0} = \mathrm{Id}$. Notice next that

$$\mathrm{d}\widehat{\Psi} = (\partial_0\widehat{\Psi}, \dots, \partial_{n-1}\widehat{\Psi}, \partial_n\widehat{\Psi}), \tag{14}$$

where each partial derivative is a vector. Since by (11), $\partial_n^2\widehat{\Psi}^i = -\widehat{\Gamma}^i_{nn} = 0$ for $x^n=0$,

$$\partial_n(\det \mathrm{d}\widehat{\Psi})|_{x^n=0} = 0.$$

To analyze $k=2$, we notice first that

$$\partial_n^3\widehat{\Psi}^i = -\partial_n\widehat{\Gamma}^i_{nn} = -\partial_n\big(\delta^i_n\partial_n\mu - \tfrac{1}{2}g^{il}\partial_l\mu\big) = -\delta^i_n\mu_{nn} + \cdots,$$

where the dots represent a term involving lower-order ∂_n-derivatives of μ. Using this in (14), we get

$$\partial_n^2(\det \mathrm{d}\widehat{\Psi})|_{x^n=0} = -\mu_{nn}|_{x^n=0}.$$

Reasoning as above, we see that

$$\partial_n^j(\det \mathrm{d}\widehat{\Psi})|_{x^n=0} = -\partial_n^j\mu|_{x^n=0} + \cdots, \tag{15}$$

where the dots represent terms involving normal derivatives of μ (possibly differentiated tangentially) up to order $j-1$.

We will analyze the normal derivatives of $\det(\hat{g}\circ\widehat{\Psi})$ in (13) now. Since $\det\hat{g} = e^{n\mu}\det g$, we get

$$\begin{aligned}\partial_n \det(\hat{g}\circ\widehat{\Psi}) &= \partial_n(e^{(n+1)\mu\circ\widehat{\Psi}}\det g\circ\widehat{\Psi})\\ &= (n+1)\mu_n\det g + \partial_n\det g = \partial_n\det g \quad \text{on } \partial M.\end{aligned} \tag{16}$$

We used the fact that $\mathrm{d}\Psi = \mathrm{Id}$ on ∂M and that $\partial_n\mathrm{d}\Psi = 0$ since $\mathrm{d}\mu = 0$ on ∂M. Therefore, $\partial_n^j\det\hat{g}\circ\widehat{\Psi} = \partial_n^j\det g$ on $x^n=0$ for $j=0,1$.

For the highest-order derivatives, notice that $\partial_n^j\widehat{\Psi}$ involves $\partial_n^{j-1}\mu$ as its highest-order normal μ-derivative, as the arguments leading to (15) show. Differentiating (16), we therefore get

$$\begin{aligned}\partial_n^j\det(\hat{g}\circ\widehat{\Psi}) &= \partial_n^j(e^{(n+1)\mu\circ\widehat{\Psi}}\det g\circ\widehat{\Psi})\\ &= (n+1)(\partial_n^j\mu)\det g + \cdots \quad \text{on } \partial M,\end{aligned} \tag{17}$$

where the dots have the same meaning as in (15).

Using (13) in combination with (15) and (17), we get

$$\partial_n^j(\det\widehat{\Psi}^*\hat{g})|_{x^n=0} = (n-1)(\partial_n^j\mu)\det g + \cdots. \tag{18}$$

To complete the proof of the lemma, we determine the normal derivatives of μ on $x^n=0$ for $j=2,\dots$. We get first $\partial_n^2(\det\widehat{\Psi}^*\hat{g})|_{x^n=0} = (n-1)\mu_{nn}|_{x^n=0}$, which needs to be equal to r_2, and can be solved for μ_{nn}. Then we can determine the tangential derivatives of the latter. After that, we can solve (17) with $j=3$ for μ_{nnn}, etc. To complete the proof, we use Borel's lemma. □

Let g and $\tilde{g}$ be two metrics satisfying (M1) with the two diffeomorphisms Ψ and $\widetilde{\Psi}$ respectively, as in Lemma 2.3. Applying Lemma 2.4 to $S=\partial M$ and $p=x_0$, we can find a metric $\hat{g} := e^{\mu}g$ with $\mu=0$,

$\partial_\nu \mu = 0$ on ∂M such that under the semigeodesic normal coordinates $(x^0, \dots, x^n)$ for g we have

$$\partial_n^j \det(\widehat{\Psi}^* \hat{g}) = \partial_n^j \det(\widetilde{\Psi}^* \tilde{g}), \quad j = 2, 3, \dots,$$

on ∂M. Notice that $(x^0, \dots, x^n)$ are also semigeodesic normal coordinates for $\tilde{g}$ by (M1).

Now consider the metrics $(\widehat{\Psi} \circ \widetilde{\Psi}^{-1})^* \hat{g}$ and $\tilde{g}$. These metrics have common semigeodesic normal coordinates (see the argument following Lemma 2.3), which are $(x^0, \dots, x^n)$. In these coordinates the choice of $\hat{g}$ yields

$$\partial_n^j \det(\widetilde{\Psi}^* \circ (\widehat{\Psi} \circ \widetilde{\Psi}^{-1})^* \hat{g}) = \partial_n^j \det(\widehat{\Psi}^* \hat{g}) = \partial_n^j \det(\widetilde{\Psi}^* \tilde{g}).$$

Thus we may replace g by $(\widehat{\Psi} \circ \widetilde{\Psi}^{-1})^* \hat{g}$ and change A, q accordingly, as in Lemma 2.1 and Lemma 2.2, without affecting $\Lambda_{g,A,q}$. We therefore can assume that g and $\tilde{g}$ satisfy not only (M1), but also:

(M2) In the common semigeodesic normal coordinates (x', x^n),

$$\partial_n^j \det g(x', 0) = \partial_n^j \det \tilde{g}(x', 0), \quad j = 2, 3, \dots.$$

Here we have identified the metrics with their coordinate representations under $\widetilde{\Psi}$.

Thirdly, we make modifications to the 1-form A. Again the modification does not change the gauge-equivalence class of $\Lambda_{g,A,q}$ due to Lemma 2.2.

Lemma 2.5. *Let (M, g) be a Lorentzian manifold with boundary as above, let A be a smooth 1-form and q be a smooth function on M. There exists a smooth function ψ with $\psi|_{\partial M} = 0$ such that in the semigeodesic normal coordinates (x', x^n), $B := A - d\psi$ satisfy*

$$\partial_n^j B_n(x', 0) = 0, \quad j = 0, 1, 2, \dots. \tag{19}$$

Proof. We can find a smooth function ψ with

$$\psi(x', 0) = 0, \qquad \partial_n^{j+1} \psi(x', 0) = \partial_n^j A_n(x', 0), \quad j = 0, 1, 2, \dots.$$

Extend it in a suitable manner so that $\psi \in C^\infty(M)$ with $\psi|_{\partial M} = 0$. Then $B = A - d\psi$ satisfies (19). ☐

As a result we may further assume:

(M3) In the common semigeodesic normal coordinates (x', x^n) of g and $\tilde{g}$,

$$\partial_n^j A_n(x', 0) = \partial_n^j \widetilde{A}_n(x', 0) = 0, \quad j = 0, 1, 2, \dots.$$

3. Boundary stability

We choose the semigeodesic coordinates (x', x^n) near x_0 so that $x_0 = 0$, ∂M locally is given by $x^n = 0$, and the interior of M is given by $x^n > 0$. Let $\xi^{0\prime}$ be a future-pointing timelike covector in $T^*_{x_0} \partial M$ at x_0. On Figure 1, the associated vector would look like v' on the left, while the covector $\xi^{0\prime}$ would have the opposite time direction, like the figure on the right. Let $\chi(x', \xi')$ be a smooth cutoff function with small enough support in $\mathcal{U}$ that is equal to 1 in a smaller conic timelike neighborhood of $(x_0, \xi^{0\prime})$. Assume also that χ is homogeneous in ξ' of order 0.

For

$$f(x') = e^{i\lambda x'\cdot\xi'}\chi(x',\xi'), \tag{20}$$

and for every $N > 0$, we would like to construct a geometric optics approximation of the outgoing solution u of (3) near x_0 in M of the form

$$u_N(x) := e^{i\lambda\phi(x,\xi')}\sum_{j=0}^{N}\frac{1}{\lambda^j}a_j(x,\xi'). \tag{21}$$

The eikonal and the transport equations below are based on the identity

$$e^{-i\lambda\phi}Pe^{i\lambda\phi} = -\lambda^2 g^{jk}(\partial_j\phi)(\partial_k\phi) + i\lambda\Box_g\phi + 2i\lambda g^{jk}\partial_j\phi(\partial_k - iA_k) + P.$$

In M near x_0, the phase function $\phi(x,\xi')$ solves the eikonal equation, which in the semigeodesic coordinates takes the form

$$g^{\alpha\beta}\partial_\alpha\phi\,\partial_\beta\phi + (\partial_n\phi)^2 = 0, \quad \phi|_{x^n=0} = x'\cdot\xi'. \tag{22}$$

With the extra condition $\partial_\nu\phi|_{\partial M} < 0$, (22) is locally uniquely solvable. Moreover, (22) implies

$$\partial_n\phi(x',0) = \xi_n(x',\xi') > 0 \quad \text{for any } (x',\xi')\in\mathcal{U}, \tag{23}$$

where

$$\xi_n(x',\xi') := \sqrt{-g^{\alpha\beta}(x',0)\xi_\alpha\xi_\beta}. \tag{24}$$

Notice that the choice of the sign of ξ_n makes ξ a lightlike future-pointing covector, pointing into M. In Figure 1, the associated vector $v = g^{-1}\xi$ looks like v_{int} on the left.

We recall briefly the method of characteristics for solving the eikonal equation. We first determine $\partial\phi$ on $x^n = 0$ to get (23) or the same equation with a negative square root. We choose one of them, and in this case our choice is determined by the requirement that $\partial\phi$ points into M; see Figure 1. Let now $(q_{x',\xi'}(s), p_{x',\xi'}(s))$ be the null bicharacteristic with $q_{x',\xi'}(0) = x'$, $p_{x',\xi'}(0) = (\xi',\xi_n)$. We think of (x',s) as local coordinates and set $\phi(x',s) = x'\cdot\xi'$. More precisely, ϕ is uniquely determined locally by the requirement to be constant along the null bicharacteristics $q_{x',\xi'}$. Moreover,

$$p(s) = \nabla_x\phi(q(s),\xi'). \tag{25}$$

Since by the Hamilton equations, $\dot q^i(s) = g^{ij}p_j(s)$, we get in particular that $g^{ij}\partial_j\phi\partial_i$ is just the derivative $\partial/\partial s$ along the null bicharacteristic.

In M near x_0, the amplitudes a_0 and a_j, $j = 1,2,\ldots$, solve the following transport equations:

$$Ta_0 = 0, \qquad a_0|_{x_n=0} = \chi, \tag{26}$$

$$iTa_j = -Pa_{j-1}, \quad a_j|_{x_n=0} = 0, \quad j\ge 1. \tag{27}$$

where the operator T is defined as

$$T := 2g^{jk}\partial_j\phi(\partial_k - iA_k) + \Box_g\phi. \tag{28}$$

We prefer to express the bicharacteristics through the geodesics

$$\Gamma(s) := (q_{x',\xi'}(s), p_{x',\xi'}(s)) = (\gamma_{x',\xi'}(s), g\dot{\gamma}_{x',\xi'}(s)).$$

Then along the bicharacteristics, we have

$$T = 2\partial_s - 2i\langle A, p(s)\rangle + \Box_g \phi = 2\mu\partial_s\mu^{-1}, \tag{29}$$

with the integrating factor μ given by

$$\mu(\Gamma(s)) = \exp\left\{-\frac{1}{2}\int_0^s (\Box_g\phi)(\gamma_{x',\xi'}(\sigma))\,\mathrm{d}\sigma\right\}\exp\left\{i\int_0^s \big\langle A(\gamma_{x',\xi'}(\sigma)), \dot{\gamma}_{x',\xi'}(\sigma)\big\rangle\,\mathrm{d}\sigma\right\}. \tag{30}$$

The amplitudes a_j, $j = 0, 1, \dots$, are supported in a neighborhood of the characteristics issued from $x_0 \in \partial M$ in the codirection $\xi(x_0)$. As a result, on some neighborhood of x_0, we have u_N solves $Pu_N = O(\lambda^{-N})$, $u|_{\partial M} = f$.

Theorem 3.1. *$\Lambda^{\text{loc}}_{g,A,q}$ is an elliptic* ΨDO *of order* 1 *in $\mathcal{U}$.*

Proof. Given $f \in \mathcal{E}'(U)$, not related to (20), with a wave-front set as in the theorem, we are looking for an outgoing solution u of $Pu = 0$ near x_0, $u = f$ on U, of the form

$$u(x) = (2\pi)^{-n}\int e^{i\phi(x,\xi')}a(x,\xi')\hat{f}(\xi')\,\mathrm{d}\xi'. \tag{31}$$

The phase ϕ solves the eikonal equation (22) and therefore coincides with ϕ there. We chose the solution which guarantees a locally outgoing u, which corresponds to the positive square root in (24). We are looking for an amplitude a of the form $a \sim \sum_{j=0}^{\infty} a_j(x,\xi')$, where a_j is homogeneous in the ξ'-variable of degree $-j$. The standard geometric optics construction leads to the transport equations (26), (27). Using the standard Borel lemma argument, we construct a convergent series for a. Then u is the microlocal solution (up to a microlocally smoothing operator applied to f) that we used to define $\Lambda^{\text{loc}}_{g,A,q}$. Then $\Lambda^{\text{loc}}_{g,A,q} f = \partial u/\partial\nu|_U$. Since $\phi = x'\cdot\xi'$ on U, we get that $\Lambda^{\text{loc}}_{g,A,q}$ is a ΨDO with symbol

$$-i\xi_n(x',\xi') - \partial_n a|_{x^n=0}.$$

In particular, for the principal symbol we get

$$\sigma_p(\Lambda^{\text{loc}}_{g,A,q})(x',\xi') = -i\xi_n = -i\sqrt{-g^{\alpha\beta}(x')\xi_\alpha\xi_\beta}. \tag{32}$$

We proceed in the same way if (x',ξ') is past pointing.

It remains to show that if we use another locally outgoing solution $\tilde{u}$, the resulting $\widetilde{\Lambda}^{\text{loc}}_{g,A,q}$ would differ by a smoothing operator. This follows by considering $v := u - \tilde{u}$ which is a locally outgoing solution with smooth boundary data, which therefore must be smooth. We omit the details. □

We prove a stable determination result on the boundary next. Let (g, A, q) and $(\tilde{g}, \widetilde{A}, \tilde{q})$ be two triples. Define

$$\delta = \big\|\Lambda^{\text{loc}}_{g,A,q} - \Lambda^{\text{loc}}_{\tilde{g},\widetilde{A},\tilde{q}}\big\|_{H^1(U)\to L^2(U)}, \tag{33}$$

where, as above, $\Lambda^{\mathrm{loc}}_{g,A,q}$ and $\Lambda^{\mathrm{loc}}_{\tilde g,\tilde A,\tilde q}$ are the local DN maps associated with (g, A, q) and $(\tilde g, \tilde A, \tilde q)$ respectively microlocally restricted to a fixed conic neighborhood $\mathcal{U}$ of a timelike future-pointing $(x_0, \xi^{0\prime}) \in T^*U$ with $x_0 \in U \subset \partial M$. As above, we assume that $\xi^{0\prime}$ is future pointing and timelike for both g and $\tilde g$, and that $\mathcal{U}$ is small enough so that it is included in the future timelike cone on T^*U for both metrics. Therefore, in the theorem below, we need to know the DN map microlocally only near a fixed timelike covector on $T^*\partial M$.

Theorem 3.2. *Let (g, A, q) and $(\tilde g, \tilde A, \tilde q)$ be replaced by their gauge equivalent triples satisfying (M1)–(M3). Then for any $\mu < 1$ and $m \geq 0$, and some open neighborhood $U_0 \Subset U$ of x_0,*

(1) $\sup_{x\in\bar U_0,|\gamma|\le m} |\partial^\gamma(g-\tilde g)| \le C\delta^{\frac{\mu}{2m}}$;

(2) $\sup_{x\in\bar U_0,|\gamma|\le m} |\partial^\gamma(A-\tilde A)| \le C\delta^{\frac{\mu}{2m+1}}$;

(3) $\sup_{x\in\bar U_0,|\gamma|\le m} |\partial^\gamma(q-\tilde q)| \le C\delta^{\frac{\mu}{2m+2}}$

are valid whenever $g, \tilde g, A, \tilde A, q, \tilde q$ are bounded in a certain C^k norm in the semigeodesic normal coordinates near x_0 with a constant $C > 0$ depending on that bound with $k = k(m, \mu)$.

Proof. We adapt the proofs in [Montalto 2014; Stefanov and Uhlmann 2005b] in the Riemannian setting. Let Γ_0 be a small conic neighborhood of $\xi^{0\prime}$. We can assume that $\chi = 1$ on $U_0 \times \Gamma_0$. Let f be as in (20). We restrict (x', ξ') to $U_0 \times \Gamma_0$ below. In addition, we normalize ξ' to have unit Euclidean length (in that coordinate system). Since $\partial_\nu = -\partial_n$, the formal Dirichlet-to-Neumann map in the boundary normal coordinates (x', x^n) is given by

$$\Lambda^{\mathrm{loc}}_{g,A,q} f(x') = -e^{i\lambda x'\cdot\xi'}\left(i\lambda\partial_n\phi(x',0,\xi') + \sum_{j=0}^{N}\frac{1}{\lambda^j}(\partial_n - iA_n)a_j(x',0,\xi)\right) + O(\lambda^{-N-1}). \tag{34}$$

The expression for $\Lambda_{\tilde g,\tilde A,\tilde q} f$ is similar, with ϕ and a_j replaced by $\tilde\phi$ and $\tilde a_j$, respectively.

The representation (34) could be derived from (21) but since u there is an approximate solution only, and we defined $\Lambda^{\mathrm{loc}}_{g,A,q}$ microlocally, we need to go back to its definition. To justify (34), notice that by [Taylor 1981, Chapter VIII.7], on the set $\chi = 1$, we have $e^{-i\lambda x'\cdot\xi'}\Lambda^{\mathrm{loc}}_{g,A,q} f$ is equal to the full symbol of $\Lambda^{\mathrm{loc}}_{g,A,q}$ with $\lambda = |\xi|$ and ξ in (34) unit.

In the following, C denotes various constants depending only on M, χ in (20), on the choice of $k \gg 1$ and on the a priori bounds of the coefficients of P in C^k. Solving for $\partial_n\phi$ (resp. $\partial_n\tilde\phi$) in (34) and taking the difference we obtain

$$\partial_n\phi - \partial_n\tilde\phi = \frac{1}{i\lambda}(\Lambda^{\mathrm{loc}}_{g,A,q} f - \Lambda^{\mathrm{loc}}_{\tilde g,\tilde A,\tilde q} f) + \frac{1}{i\lambda}\sum_{j=0}^{N}\frac{1}{\lambda^j}\big[(\partial_n a_j - \partial_n\tilde a_j) - i(A_n a_j - \tilde A_n\tilde a_j)\big] + O(\lambda^{-N-1})$$

in $L^2(U_0)$. Integrating in U_0 yields

$$\|\partial_n\phi - \partial_n\tilde\phi\|_{L^2(U_0)} \le \frac{C}{\lambda}\delta\|f\|_{H^1(U_0)} + \frac{C}{\lambda}. \tag{35}$$

The choice of f in (20) indicates that $\|f\|_{H^1(U_0)} \le C\lambda$. Thus, taking the limit $\lambda \to \infty$ yields

$$\|\xi_n - \tilde{\xi}_n\|_{L^2(U_0)} = \|\partial_n\phi - \partial_n\tilde{\phi}\|_{L^2(U_0)} \le C\delta. \tag{36}$$

From relation (24) we have

$$\left\|(g^{\alpha\beta} - \tilde{g}^{\alpha\beta})\xi_\alpha\xi_\beta\right\|_{L^2(U_0)} = \|\xi_n^2 - \tilde{\xi}_n^2\|_{L^2(U_0)} = \left\|(\partial_n\phi)^2 - (\partial_n\tilde{\phi})^2\right\|_{L^2(U_0)} \le C\delta. \tag{37}$$

We use the following argument here and in several places below: a quadratic form $h^{\alpha\beta}\xi_\alpha\xi_\beta$ is uniquely determined for ξ' in any fixed-in-advance open set Γ on the unit sphere. In fact, one can choose $n(n-1)/2$ vectors ξ' in Γ and then the recovery is done by inverting an isomorphism on $\mathbb{R}^{\frac{n(n-1)}{2}}$, and is therefore stable; see [Dairbekov et al. 2007, Lemma 3.3]. Therefore, (37) implies $\|g - \tilde{g}\|_{L^2(U_0)} \le C\delta$. By interpolation estimates in Sobolev space and Sobolev embedding theorems, we have for any $m \ge 0$ and $\mu < 1$ that

$$\|g - \tilde{g}\|_{C^m(\bar{U}_0)} \le C\delta^\mu \tag{38}$$

provided $k \gg 1$ is sufficiently large.

Second, we show that the first-order normal derivatives of g and the 1-form can be stably determined on the boundary. From (34) we have

$$(\partial_n - i\widetilde{A}_n)\tilde{a}_0 - (\partial_n - iA_n)a_0 = e^{-i\lambda x'\cdot\xi'}(\Lambda_{g,A,q}f - \Lambda_{\tilde{g},\widetilde{A},\tilde{q}}f) + i\lambda(\partial_n\phi - \partial_n\tilde{\phi}) + \sum_{j=1}^{N}\frac{1}{\lambda^j}(\partial_n a_j - \partial_n\tilde{a}_j) + O\Big(\frac{1}{\lambda^{N+1}}\Big) \quad \text{in } L^2(U_0).$$

Estimate as in (35) to obtain

$$\left\|(\partial_n - iA_n)a_0 - (\partial_n - i\widetilde{A}_n)\tilde{a}_0\right\|_{L^2(U_0)} \le C\Big(\delta + \lambda\delta + \frac{1}{\lambda}\Big),$$

which holds for all $\lambda > 0$. In particular, we may choose $\lambda = \delta^{-\frac{1}{2}}$ to minimize the right-hand side; then

$$\left\|(\partial_n - iA_n)a_0 - (\partial_n - i\widetilde{A}_n)\tilde{a}_0\right\|_{L^2(U_0)} \le C\delta^{\frac{1}{2}}. \tag{39}$$

In order to estimate the difference of first-order normal derivatives of the metrics, we consider the transport equation in (26). Since $\chi \equiv 1$ for $x \in U_0$, it follows from the boundary condition in (26) that $\partial_\alpha a_0 = \partial_\alpha\chi = 0$ for $\alpha = 0, \dots, n-1$. Moreover, $g^{nj} = \delta^{nj}$ in the semigeodesic coordinates; thus the transport equation in (26) becomes

$$2\xi_n(\partial_n - iA_n)a_0 - 2iA^\alpha\xi_\alpha + \frac{1}{\sqrt{-\det g}}\partial_n(\sqrt{-\det g}\,\partial_n\phi) + Q(g) = 0, \tag{40}$$

where, as before, Greek indices range from 0 to $n-1$ (but not n). Here $A^\alpha := g^{\alpha\beta}A_\beta$, and $Q(g)$ is a linear combination of tangential derivatives of g, which is defined as follows:

$$Q(g) := \frac{1}{\sqrt{-\det g}}\partial_\alpha(\sqrt{-\det g}\,g^{\alpha\beta})\xi_\beta,$$

where we have used that $\partial_\beta \phi = \xi_\beta$ in U_0, $\beta = 0, \dots, n-1$. As a consequence of (38),

$$Q(g) - Q(\tilde{g}) = O(\delta^{\frac{1}{2}}). \tag{41}$$

Therefore, combining (39), (40) and (41) we obtain

$$\frac{1}{\sqrt{-\det g}}\partial_n(\sqrt{-\det g}\,\partial_n\phi) - \frac{1}{\sqrt{-\det \tilde{g}}}\partial_n(\sqrt{-\det \tilde{g}}\,\partial_n\tilde{\phi}) - 2i(A^\alpha - \widetilde{A}^\alpha)\xi_\alpha = O(\delta^{\frac{1}{2}}).$$

Notice that

$$\frac{1}{\sqrt{-\det g}}\partial_n(\sqrt{-\det g}\,\partial_n\phi) = \frac{1}{2\det g}\partial_n \det g\,\partial_n\phi + \partial_n^2\phi = \frac{\xi_n}{2\det g}\partial_n \det g - \frac{1}{2\xi_n}\partial_n g^{\alpha\beta}\xi_\alpha\xi_\beta$$

is an even function of ξ'. Here in the computation, $\partial_n\phi$ is substituted by ξ_n due to (23) and $\partial_n^2\phi$ is calculated by differentiating the eikonal equation (22). Separating the even and odd parts in ξ' we conclude

$$\left(\frac{\xi_n}{2\det g}\partial_n \det g - \frac{1}{2\xi_n}\partial_n g^{\alpha\beta}\xi_\alpha\xi_\beta\right) - \left(\frac{\tilde{\xi}_n}{2\det \tilde{g}}\partial_n \det \tilde{g} - \frac{1}{2\tilde{\xi}_n}\partial_n \tilde{g}^{\alpha\beta}\xi_\alpha\xi_\beta\right) = O(\delta^{\frac{1}{2}}), \tag{42}$$

$$(A^\alpha - \widetilde{A}^\alpha)\xi_\alpha = O(\delta^{\frac{1}{2}}). \tag{43}$$

From the odd part (43), varying ξ' locally, we get

$$\|A - \widetilde{A}\|_{L^2(U_0)} \le C\delta^{\frac{1}{2}}. \tag{44}$$

To deal with the even part, notice (42) states that

$$\frac{\xi_n}{2\det g}\partial_n \det g - \frac{1}{2\xi_n}\partial_n g^{\alpha\beta}\xi_\alpha\xi_\beta$$

is stably determined of order $O(\delta^{\frac{1}{2}})$. As ξ_n is stably determined on U_0, see (36), their product

$$\begin{aligned}\frac{\xi_n^2}{2\det g}\partial_n \det g - \frac{1}{2}\partial_n g^{\alpha\beta}\xi_\alpha\xi_\beta &= -\frac{1}{2\det g}(\partial_n \det g)g^{\alpha\beta}\xi_\alpha\xi_\beta - \frac{1}{2}\partial_n g^{\alpha\beta}\xi_\alpha\xi_\beta \\ &= -\frac{1}{2}\frac{1}{2\det g}\partial_n(\det g \cdot g^{\alpha\beta})\xi_\alpha\xi_\beta\end{aligned}$$

is also stably determined. Since $\det g$ is known to be stable and away from zero, it follows that $\partial_n h^{\alpha\beta}$ is stable where $h^{\alpha\beta} := (\det g)g^{\alpha\beta}$. Hence, the normal derivative of $g = (\det h)^{\frac{1}{1-n}}h$ is also stably determined; that is,

$$\|\partial_n g - \partial_n \tilde{g}\|_{L^2(U_0)} \le C\delta^{\frac{1}{2}}. \tag{45}$$

Using interpolation and Sobolev embedding theorems, we obtain from (45) and (44) that for any $m \ge 0$ and $\mu < 1$,

$$\|\partial_n g - \partial_n \tilde{g}\|_{C^m(\bar{U}_0)} + \|A - \widetilde{A}\|_{C^m(\bar{U}_0)} \le C\delta^{\frac{\mu}{2}} \tag{46}$$

provided $k \gg 1$ is sufficiently large.

Next we show that the second-order normal derivatives of g, the first-order normal derivatives of A, and the values of q can be stably determined on the boundary. By (34) up to λ^{-1} we obtain

$$\left\|(\partial_n - iA_n)a_1 - (\partial_n - i\widetilde{A}_n)\tilde{a}_1\right\|_{L^2(U_0)} \le C(\lambda^2\delta + \lambda\delta^{\frac{1}{2}} + \lambda^{-1}).$$

Choose $\lambda = \delta^{-\frac{1}{4}}$ to minimize the right-hand side. Then

$$\left\|(\partial_n - iA_n)a_1 - (\partial_n - i\widetilde{A}_n)\tilde{a}_1\right\|_{L^2(U_0)} \le C\delta^{\frac{1}{4}}. \tag{47}$$

Consider the transport equation (27) for a_1. In the semigeodesic coordinates this equation takes the form

$$2i\xi_n(\partial_n - iA_n)a_1 = -\partial_n^2 a_0 + q + O(\delta^{\frac{1}{2}}), \tag{48}$$

where $O(\delta^{\frac{1}{2}})$ represents the stably determined terms of order $O(\delta^{\frac{1}{2}})$. (In fact, $a_1 = 0$ in these expressions by the boundary condition in (27), but it is left here for the convenience of tracking the corresponding terms.) From the estimates (36), (45) and (48) it follows that

$$(-\partial_n^2 a_0 + \partial_n^2\tilde{a}_0) + (q - \tilde{q}) = O(\delta^{\frac{1}{4}}). \tag{49}$$

To obtain an expression of $\partial_n^2 a_0$, we differentiate the transport equation in (26) and evaluate it on U_0:

$$\begin{aligned}\partial_n^2 a_0 &= -\frac{1}{4\det g}\partial_n^2 \det g - \frac{1}{2\xi_n}\partial_n^3\phi + \frac{i}{\xi_n}g^{\alpha\beta}\partial_n A_\alpha\xi_\beta + O(\delta^{\frac{1}{2}})\\ &= -\frac{1}{4\det g}\partial_n^2 \det g + \frac{1}{4\xi_n^2}\partial_n^2 g^{\alpha\beta}\xi_\alpha\xi_\beta + \frac{i}{\xi_n}g^{\alpha\beta}\partial_n A_\alpha\xi_\beta + O(\delta^{\frac{1}{2}}),\end{aligned}$$

where the $O(\delta^{\frac{1}{2}})$ terms are estimated by (38) and (46) and we have used that $\partial_n A_n(x', 0) = 0$ in (M3). Inserting this into (49) and separating the even and odd parts in ξ' gives (notice that $\xi_n = \sqrt{-g^{\alpha\beta}\xi_\alpha\xi_\beta}$ is an even function of ξ'):

$$\left(\frac{1}{4\det g}\partial_n^2\det g - \frac{1}{4\det\tilde{g}}\partial_n^2\det\tilde{g} - \frac{1}{4\xi_n^2}\partial_n^2 g^{\alpha\beta}\xi_\alpha\xi_\beta + \frac{1}{4\tilde{\xi}_n^2}\partial_n^2\tilde{g}^{\alpha\beta}\xi_\alpha\xi_\beta\right) + (q - \tilde{q}) = O(\delta^{\frac{1}{4}}), \tag{50}$$

$$-\frac{i}{\xi_n}g^{\alpha\beta}\partial_n A_\alpha\xi_\beta + \frac{i}{\tilde{\xi}_n}\tilde{g}^{\alpha\beta}\partial_n\widetilde{A}_\alpha\xi_\beta = O(\delta^{\frac{1}{4}}). \tag{51}$$

To deal with (51), we multiply the two terms by ξ_n and $\tilde{\xi}_n$ respectively. This is valid since ξ_n is stably determined in (36). By the argument following (37),

$$\|\partial_n A_\alpha - \partial_n\widetilde{A}_\alpha\|_{L^2(U_0)} \le C\delta^{\frac{1}{2}}.$$

To deal with (50), recall the following matrix identity which is valid for any invertible matrix S:

$$\partial\log|\det S| = \mathrm{tr}(S^{-1}\partial S).$$

Taking $S = g^{\alpha\beta}$ and applying ∂_n^{j-1} we see that

$$\partial_n^j\log(-\det g^{\alpha\beta}) = \partial_n^{j-1}(g_{\alpha\beta}\partial_n g^{\alpha\beta}), \quad j = 1, 2, \ldots.$$

For $j = 2$, it gives

$$g_{\alpha\beta}\partial_n^2 g^{\alpha\beta} = \partial_n^2 \log(-\det g^{\alpha\beta}) - \partial_n g_{\alpha\beta}\partial_n g^{\alpha\beta}.$$

The right-hand side is stably determined by (M2) and (45); we thus get on U_0 that

$$g_{\alpha\beta}\partial_n^2 g^{\alpha\beta} - \tilde{g}_{\alpha\beta}\partial_n^2 \tilde{g}^{\alpha\beta} = O(\delta^{\frac{1}{2}}). \tag{52}$$

On the other hand, remember that the two metrics g and $\tilde{g}$ have been modified to satisfy (M2); thus by (37)

$$\frac{1}{4\det g}\partial_n^2 \det g - \frac{1}{4\det \tilde{g}}\partial_n^2 \det \tilde{g} = \Big(\frac{1}{4\det g} - \frac{1}{4\det \tilde{g}}\Big)\partial_n^2 \det g = O(\delta).$$

This together with (50) gives

$$\Big(-\frac{1}{4\xi_n^2}\partial_n^2 g^{\alpha\beta}\xi_\alpha\xi_\beta + \frac{1}{4\tilde{\xi}_n^2}\partial_n^2 \tilde{g}^{\alpha\beta}\xi_\alpha\xi_\beta\Big) + (q-\tilde{q}) = O(\delta^{\frac{1}{4}}). \tag{53}$$

Again we multiply the terms without the tilde by ξ_n^2 and those with it by $\tilde{\xi}_n^2$; using (24) we have

$$(\partial_n^2 g^{\alpha\beta} + 4q g^{\alpha\beta} - \partial_n^2 \tilde{g}^{\alpha\beta} - 4\tilde{q}\tilde{g}^{\alpha\beta})\xi_\alpha\xi_\beta = O(\delta^{\frac{1}{4}}).$$

By the argument following (37),

$$(\partial_n^2 g^{\alpha\beta} + 4q g^{\alpha\beta}) - (\partial_n^2 \tilde{g}^{\alpha\beta} + 4\tilde{q}\tilde{g}^{\alpha\beta}) = O(\delta^{\frac{1}{4}}).$$

Multiplying those terms without the tilde by $g_{\alpha\beta}$, those with the tilde by $\tilde{g}_{\alpha\beta}$, and then summing up in α, β yields

$$(g_{\alpha\beta}\partial_n^2 g^{\alpha\beta} + 4nq) - (\tilde{g}_{\alpha\beta}\partial_n^2 \tilde{g}^{\alpha\beta} + 4n\tilde{q}) = O(\delta^{\frac{1}{4}}).$$

From (52) we come to the conclusion that

$$\|q-\tilde{q}\|_{L^2(U_0)} \le C\delta^{\frac{1}{4}}.$$

Inserting this into (53) and using the argument following (37),

$$\|\partial_n^2 g^{\alpha\beta} - \partial_n^2 \tilde{g}^{\alpha\beta}\|_{L^2(U_0)} \le C\delta^{\frac{1}{4}}.$$

Putting the estimates on g, A, q together, we have established

$$\|\partial_n^2 g - \partial_n^2 \tilde{g}\|_{L^2(U_0)} + \|\partial_n A_\alpha - \partial_n \widetilde{A}_\alpha\|_{L^2(U_0)} + \|q-\tilde{q}\|_{L^2(U_0)} \le C\delta^{\frac{1}{4}}.$$

As before, interpolation and the Sobolev embedding theorem lead to

$$\|\partial_n^2 g - \partial_n^2 \tilde{g}\|_{C^m(\bar{U}_0)} + \|\partial_n A_\alpha - \partial_n \widetilde{A}_\alpha\|_{C^m(\bar{U}_0)} + \|q-\tilde{q}\|_{C^m(\bar{U}_0)} \le C\delta^{\frac{\mu}{4}}$$

for $m > 0$ and $\mu < 1$. Repeating this type of argument will establish the stability for higher-order derivatives of g, A, q on U_0. □

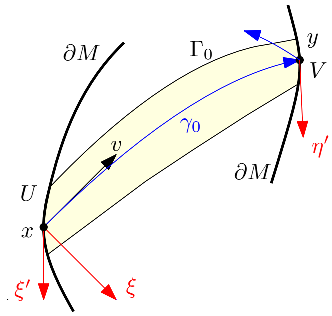

Figure 2. The solution u.

4. Interior stability

The semiglobal microlocal solution. We construct the semiglobal microlocal solution u sketched in the Introduction in the paragraph following (5) and used to define $\Lambda^{\mathrm{gl}}_{g,A,q}$. We recall the assumptions. We fix a timelike $(x_0,\xi^{0\prime})\in T^*\partial M\setminus 0$ and a small conic neighborhood $\mathcal{U}$ of it. We choose a local orientation so that $(x_0,\xi^{0\prime})$ is future pointing. Then there is a unique lightlike $(x_0,\xi^0)\in T^*M\setminus 0$ which projects orthogonally to $(x_0,\xi^{0\prime})$. Let γ_0 be the zero bicharacteristic issued from $(x_0,\xi^{0\prime})$ extended until it hits $T^*\partial M$ again, transversally, by assumption, at some (y_0,η^0) with projection $(y_0,\eta^{0\prime})=\mathcal{L}(x_0,\xi^{0\prime})\in T^*\partial M$. Let $\mathcal{V}=\mathcal{L}(\mathcal{U})$ and denote by U and V the projections $\pi(\mathcal{U})$, $\pi(\mathcal{V})$ of $\mathcal{U}$ and $\mathcal{V}$ onto the base, i.e., their "x-parts". Denote by Γ the union of all zero bicharacteristics issued "from $\mathcal{U}$", i.e., from all future-pointing (x,ξ) with $x\in\partial M$ which have projections on the boundary in $\mathcal{U}$. Let $\Gamma_0:=\pi(\Gamma)\subset M$ be the projection of Γ onto the base; see Figure 2. We assume below, for convenience, that (M,g) is embedded in a slightly larger manifold.

Next proposition says that the microlocal solution u used in the Introduction to define $\Lambda^{\mathrm{gl}}_{g,A,q}$ is well defined.

Proposition 4.1. *If $\mathcal{U}$ is small enough, then for every $f\in\mathcal{E}'(\partial M)$ with* $\mathrm{WF}(f)\in\mathcal{U}$ *there exists a distribution u defined in a neighborhood of $\bar\Gamma_0$ so that $Pu\in C^\infty(\Gamma_0)$, $u|_U-f\in C^\infty(U)$ and $u|_V\in C^\infty$. Moreover, u is unique up to a smooth function in Γ_0.*

Proof. We are looking for a solution u^{inc} of the form (31) with f having a wave-front set in $\mathcal{U}$. Past-pointing codirections can be handled the same way. The solution is the same as in Theorem 3.1 but we are now trying to extend it as far as possible away from ∂M. We know that microlocally, u^{inc} is supported in a small neighborhood of the null bicharacteristic (projecting to a null geodesic on M) issued from (x_0,ξ^0) with ξ^0 future pointing with a projection $\xi^{0\prime}$ on the boundary; i.e., $\xi^0=(\xi^{0\prime},\xi_n(x_0,\xi^0))$, where ξ_n is given by (24). See Figure 1. This follows from the general propagation of singularities theory but in this particular case it can be derived from the fact that T in (28) has its principal part a vector field along such null geodesics, and $\mathrm{WF}(u^{\mathrm{inc}})$ can be analyzed directly with the aid of (31).

Such a solution is guaranteed to exist only near some neighborhood of x_0 because the eikonal equation may not be globally solvable. On the other hand, the solution is still a global FIO applied to the boundary data f. Indeed, it can also be viewed as a superposition of a finite number of local FIOs, each one having a representation of the kind (31). We construct u^{inc} first near ∂M; call it u_1. Then we restrict it to a timelike hypersurface S_1 intersecting the null geodesic $\pi(\gamma_0)$ transversely and we chose S_1 so that the geometric optics construction is still valid along $\pi(\gamma_0)$ until it hits S_1, and a bit beyond it. We take the boundary data at S_1, and solve a new similar problem, by taking the outgoing solution (the future-pointing cone on S_1 is the one determined by $\mathrm{WF}(u_1|_{S_1})$), etc. By compactness arguments, we can cover the whole null geodesic (the projection of γ_0 to the base) until it hits ∂M again. This construction provides solutions (modulo smooth terms) $u_1, \dots, u_k$, each one defined in an open set Γ_k, where $\bigcup_k \Gamma_k$ covers $\bar{\Gamma}_0$. Without loss of generality we may assume that the only intersections of the Γ_k's happen among consecutive ones. Then on S_k, near the intersection with $\pi(\gamma_0)$, we have two microlocal solutions: u_k and u_{k+1}. They have the same traces on S_k modulo a smooth function.

Next, in their common domain of definition, u_k and u_{k+1} coincide up to a smooth function. Indeed, the difference v has smooth trace on S_k and it is outgoing. By the last paragraph of the proof of Theorem 3.1, v is smooth near S_k.

We choose a partition of unity $1 = \sum_k \chi_k$ near $\bar{\Gamma}_0$ subordinate to that cover and set $u^{\text{inc}} = \sum_k \chi_k u_k$. The latter is a microlocal solution (i.e., a solution up to smooth errors) in a neighborhood of $\bar{\Gamma}_0$. Indeed, this is not completely obvious only when $\operatorname{supp} \chi_k$ and $\operatorname{supp} \chi_{k+1}$ intersect but then $u_{k+1} = u_k$ modulo C^∞ and therefore, near such a point, $u^{\text{inc}} = \chi_k u_k + \chi_{k+1} u_{k+1} = u_k$ modulo C^∞, which is a microlocal solution.

We use this argument several times below. This construction is similar to that in [Duistermaat 1996], where it is shown that the Cauchy problem on a spacelike surface gives rise to a global FIO. As a result, one gets a microlocal solution u^{inc} in a neighborhood of $\bar{\Gamma}_0$ (not satisfying the needed boundary conditions on V yet) as a composition of a finite number of FIOs.

We need to reflect u^{inc} at V to satisfy the zero boundary condition. We write the solution u as the sum of the incident wave u^{inc} and the reflected wave u^{ref}: $u = u^{\text{inc}} + u^{\text{ref}}$. The construction of u^{ref} is similar — we start with boundary data $-u^{\text{inc}}|_V$ on V and singularities which propagate into M into the future (the past-future orientation near V is determined by declaring the singularities of u^{inc} on $\mathcal{V}$ coming from the past). We refer to (57) below and the construction following it for more details. The solution u^{ref} needs to be extended to a small neighborhood of the geodesics near γ_0 reflected at V until they leave Γ_0. By choosing $\mathcal{U}$ small enough, we guarantee that the reflected geodesics do not hit ∂M again.

Finally, we prove the uniqueness statement. If u_1 and u_2 are two such solutions, then $v := u_1 - u_2$ is smooth on both U and V. A priori, v can be only singular along bicharacteristics close to γ_0 or its reflection from V. By the argument we used above, v must be smooth in Γ_0 with the possible exception of some neighborhood of V in M, where u^{ref} might be nontrivial. Near V, we know v has smooth Cauchy data. An (easier) adaptation of the same argument shows that v has to be smooth near V as well. Indeed, otherwise, for v, extended as zero outside M, we would get that Pv has singularities conormal to V only, and the microlocal propagation of singularities theorem then would yield that v has no singularities near γ_0 or its reflection. □

Having constructed u, then we define $\Lambda^{\text{gl}}_{g,A,q}$ as in (4) but with the so-constructed u. The uniqueness part of the proposition shows that $\Lambda^{\text{gl}}_{g,A,q}$ is defined up to a smoothing operator.

$\Lambda^{\mathrm{gl}}_{g,A,q}$ recovers the lens relation $\mathcal{L}$ in a stable way.

Theorem 4.2. *Under the assumptions in the Introduction, $\Lambda^{\mathrm{gl}}_{g,A,q}$ is an elliptic FIO of order 1 associated with the (canonical) graph of $\mathcal{L}$.*

Note that we excluded lightlike covectors in $\mathrm{WF}(f)$. This excludes bicharacteristics (geodesics) tangent to ∂M carrying singularities of u. This is where the two Lagrangians (one of them being the diagonal) intersect. We also restricted u to the first reflection shortly after that. Without that, the canonical relations would contain powers of $\mathcal{L}$. The theorem is a direct consequence of the geometric optics construction and propagation of singularities results for the wave equation and can be considered as essentially known.

As a consequence of Theorem 4.2, for every s, we have $\Lambda^{\mathrm{loc}}_{g,A,q}$ maps $H^s(U)$ into $H^{s-1}(U)$ and $\Lambda^{\mathrm{gl}}_{g,A,q}$ maps $H^s(U)$ into $H^{s-1}(V)$. Fixing $s=1$, one may conclude that the natural norms for those two operators are the $H^1\to L^2$ ones. While both operators are bounded in those norms, their dependence on the metric g is not necessarily continuous if we stay in those norms. For $\Lambda^{\mathrm{loc}}_{g,A,q}$, we will see that the principal symbol (and the whole one, in fact) depends continuously on g; and in fact the whole operator does, as well. On the other hand, while the canonical relation of $\Lambda^{\mathrm{gl}}_{g,A,q}$ depends continuously on g, the operator itself does not. This observation was used in [Bao and Zhang 2014]; see also [Stefanov et al. 2016] for a discussion.

Proof of Theorem 4.2. We will analyze first the map $F: f\mapsto u^{\mathrm{inc}}|_S$, where S is a timelike surface as in the proof of Proposition 4.1, and (31) for $u=u_{\mathrm{inc}}$ is valid all the way to it, and a bit beyond it.

Change the coordinates x so that $S=\{x^n=1\}$. This can be done if S is close enough to ∂M. Then (31) with $x=(x',1)$ is a local representation of the FIO F and its canonical relation is given by (see, e.g., [Taylor 1981, Chapter VIII])

$$(\nabla_{\xi'}\phi|_{x^n=1},\xi')\mapsto(x',\nabla_{x'}\phi|_{x^n=1}).$$

By (25), with the momentum p projected to $T^*\{x^n=1\}$, we get that this is the lens relation $\mathcal{L}_1$ from $\mathcal{U}\subset T^*\partial M$ to T^*S (instead of the image being on $T^*\partial M$ again).

We can repeat this finitely many times by choosing S_1, S_2, etc., to get a composition of finitely many canonical relations, starting with $\mathcal{L}_1$; then $\mathcal{L}_2$ maps data on T^*S_1 to T^*S_2, etc. That composition of, say m of them, gives the lens relation from ∂M to S_m. In the final step, we need to take the normal derivative. This shows that the map $f\mapsto\partial_\nu u^{\mathrm{inc}}|_V$ is an FIO of the claimed type.

To prove this for $\Lambda^{\mathrm{gl}}_{g,A,q}$, we need to add $\partial_\nu u^{\mathrm{ref}}|_V$. The latter has an oscillatory representation of the same kind with a different phase; see (57). Its normal derivative on V is the same however and the principal symbol is the same as that of $\partial_\nu u^{\mathrm{ref}}|_V$; see (58) below. □

To prove stable recovery of the lens relation $\mathcal{L}$, we recall that the $H^1\to L^2$ norm of the DN maps is not suitable for measuring how close the canonical relations $\mathcal{L}$ and $\widetilde{\mathcal{L}}$ of the FIOs $\Lambda^{\mathrm{gl}}_{g,A,q}$ and $\Lambda^{\mathrm{gl}}_{\tilde g,\tilde A,\tilde q}$ are. Instead, we formulate stability based on measuring propagation of singularities. Given a properly supported ΨDO R on ∂M near (y_0,η^0), with a principal symbol r_0, we consider $\Lambda^*R\Lambda$, where $\Lambda=\Lambda^{\mathrm{gl}}_{g,A,q}$. By the Egorov theorem, this is actually a ΨDO near (x_0,ξ_0) with a principal symbol $(r_0\circ\mathcal{L})\lambda_0$, where λ_0 is the principal symbol of $\Lambda\Lambda^*$ which depends on g. In this way, we do not recover $\mathcal{L}$ directly; instead we recover functions of $\mathcal{L}$ for various choices of r_0, multiplied by λ_0. Choosing a finite number of R's satisfying some nondegeneracy assumption, we can apply the implicit function theorem to recover $\mathcal{L}$ locally. In fact, we choose below the differential operators

$$\{R_j\} = \{1, y^0, \dots, y^{n-1}, \partial/\partial y^0, \dots, \partial/\partial y^{n-1}\}. \tag{54}$$

Theorem 4.3. *Let $(y^0, \dots, y^{n-1})$ be local coordinates on ∂M near y_0. Let*

$$\begin{gathered}\sum_{j=0}^{n-1} \|\Lambda^* y^j \Lambda - \widetilde{\Lambda}^* y^j \widetilde{\Lambda}\|_{H^2(U)\to L^2(U)} \le \delta, \quad \|\Lambda^*\Lambda - \widetilde{\Lambda}^*\widetilde{\Lambda}\|_{H^2(U)\to L^2(U)} \le \delta,\\ \sum_{j=0}^{n-1} \left\| \Lambda^* \frac{\partial}{\partial y^j}\Lambda - \widetilde{\Lambda}^* \frac{\partial}{\partial y^j}\widetilde{\Lambda}\right\|_{H^3(U)\to L^2(U)} \le \delta,\end{gathered} \tag{55}$$

with $\Lambda := \Lambda^{\mathrm{gl}}_{g,A,q}$, $\widetilde{\Lambda} := \widetilde{\Lambda}^{\mathrm{gl}}_{\tilde g,\widetilde{A},\tilde q}$. Assume that (g, A, q) and $(\tilde g, \widetilde{A}, \tilde q)$ are ε-close to a fixed triple (g_0, A_0, q_0) in a certain C^k norm in the semigeodesic normal coordinates near x_0 and near y_0. Then there exist $k > 0$ and $\mu \in (0,1)$ so that

$$|(\mathcal{L} - \widetilde{\mathcal{L}})(x, \xi')| \le C\delta^{\mu}\sqrt{-g(\xi', \xi')} \quad \textit{for all } (x, \xi') \in \mathcal{U}, \tag{56}$$

if $\mathcal{U}$ and $\varepsilon > 0$ are small enough.

A few remarks:

(a) The square-root term is just a homogeneity factor.

(b) The cotangent bundle $T^*\partial M$ is not a linear space; therefore the difference $\mathcal{L} - \widetilde{\mathcal{L}}$ makes sense in fixed coordinates only.

(c) The norms in (55) are the natural one since the operators we subtract there are ΨDOs of orders 2 and 3, respectively.

(d) The norms in (55) are equivalent to studying the quadratic forms $(\Lambda f, R_j \Lambda f) - (\widetilde{\Lambda} f, R_j \widetilde{\Lambda} f)$.

(e) One could reduce the number of the R_j's to $2n-2$; in fact, $R_0 = 1$ in (55) is not needed, as it follows from Remark 4.8, since we can recover η'/η_n and use the fact $\eta = (\eta', \eta_n)$ is a null covector.

We prove Theorem 4.3 at the end of this section.

Stable recovery of the light ray transforms of A and q. Let, as in the Introduction, $\xi^0 \in T_{x_0}M \setminus 0$ be the future-pointing lightlike covector whose projection on $T^*\partial M \setminus 0$ is the timelike covector $\xi^{0\prime}$ as in the definition of the semiglobal DN map. Let $\gamma_0 := \gamma_{x_0,\xi^{0\prime}}$ be the lightlike geodesic issued from (x_0, ξ^0) which intersects ∂M at another point y_0. Let V be a neighborhood of y_0 containing all endpoints of future-pointing geodesics issued from $\bar{\mathcal{U}}$. Choose and fix any parametrization of the lightlike geodesics close to γ_0 by normalizing ξ'. This defines a hypersurface $\mathcal{U}_0$ in $\mathcal{U}$. The theorem below holds if $\mathcal{U}$ is a small enough neighborhood of $(x_0, \xi^{0\prime})$, and therefore $\mathcal{U}_0$ is small enough, as well. Then L_1 and L_0 are well defined on $\mathcal{U}_0$.

Theorem 4.4. *Fix a Lorentzian metric g and (x_0, ξ^0) satisfying the assumptions above. Let (A, q) and $(\widetilde{A}, \tilde{q})$ be two pairs of magnetic and electric potentials. Define $\delta := \|\Lambda^{\mathrm{gl}}_{g,A,q} - \Lambda^{\mathrm{gl}}_{g,\widetilde{A},\tilde{q}}\|_{H^1(U)\to L^2(V)}$. Then*:

(a) *For any $\mu < 1$ and $m \geq 0$, the following estimates are valid for some integer N whenever g, A, $\widetilde{A}$, q, $\tilde{q}$ are bounded in a certain C^k norm*:

$$\|L_1(A-\widetilde{A}) - 2\pi N\|_{C^m(\bar{\mathcal{U}}_0)} \leq C\delta^\mu.$$

(b) *Under the a priori condition $\|A-\widetilde{A}\|_{C^1(M)} \leq \delta_1$ for some $\delta_1 > 0$, for any $0 < \mu < \mu'$ and $m \geq 0$, the following estimate is valid whenever g, A, $\widetilde{A}$, q, $\tilde{q}$ are bounded in a certain C^k norm*:

$$\|L_0(q-\tilde{q})\|_{C^m(\bar{\mathcal{U}}_0)} \leq C(\delta^\mu + \delta_1^\mu).$$

If there are no conjugate points along γ_0, the proof can be done using a geometric optics construction of the kind (21) but with a different phase in (21) all the way along that geodesic and taking the normal derivative in V. Since we do not want to assume there are no conjugate points along γ_0, we will proceed in a somewhat different way.

The fact that we cannot rule out the case $N \neq 0$ based on those arguments can be considered as a manifestation of the Aharonov–Bohm effect. If $\widetilde{A}$ and A are a priori close, then $N = 0$.

We start with a preparation for the proof of the theorem. Consider first the geometric optics parametrix of the kind (31) of the outgoing solution u like in the previous section. We assume that the boundary condition f has a wave-front set in the timelike cone on the boundary, and for simplicity, assume that it is in the future-pointing one ($\tau < 0$ in local coordinates for which $\partial/\partial t$ is future pointing). Assume at this point that the construction is valid in some neighborhood of the maximal γ_0. We microlocalize all calculations below there. All inverses like D^{-1}, etc., below are microlocal parametrices and the equalities between operators are modulo smoothing operators in the corresponding conic microlocal neighborhoods depending on the context.

The construction is the same as that in the previous section, but this time the outgoing solution u is constructed near the bicharacteristic issued from $(x_0, \xi^{0\prime})$ all the way to y_0. Since the solution can reach the other side of the boundary, we need to reflect it at the boundary to satisfy the zero boundary condition. We write the solution u as the sum of the incident wave u^{inc} and the reflected wave u^{ref}: $u = u^{\mathrm{inc}} + u^{\mathrm{ref}}$ where

$$\begin{aligned} u^{\mathrm{inc}}(x) &= (2\pi)^{-n} \int e^{i\phi(x,\xi')}(a_0^{\mathrm{inc}} + a_1^{\mathrm{inc}} + R^{\mathrm{inc}})(x,\xi')\hat{f}(\xi')\,\mathrm{d}\xi', \\ u^{\mathrm{ref}}(x) &= (2\pi)^{-n} \int e^{i\phi^{\mathrm{ref}}(x,\xi')}(a_0^{\mathrm{ref}} + a_1^{\mathrm{ref}} + R^{\mathrm{ref}})(x,\xi')\hat{f}(\xi')\,\mathrm{d}\xi'. \end{aligned} \tag{57}$$

Here the phase function ϕ^{ref} solves the same eikonal equation as ϕ does but satisfies the boundary condition $\phi^{\mathrm{ref}}|_V = \phi$. It differs from ϕ by the sign of its (exterior) normal derivative $\partial\phi/\partial\nu = -\partial\phi^{\mathrm{ref}}/\partial\nu > 0$ on V. The amplitudes are of orders 0 and -1, respectively, and satisfy

$$\begin{aligned} T^{\mathrm{inc}} a_0^{\mathrm{inc}} &= 0, & a_0^{\mathrm{inc}}|_U &= \chi, \\ T^{\mathrm{ref}} a_0^{\mathrm{ref}} &= 0, & a_0^{\mathrm{ref}}|_V &= -a_0^{\mathrm{inc}}|_V, \\ iT^{\mathrm{inc}} a_1^{\mathrm{inc}} &= -P a_0^{\mathrm{inc}}, & a_1^{\mathrm{inc}}|_U &= 0, \\ iT^{\mathrm{ref}} a_1^{\mathrm{ref}} &= -P a_0^{\mathrm{ref}}, & a_1^{\mathrm{ref}}|_V &= -a_1^{\mathrm{inc}}|_V, \end{aligned}$$

where T^{inc} and T^{ref} are the transport operators defined in (28), related to the corresponding phase function, and the remainder terms are of order -2. Replace A and $\widetilde{A}$ with their gauge-equivalent field satisfying (M3) on V. This does not change their light ray transforms. A direct computation, which can be justified as (34), yields

$$\Lambda^{\mathrm{gl}}_{g,A,q} f = (2\pi)^{-n} \int e^{i\phi(x,\xi')} \big(2i(\partial_\nu\phi)a_0^{\mathrm{inc}} + 2i(\partial_\nu\phi)a_1^{\mathrm{inc}} + \partial_\nu(a_0^{\mathrm{inc}} + a_0^{\mathrm{ref}}) + a_{-1}\big)\hat{f}(\xi')\,\mathrm{d}\xi', \quad (58)$$

where a_{-1} is of order -1 and ϕ and the amplitudes are restricted to $x \in V$.

The expression (58) allows us to factorize $\Lambda^{\mathrm{gl}}_{g,A,q}$ as $\Lambda^{\mathrm{gl}}_{g,A,q} = 2N_0 D$ modulo FIOs of order 0 associated with the same canonical relation, where Df is the trace of u^{inc} on V (a "Dirichlet-to-Dirichlet map") and N_0 is the DN map $\Lambda^{\mathrm{loc}}_{g,0,0}$ but localized in V. Note that replacing A and q in N_0 by zeros or not contributes to lower-order error terms. Let D_0 be the operator D related to $A = 0$, $q = 0$. Let N_0^{-1} and D_0^{-1} be microlocal parametrices of those operators which are actually parametrices of the local Neumann-to-Dirichlet map and the incoming Dirichlet-to-Dirichlet one from V to U. Then

$$D_0^{-1} N_0^{-1} \Lambda^{\mathrm{gl}}_{g,A,q} = 2D_0^{-1} D \mod S^{-1} \quad (59)$$

is a ΨDO of order 0.

In the next lemma, we do not assume that the geometric optics construction is valid along the whole γ_0.

Lemma 4.5. *The operator $D_0^{-1} N_0^{-1} \Lambda^{\mathrm{gl}}_{g,A,q}$ is a ΨDO of order 0 in $\mathcal{U}$ with principal symbol*

$$2\exp\{iL_1 A(\gamma_{x',\xi'})\}, \quad (60)$$

where $\gamma_{x',\xi'}$ is the future-pointing lightlike geodesic issued from x' in direction ξ with projection ξ'.

Proof. By (59), we need to find the principal symbol of $D_0^{-1} D$.

The transport equation for a_0^{inc} is

$$\big[2g^{jk}(\partial_j\phi)(\partial_k - iA_k) + \Box_g\phi\big]a_0^{\mathrm{inc}} = 0, \quad a_0^{\mathrm{inc}}|_U = 1.$$

As explained right after (25), $g^{jk}(\partial_j\phi)\partial_k$ is the tangent vector field along the null geodesic $\gamma_{x',\xi'}$. Therefore, with $\Gamma(s) := (\gamma_{x',\xi'}(s), g\dot{\gamma}_{x',\xi'}(s))$, as before, on the set $\chi = 1$ we get $a_0^{\mathrm{inc}} = \mu$; see (30). That is,

$$a_0^{\mathrm{inc}}(\Gamma(s)) = \exp\left\{-\frac{1}{2}\int_0^s (\Box_g\phi)(\Gamma(\sigma))\,\mathrm{d}\sigma\right\} \exp\left\{i\int_0^s A_k \circ \gamma_{x',\xi'}(\sigma)\dot{\gamma}^k_{x',\xi'}(\sigma)\,\mathrm{d}\sigma\right\}. \quad (61)$$

Take $s = s(x,\xi)$ so that $\gamma_{x',\xi'}(s) \in V$ to get

$$(a_0^{\mathrm{inc}} \circ \mathcal{L})(x',\xi') = \exp\big\{-\tfrac{1}{2}L(\Box_g\phi)(x',\xi') + iL_1A(x',\xi')\big\},$$

where we use the coordinates (x',ξ') to parametrize the lightlike geodesics locally, and the definition of $L(\Box_g\phi)$ is clear from (61).

To construct a representation for D_0^{-1}, note first that when $A = 0$, the term involving $L_1 A$ is missing above. We look for a parametrix of the incoming solution of $\Box_g u = 0$ with boundary data $u = h$ on V with $\mathrm{WF}(h) \subset \mathcal{V}$ of the form

$$u(x) = (2\pi)^{-n} \int e^{i\phi(x,\xi')} b(x,\xi')\hat{f}(\xi')\,\mathrm{d}\xi', \quad (62)$$

where ϕ is the same phase as in the first equation in (57) and f, not related to (20), depends on h as below. The amplitude b solves the transport equation along the same bicharacteristics (with different coefficients since $A = 0$, $q = 0$) with the initial condition

$$b|_V = a^{\mathrm{inc}}|_V,$$

where a^{inc} is the full amplitude in the first equation in (57). Restricted to V, the map $f \to u|_V$ is just Df. Then to satisfy $u = h$ on V, we need to solve $Df = h$, i.e., to take $f = D^{-1}h$ microlocally.

To illustrate the argument below better, suppose that we are solving the ODE

$$y' + ay = 0, \quad y(0) = 1,$$

from $t = 0$ to $t = 1$, where $a = a(t)$. Then we solve

$$y_1' + a_1 y_1 = 0, \quad y_1(1) = y(1),$$

where $a_1 = a_1(t)$. A direct calculation yields

$$y(t) = \exp\left\{-\int_0^t a(s)\,\mathrm{d}s\right\}, \quad y_1(t) = \exp\left\{-\int_1^t a_1(s)\,\mathrm{d}s\right\} y(1).$$

In particular,

$$y_1(0) = \exp\left\{-\int_0^1 (a_1(s) - a(s))\,\mathrm{d}s\right\}.$$

We apply those argument to the transport equation to get

$$b|_U = \exp\{iL_1 A(\gamma_{x',\xi'})\}.$$

Then

$$D_0^{-1} Df = (2\pi)^{-n} \int e^{x'\cdot\xi'} \exp\{iL_1 A(\gamma_{x',\xi'})\} \hat{f}(\xi')\,\mathrm{d}\xi'.$$

This proves the lemma under the assumption that the geometric optics construction is valid in a neighborhood of γ_0.

To prove the theorem in the general case, we use the partition argument we used in Proposition 4.1. Let $S_1, \dots, S_k$ be small timelike surfaces intersecting γ_0 in increasing order from U to V so that the geometric optics construction is valid in a neighborhood of each segment of γ_0 cut by two consecutive surfaces of the sequence $\{U, S_1, \dots, S_k, V\}$. This determines Dirichlet-to-Dirichlet maps D_1 from U to S_1, then D_2, from S_1 to S_2, etc., until D_{k+1} from S_k to V. Then $D = D_{k+1}D_k \cdots D_1$. Similarly, $D_0 = D_{0,k+1}D_{0,k} \cdots D_{0,1}$. Then (59) is still valid and takes the form

$$D_0^{-1} N_0^{-1} \Lambda^{\mathrm{gl}}_{g,A,q} = 2 D_{0,1}^{-1} \cdots D_{0,k}^{-1} D_{0,k+1}^{-1} D_{k+1} D_k \cdots D_1 \mod S^{-1}.$$

By the first part of the proof, $D_{0,k+1}^{-1} D_{k+1}$ is a ΨDO on V with principal symbol $\exp\{iL_1^{(k+1)}A\}$, where $L_1^{(k+1)}$ is the light ray transform L_1 restricted to geodesics between S_k and V. Then we apply Egorov's theorem, see [Hörmander 1985b, Theorem 25.2.5], to conclude that $D_{0,k}^{-1}(D_{0,k+1}^{-1}D_{k+1})D_k$ is a ΨDO with a principal symbol that of $D_{0,k+1}^{-1}D_{k+1}$, pulled back by $\mathcal{C}_{k+1}$, the canonical relation between S_k

and V, multiplied by the principal symbol of $D_{0,k}^{-1}D_k$. The result is then (60) without the factor of 2 with the integration between S_k (through S_{k+1}) to V. Repeating this argument several times, we complete the proof of the lemma. □

Stability of the light ray transform of the magnetic field.

Proof of Theorem 4.4(a). We have

$$\|D_0^{-1}N_0^{-1}(\Lambda^{\mathrm{gl}}_{g,A,q}-\Lambda^{\mathrm{gl}}_{g,\tilde A,\tilde q})\|_{H^1(U)} \le C\|\Lambda^{\mathrm{gl}}_{g,A,q}-\Lambda^{\mathrm{gl}}_{g,\tilde A,\tilde q}\|_{H^1(U)\to L^2(V)} = C\delta. \tag{63}$$

Set $R := D_0^{-1}N_0^{-1}(\Lambda^{\mathrm{gl}}_{g,A,q}-\Lambda^{\mathrm{gl}}_{g,\tilde A,\tilde q})$. By Lemma 4.5, R is a ΨDO in $\mathcal{U}$ of order 0 with principal symbol

$$r_0(x',\xi') = 2\exp\{iL_1(\tilde A - A)(\gamma_{x',\xi'})\},$$

and we have $\|R\|_{H^1(V)} \le C\delta$, by (63). We need to derive that r_0 is "small" in $\mathcal{U}$, as well. We essentially did that in the proof of Theorem 3.2. Choose f as in (20). By [Taylor 1981, Chapter VIII.7], on the set $\chi = 1$, we know $e^{-i\lambda x'\cdot\xi'}Rf$ is equal to the full symbol of $\Lambda^{\mathrm{loc}}_{g,A,q}$ with $\lambda = |\xi|$ and ξ in (34) bounded, say, unit. Therefore,

$$r_0(x',\xi') = e^{-i\lambda x'\cdot\xi'}Rf + O(1/\lambda) \tag{64}$$

in C^k for every k. Since $\|f\|_{L^2} = C$ and $\|f\|_{H^1} \sim \lambda$, (63) yields

$$\|r_0(\,\cdot\,,\xi')\|_{H^1(U)} \le C\lambda\delta + C/\lambda,$$

uniformly for ξ in some neighborhood of $\xi^{0\prime}$. With a little more effort one can remove λ from $C\lambda\delta$ but this is not needed. Take $\lambda = \delta^{-\frac12}$ to get

$$\|\exp\{iL_1(\tilde A - A)(\gamma_{x',\xi'})\}\|_{H^1(\bar{\mathcal{U}}')} \le C\delta^{\frac12}.$$

Using interpolation estimates, we can replace the H^1 norm by any other one at the expense of lowering the exponent on the right from $\frac12$ to another positive one, if k in Theorem 4.4 is large enough. Since $|e^{iz}-1| < \varepsilon$ implies $|z - 2\pi N| < C\varepsilon$ for some integer N, this proves part (a) of the theorem. □

Stability of the light ray transform of the potential.

Proof of Theorem 4.4(b). First, we will reduce the problem to the case $\tilde A = A$. For $\Lambda^{\mathrm{gl}}_{g,\tilde A,\tilde q} - \Lambda^{\mathrm{gl}}_{g,A,\tilde q}$, we get a representation as in (58) with a principal symbol with seminorms $O(\delta_1^{\mu'})$, since we can use interpolation estimates to estimate the higher derivatives of $\tilde A - A$. Apply a parametrix $(\Lambda^{\mathrm{gl}}_{g,A,\tilde q})^{-1}$ to that difference to get a ΨDO Q of order 0 microlocally supported in $\mathcal{U}$. If the geometric optics construction is valid all the way from U to V, we get as in the proof of (a) that $Qf = O(\delta_1^{\mu'}) + O(1/\lambda)$ in H^1. This implies the same estimate for $\|(\Lambda^{\mathrm{gl}}_{g,\tilde A,\tilde q} - \Lambda^{\mathrm{gl}}_{g,A,\tilde q})f\|_{L^2}$. In the general case, we can prove the same estimate as in the proof of (a). We will use this later and for now, we assume $\tilde A = A$.

Lemma 4.6. *The operator $D^{-1}N_0^{-1}(\Lambda^{\mathrm{gl}}_{g,A,\tilde q} - \Lambda^{\mathrm{gl}}_{g,A,q})$ is a ΨDO of order -1 on U with principal symbol*

$$2[L_0(\tilde q - q)]\circ\gamma_{x',\xi'}, \tag{65}$$

where $\gamma_{x',\xi'}$ is the future-pointing lightlike geodesic issued from x' in direction ξ with projection ξ'.

Proof. Assume first that the geometric optics construction is valid in a neighborhood of the whole γ_0. In the amplitude

$$-2i(\partial_\nu\phi)a_0^{\mathrm{inc}} - 2i(\partial_\nu\phi)a_1^{\mathrm{inc}} + \partial_\nu(a_0^{\mathrm{inc}} + a_0^{\mathrm{ref}}) + a_{-1}$$

in (58), the terms $-2i(\partial_\nu\phi)a_0^{\mathrm{inc}}$ and $\partial_\nu(a_0^{\mathrm{inc}} + a_0^{\mathrm{ref}})$ do not depend on q; see (61). The other two terms depend on q but they are of different orders. Therefore,

$$(\Lambda^{\mathrm{gl}}_{g,A,\tilde q} - \Lambda^{\mathrm{gl}}_{g,A,q})f = (2\pi)^{-n}\int e^{i\phi(x,\xi')}\big(-2i(\partial_\nu\phi)(\tilde a_1^{\mathrm{inc}} - a_1^{\mathrm{inc}}) + (a_{-1} - \tilde a_{-1})\big)\hat f(\xi')\,\mathrm{d}\xi'|_V. \quad (66)$$

The order of the FIO above is zero. As in the previous proof, we can represent this as a composition of $2N_0$ with the operator $\widetilde D - D$ (the difference of two such Dirichlet-to-Dirichlet maps):

$$\Lambda^{\mathrm{gl}}_{g,A,\tilde q} - \Lambda^{\mathrm{gl}}_{g,A,q} = 2N_0(\widetilde D - D) \quad (67)$$

modulo FIOs of order -2. That operator $\widetilde D - D$ is an FIO with a symbol, compare with (66),

$$\sigma(\widetilde D - D) = -2i(\tilde a_1^{\mathrm{inc}} - a_1^{\mathrm{inc}}) + a_{-2}, \quad (68)$$

with a_{-2} of order -2.

To compute a_1^{inc}, recall the transport equation for a_1^{inc}

$$\big[2g^{jk}\partial_j\phi(\partial_k - iA_k) + \square_g\phi\big]a_1^{\mathrm{inc}} = iPa_0^{\mathrm{inc}}, \quad a_1^{\mathrm{inc}}|_U = 0, \quad (69)$$

where

$$iPa_0^{\mathrm{inc}} = iP_{g,A,0}a_0^{\mathrm{inc}} + iqa_0^{\mathrm{inc}}.$$

The first term on the right is independent of q. By (29), (30), with $\Gamma(s)$ as in (61), we get

$$\begin{aligned} a_1^{\mathrm{inc}}(\Gamma(s)) &= \frac{ia_0^{\mathrm{inc}}}{2}\int_0^s \frac{1}{a_0^{\mathrm{inc}}}[P_{g,A,0}a_0^{\mathrm{inc}} + qa_0^{\mathrm{inc}}]\circ\Gamma(\sigma)\,\mathrm{d}\sigma \\ &= \frac{ia_0^{\mathrm{inc}}}{2}\int_0^s \left[\frac{1}{a_0^{\mathrm{inc}}}P_{g,A,0}a_0^{\mathrm{inc}} + q\right]\circ\Gamma(\sigma)\,\mathrm{d}\sigma. \end{aligned} \quad (70)$$

The potential q depends on x only, so $q\circ\Gamma(s) = q\circ\gamma(s)$. In (70), only the last term depends on q and is an integral of q over lightlike geodesics multiplied by an elliptic factor. Note that the integral, as well as a_1^{inc}, are homogeneous of order -1 in ξ', as they should be.

We go back to (68) now. Using (70), the terms involving $P_{g,A,0}$ and $P_{g,\widetilde A,0}$ cancel below and we get

$$\sigma(\widetilde D - D)\circ\mathcal{L} = ia_0^{\mathrm{inc}}L_0(\tilde q - q) + a_{-2}, \quad (71)$$

where a_{-2} is a symbol of order -2, different from the one above.

Similarly to (59), we have

$$D^{-1}N_0^{-1}(\Lambda^{\mathrm{gl}}_{g,A,\tilde q} - \Lambda^{\mathrm{gl}}_{g,A,q}) = 2D^{-1}(\widetilde D - D) \bmod S^{-2}. \quad (72)$$

Therefore, we need to compute the principal symbol of $2D^{-1}(\widetilde D - D)$. Let R be a ΨDO in U with principal symbol r_{-1} given by (65). Then, in $\mathcal{U}$, DR is an FIO of the type (62) with $x\in V$ with the same phase function and a principal amplitude b_0 solving $Tb_0 = 0$, $b_0|_U = r_{-1}$. By (29), the solution restricted to $x \in V$ is given by $\mu r_{-1} \circ \mathcal{L}^{-1}|_V$. Recall that $\mu = a_0^{\text{inc}}$. By (71), this is $2\sigma(\tilde{D} - D)$ modulo symbols of order -2. Therefore, $DR = 2(\tilde{D} - D)$ modulo FIOs of order -2. This proves the lemma under the assumption that the geometric optic construction is valid along the whole γ_0.

In the general case, we repeat the arguments of Lemma 4.5. We represent D and $\tilde{D}$ as a composition $D = D_{k+1} \cdots D_1$, and similarly for $\tilde{D}$. We will do the first step. Consider $2(D_2D_1)^{-1}(\tilde{D}_2\tilde{D}_1 - D_2D_1)$. We have

$$2(D_2D_1)^{-1}(\tilde{D}_2\tilde{D}_1 - D_2D_1) = 2D_1^{-1}D_2^{-1}\big((\tilde{D}_2 - D_2)\tilde{D}_1 + D_2(\tilde{D}_1 - D_1)\big)$$
$$= D_1^{-1}R_2\tilde{D}_1 + R_1 = D_1^{-1}R_2D_1 + R_1$$

modulo FIOs of order -2, where $R_j = 2D_j^{-1}(\tilde{D}_j - D_j)$, $j = 1, 2$. We apply Egorov's theorem to $D_1^{-1}R_2D_1$ to conclude that it is a ΨDO on U with a principal symbol equal to the sum of two terms as in (65) with L_0 taken over the geodesic segments between U and S_1 first, and S_1 and S_2 second. The sum is equal to (65) with L_0 taken over the union of those segments. Repeating this argument to include D_2, etc., completes the proof of the lemma. □

We finish the proof of part (b) as we did that for part (a). Set

$$R = D^{-1}N_0^{-1}(\Lambda^{\text{gl}}_{g,A,\tilde{q}} - \Lambda^{\text{gl}}_{g,A,q}).$$

It is a ΨDO of order -1 rather than of order 0 as in (a). The analog of (63) is still true. If, as above, r_{-1} is the principal symbol of R, then by Lemma 4.6,

$$r_{-1}(x', \xi') = -2i[L_0(\tilde{q} - q)] \circ \gamma_{x',\xi'} = \lambda e^{-i\lambda x'\cdot\xi'} Rf + O(1/\lambda)$$

with f as in (20); compare with (64). Then

$$\|r_{-1}(\cdot, \xi')\|_{H^1(U)} \le C\lambda^2\delta + C/\lambda.$$

Choose $\lambda = \delta^{-\frac{1}{3}}$ to get $\|r_{-1}(\cdot, \xi')\|_{H^1(U)} \le C\delta^{\frac{1}{3}}$. This completes the proof of the theorem. □

Proof of the stable recovery of the lens relation.

Proof of Theorem 4.3. We use the notation above. Recall the remark preceding Theorem 4.3 above. The operator $\Lambda^* P \Lambda$ is a ΨDO with a principal symbol $(p_0 \circ \mathcal{L})\lambda_0$. Take $p = p_0 = 1$ as in (54) to recover λ_0 first. Knowing the latter, we recover $p_j \circ \mathcal{L}$ for $j = 1, \dots, 2n-1$; see (54). That gives us (y, η') in (5) as functions of (x, ξ'). Therefore, we reduce the stability problem to the following: show that the principal symbol of a ΨDO A of order m is determined by $A : H^m \to L^2$ in a stable way which is resolved by the lemma below, see also (35), (36). Note that the lemma is a bit more general than what we need since $\{P_j\}$ are simple multiplication and differentiation operators.

Lemma 4.7. *Let Q be ΨDO in $\mathbb{R}^n$ with kernel supported in $K \times K$, where $K \subset \mathbb{R}^n$ is compact. Let q_m be its principal symbol homogeneous of order m. Then*

$$\|q_m(\cdot, \xi)\|_{L^2} \le C|\xi|^m \|Q\|_{H^m \to L^2}$$

for all $\xi \neq 0$ with $C > 0$ depending on K only.

Proof. Take $f = e^{ix\cdot\xi}\chi(x)$, where $\chi \in C_0^\infty$ equals 1 in a neighborhood K. Then for x in a neighborhood of K, we have $Qf(x) = e^{ix\cdot\xi}(q_m(x,\xi) + r(x,\xi))$ with $r \in S^{m-1}$. We have

$$\frac{|\xi|^m}{C} \leq \|f\|_{H^m} \leq C|\xi|^m$$

for $|\xi| \geq 1$. Therefore, for such ξ,

$$\frac{C_1\|Qf\|_{L^2}}{\|f\|_{H^m}} \geq \left\|\frac{q_m(\cdot,\xi)}{|\xi|^m}\right\|_{L^2} - \frac{C_2}{|\xi|}.$$

Take the limit $|\xi| \to \infty$ along radial rays to complete the proof. □

We complete the proof of Theorem 4.3 with the aid of Lemma 4.7. We recover first the L^2-norms with respect to x of $\mathcal{L}(x,\xi) - \widetilde{\mathcal{L}}(x,\xi)$ uniformly in ξ (in fixed coordinates); we can choose $\mu = 1$ then. Using standard interpolation estimates, we can estimate the L^∞ norm of $\mathcal{L}(x,\xi) - \widetilde{\mathcal{L}}(x,\xi)$ with $\mu < 1$ in (56), using the a priori bounds on g and $\tilde{g}$ in some C^k, $k \gg 1$, which imply similar bounds on $\mathcal{L}$ and $\widetilde{\mathcal{L}}$. □

Remark 4.8. The symbol λ_0 can be computed. Since we do not use this formula, we will sketch the proof only. Using Green's formula, as in the proof of [Stefanov and Uhlmann 1998, Proposition 2.1], we can show that $2N_{\mathcal{V}} \cong D^*\Lambda$, where $\cong$ stands for equality modulo lower-order terms, and $N_{\mathcal{V}}$ is N above with the subscript $\mathcal{V}$ indicating that it acts microlocally in that set. The same proof implies that Λ^* is the DN map associated with the incoming solution, i.e., the one which starts from $\mathcal{V}$ microlocally and hits $\mathcal{U}$. Therefore, $\Lambda^* \cong 2N_{\mathcal{U}}D^{-1}$, where $N_{\mathcal{U}}$ now acts in $\mathcal{U}$. Those two identities and Egorov's theorem imply $\lambda_0 = -4(\xi_n \circ \mathcal{L})\xi_n$, where ξ_n is the function defined in (32).

5. Applications and examples

We start with a partial but still general enough case. We follow [Hörmander 1985a, §24.1]. Let M be a Lorentzian manifolds with a timelike boundary ∂M. Assume that t is a real-valued smooth function on M so that the level surfaces $t =$ const. are compact and spacelike. For every $a < b$, the (compact) "cylinder" $M_{ab} = \{a \leq t \leq b\}$ (assuming $[a,b]$ is in the range of t) has a boundary consisting of the spacelike surfaces $t^{-1}(a)$, $t^{-1}(b)$ and $\partial M \cap M_{ab}$ which intersect transversely. This is a generalization of $[0,T]\times\Omega$ in the Riemannian case. By [Hörmander 1985a, Theorem 24.1.1], the following problem is well posed:

$$Pu = 0 \quad \text{in } M, \qquad u|_{t<a} = 0, \qquad u|_{\partial M} = f$$

with $f \in H^s(\partial M)$, $s \geq 1$, $f = 0$ for $t < a$, with a unique solution $u \in H^s(M)$ vanishing for $t < a$. Moreover, the map $f \mapsto u$ is continuous. Then the Dirichlet-to-Neumann map $\Lambda_{g,A,q}$ defined as in (4), is well defined.

Let $x_0 \in U_0 \Subset U \subset \partial M$ be as in Theorem 3.2. Let χ be a properly supported ΨDO cutoff of order 0 localizing near some timelike covector over $x_0 \in U_0$. Since there is a globally defined time function, there are no periodic lightlike geodesics. Then $\chi\Lambda_{g,A,q}\chi$ can be taken as $\Lambda^{\text{loc}}_{g,A,q}$ and Theorem 3.2 applies. If we know a priori that $\Lambda_{g,A,q} : H^1_{(0)}(\partial M) \to L^2(\partial M)$ is continuous, where the subscript (0) indicates functions vanishing for $t = 0$, then we can replace $\Lambda^{\text{loc}}_{g,A,q}$ by $\Lambda_{g,A,q}$ in (33) and therefore, in Theorem 3.2.

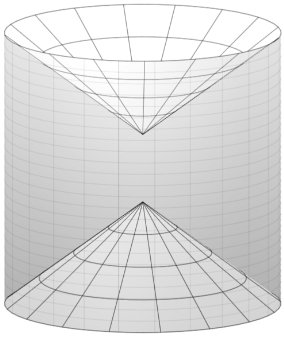

Figure 3. The DN map, with g Minkowski, on the lateral boundary of the cylinder determines a potential and a magnetic field up to $d\phi$ inside the cylinder but outside the two characteristic cones.

Similarly, with suitable ΨDO cutoffs χ_1 and χ_2, we can take $\Lambda^{\mathrm{gl}}_{g,A,q} = \chi_1 \Lambda_{g,A,q} \chi_2$, under the assumptions of Theorem 4.4. And again, if we know that $\Lambda_{g,A,q} : H^1_{(0)}(\partial M) \to L^2(\partial M)$ is continuous, we can remove the cutoffs. The results with the cutoffs are actually stronger.

Some special subcases are discussed below. They recover and extend the uniqueness results in [Stefanov 1989; Ramm and Sjöstrand 1991; Ramm and Rakesh 1991; Waters 2014; Salazar 2013; Ben Aïcha 2015; Bellassoued and Ben Aïcha 2017], and some of the stability results there. Using the results in this paper with the support theorems about the light ray transform in [Stefanov 2017; RabieniaHaratbar 2017], we can get new partial data results.

Example 5.1. Let q be a unknown potential but assume that the metric and the magnetic fields are known. Restrict the DN map to M_{ab} for some $a < b$. Then we can recover $L_0 q$ in a stable way as in Theorem 4.4 over all timelike geodesics intersecting the lateral boundary transversely at their endpoints. If g is real-analytic, then we can apply the results in [Stefanov 2017] to recover q in the set covered by those geodesics under an additional foliation condition. Note that in contrast, the results in [Eskin 2017] require A and q to be analytic in time.

Example 5.2. In the example above, assume that g is Minkowski, and $M_{ab} = [0, T] \times \bar{\Omega}$ for some bounded smooth $\Omega \subset \mathbb{R}^n$. By Theorem 3.2, we can recover $L_1 A$ and $L_0 q$ over all lightlike geodesics (lines) $l_{z,\theta} = \{(t, x) = (s, z + s\theta) : s \in \mathbb{R}\}$, $(z, \theta) \in \mathbb{R}^n \times S^{n-1}$, not intersecting the top and the bottom of the cylinder. By [Stefanov 2017], we can recover q in the set covered by those lines. By [RabieniaHaratbar 2017], we can recover A up to $d\phi$, $\phi = 0$ on $[0, T] \times \partial\Omega$ in that set as well.

For example, if Ω is the ball $B(0, 1) = \{x : |x| < 1\}$, the DN map recovers uniquely q and A, up to a gauge transform, in the cylinder $[0, T] \times \bar{B}(0, 1)$ with the upward characteristic cone with base $\{0\} \times B(0, 1)$ and the downward with base $\{T\} \times B(0, 1)$ removed; see Figure 3. If $T \le 2$, those two cones intersect; otherwise they do not but the result holds in both cases. This is the possibly reachable region from $[0, T] \times \partial\Omega$; thus the results are sharp since no information about the complement can be obtained by the finite speed of propagation.

This extends further the uniqueness part of the results in [Stefanov 1989; Ramm and Sjöstrand 1991; Ramm and Rakesh 1991; Waters 2014; Salazar 2013; Ben Aïcha 2015; Bellassoued and Ben Aïcha 2017]. Using the stability estimate in [Begmatov 2001] about L_0, and the logarithmic estimate for L_1 in [Salazar 2014], we can use Theorem 4.4 to recover the results in [Salazar 2014]. One important improvement however is that for uniqueness, we do not assume that A and q are known outside $[0, T]$, or that $T = \infty$ because the uniqueness results in [Stefanov 2017; RabieniaHaratbar 2017] do not require the function or the vector field to be compactly supported in time.

Example 5.3. A partial-data case of Example 5.2 is the following. Let $\Gamma \subset \partial\Omega$ be relatively open, and assume that $\partial\Omega$ is strictly convex. Assume that we know the DN map for f supported in $[0, T] \times \Gamma$, and we measure Λf there, as well. Then we can recover q (for all $n \geq 2$) and A for $n \geq 3$, up to a gauge transform, in the set covered by the lightlike lines hitting $[0, T] \times \partial\Omega$ in $[0, T] \times \Gamma$ at their both endpoints. When $n = 2$, the recovery of A up to a potential $\mathrm{d}\phi$ requires that if we know $L_1 A$ for some lightlike $l_{z,\theta}$, we also know it for $l_{z,-\theta}$, see [RabieniaHaratbar 2017], and this is the reason we restricted n to $n \geq 3$. Those local uniqueness results for the DN maps are new.

Example 5.4. In a recent work [Bellassoued and Ben Aïcha 2017], an inverse problem for the wave operator

$$P := \partial_t^2 + a(t, x)\partial_t - \Delta + b(t, x)$$

with real-valued a, b is studied. The coefficient b causes absorption. We do not restrict A and q to be real-valued, so we can take

$$A = \left(\frac{i}{2}a(t, x), 0, \dots, 0\right), \quad q = -\tfrac{1}{2} i\,\partial_t a(t, x) + b(t, x);$$

then P in (1) is the same as the one above. Then Theorem 4.4 proves unique recovery of A, q up to the gauge transform $A \mapsto A - \mathrm{d}\psi$ with $\psi = 0$ on $[0, T] \times \partial\Omega$. Since A is restricted to the class of covector fields with spatial components zero, we must have $\psi = \psi(t)$. However, then $\psi = 0$ for $x \in \partial\Omega$ implies $\psi \equiv 0$. Therefore, the logarithmic and the double logarithmic stability estimates in [Bellassoued and Ben Aïcha 2017] for a and for b which are about the DN map can be obtained by Theorem 4.4 combined with the stability estimates in [Begmatov 2001; Salazar 2014]. We can get new uniqueness results however with partial data as in the previous examples. In the Riemannian case studied by Montalto [2014] we can allow an absorption term as well to obtain, up to a gauge transform, stable recovery of a Riemannian simple metric in a generic class, a magnetic field, a potential and an absorption term from the DN map.

Acknowledgments

We would like to thank Matti Lassas for providing some of the references and for the useful discussions. We also thank the referee for the multiple suggestions which helped greatly improve the readability of this paper.

References

[Bao and Zhang 2014] G. Bao and H. Zhang, "Sensitivity analysis of an inverse problem for the wave equation with caustics", *J. Amer. Math. Soc.* **27**:4 (2014), 953–981. MR Zbl

[Begmatov 2001] A. K. Begmatov, "A certain inversion problem for the ray transform with incomplete data", *Sibirsk. Mat. Zh.* **42**:3 (2001), 507–514. In Russian; translated in *Siberian Math. J.* **42**:3 (2001), 428–434. MR Zbl

[Belishev 1987] M. I. Belishev, "An approach to multidimensional inverse problems for the wave equation", *Dokl. Akad. Nauk SSSR* **297**:3 (1987), 524–527. In Russian. MR Zbl

[Belishev 2007] M. I. Belishev, "Recent progress in the boundary control method", *Inverse Problems* **23**:5 (2007), R1–R67. MR Zbl

[Belishev and Kurylev 1992] M. I. Belishev and Y. V. Kurylev, "To the reconstruction of a Riemannian manifold via its spectral data (BC-method)", *Comm. Partial Differential Equations* **17**:5-6 (1992), 767–804. MR Zbl

[Bellassoued and Ben Aïcha 2017] M. Bellassoued and I. Ben Aïcha, "Stable determination outside a cloaking region of two time-dependent coefficients in an hyperbolic equation from Dirichlet to Neumann map", *J. Math. Anal. Appl.* **449**:1 (2017), 46–76. MR Zbl

[Bellassoued and Dos Santos Ferreira 2011] M. Bellassoued and D. Dos Santos Ferreira, "Stability estimates for the anisotropic wave equation from the Dirichlet-to-Neumann map", *Inverse Probl. Imaging* **5**:4 (2011), 745–773. MR Zbl

[Ben Aïcha 2015] I. Ben Aïcha, "Stability estimate for a hyperbolic inverse problem with time-dependent coefficient", *Inverse Problems* **31**:12 (2015), art. id. 125010. MR Zbl

[Boman and Quinto 1993] J. Boman and E. T. Quinto, "Support theorems for Radon transforms on real analytic line complexes in three-space", *Trans. Amer. Math. Soc.* **335**:2 (1993), 877–890. MR Zbl

[Cooper and Strauss 1984] J. Cooper and W. Strauss, "The leading singularity of a wave reflected by a moving boundary", *J. Differential Equations* **52**:2 (1984), 175–203. MR Zbl

[Dairbekov et al. 2007] N. S. Dairbekov, G. P. Paternain, P. Stefanov, and G. Uhlmann, "The boundary rigidity problem in the presence of a magnetic field", *Adv. Math.* **216**:2 (2007), 535–609. MR Zbl

[Dos Santos Ferreira et al. 2009] D. Dos Santos Ferreira, C. E. Kenig, M. Salo, and G. Uhlmann, "Limiting Carleman weights and anisotropic inverse problems", *Invent. Math.* **178**:1 (2009), 119–171. MR Zbl

[Duistermaat 1996] J. J. Duistermaat, *Fourier integral operators*, Progress in Mathematics **130**, Birkhäuser, Boston, 1996. MR Zbl

[Eskin 2017] G. Eskin, "Inverse problems for general second order hyperbolic equations with time-dependent coefficients", *Bull. Math. Sci.* **7**:2 (2017), 247–307. MR Zbl

[Eskin and Ralston 2010] G. Eskin and J. Ralston, "The determination of moving boundaries for hyperbolic equations", *Inverse Problems* **26**:1 (2010), art. id. 015001. MR Zbl

[Greenleaf and Uhlmann 1989] A. Greenleaf and G. Uhlmann, "Nonlocal inversion formulas for the X-ray transform", *Duke Math. J.* **58**:1 (1989), 205–240. MR Zbl

[Greenleaf and Uhlmann 1990a] A. Greenleaf and G. Uhlmann, "Composition of some singular Fourier integral operators and estimates for restricted X-ray transforms", *Ann. Inst. Fourier* (*Grenoble*) **40**:2 (1990), 443–466. MR Zbl

[Greenleaf and Uhlmann 1990b] A. Greenleaf and G. Uhlmann, "Microlocal techniques in integral geometry", pp. 121–135 in *Integral geometry and tomography* (Arcata, CA, 1989), edited by E. Grinberg and E. T. Quinto, Contemp. Math. **113**, Amer. Math. Soc., Providence, RI, 1990. MR Zbl

[Hörmander 1985a] L. Hörmander, *The analysis of linear partial differential operators, III: Pseudodifferential operators*, Grundlehren der Mathematischen Wissenschaften **274**, Springer, 1985. MR Zbl

[Hörmander 1985b] L. Hörmander, *The analysis of linear partial differential operators, IV: Fourier integral operators*, Grundlehren der Mathematischen Wissenschaften **275**, Springer, 1985. MR Zbl

[Isakov and Sun 1992] V. Isakov and Z. Q. Sun, "Stability estimates for hyperbolic inverse problems with local boundary data", *Inverse Problems* **8**:2 (1992), 193–206. MR Zbl

[Kian 2016] Y. Kian, "Recovery of time-dependent damping coefficients and potentials appearing in wave equations from partial data", *SIAM J. Math. Anal.* **48**:6 (2016), 4021–4046. MR Zbl

[Kian et al. 2018] Y. Kian, L. Oksanen, E. Soccorsi, and M. Yamamoto, "Global uniqueness in an inverse problem for time fractional diffusion equations", *J. Differential Equations* **264**:2 (2018), 1146–1170. MR Zbl

[Kurylev et al. 2014a] Y. Kurylev, M. Lassas, and G. Uhlmann, "Inverse problems for Lorentzian manifolds and non-linear hyperbolic equations", preprint, 2014. arXiv

[Kurylev et al. 2014b] Y. Kurylev, M. Lassas, and G. Uhlmann, "Inverse problems in spacetime, I: Inverse problems for Einstein equations (extended preprint version)", preprint, 2014. arXiv

[Lassas et al. 2016] M. Lassas, G. Uhlmann, and Y. Wang, "Inverse problems for semilinear wave equations on Lorentzian manifolds", preprint, 2016. arXiv

[Lassas et al. 2017] M. Lassas, L. Oksanen, P. Stefanov, and G. Uhlmann, "On the inverse problem of finding cosmic strings and other topological defects", *Comm. Math. Phys.* (online publication November 2017).

[Montalto 2014] C. Montalto, "Stable determination of a simple metric, a covector field and a potential from the hyperbolic Dirichlet-to-Neumann map", *Comm. Partial Differential Equations* **39**:1 (2014), 120–145. MR Zbl

[Palais 1960] R. S. Palais, "Extending diffeomorphisms", *Proc. Amer. Math. Soc.* **11** (1960), 274–277. MR Zbl

[Petrov 1969] A. Z. Petrov, *Einstein spaces*, Pergamon Press, Oxford, 1969. MR Zbl

[RabieniaHaratbar 2017] S. RabieniaHaratbar, "Support theorem for the light ray transform on Minkoswki spaces", preprint, 2017. To appear in *Inverse Probl. Imaging*. arXiv

[Ramm and Rakesh 1991] A. G. Ramm and Rakesh, "Property C and an inverse problem for a hyperbolic equation", *J. Math. Anal. Appl.* **156**:1 (1991), 209–219. MR Zbl

[Ramm and Sjöstrand 1991] A. G. Ramm and J. Sjöstrand, "An inverse problem of the wave equation", *Math. Z.* **206**:1 (1991), 119–130. MR Zbl

[Salazar 2013] R. Salazar, "Determination of time-dependent coefficients for a hyperbolic inverse problem", *Inverse Problems* **29**:9 (2013), art. id. 095015. MR Zbl

[Salazar 2014] R. Salazar, "Stability estimate for the relativistic Schrödinger equation with time-dependent vector potentials", *Inverse Problems* **30**:10 (2014), art. id. 105005. MR Zbl

[Stefanov 1989] P. D. Stefanov, "Uniqueness of the multi-dimensional inverse scattering problem for time dependent potentials", *Math. Z.* **201**:4 (1989), 541–559. MR Zbl

[Stefanov 1991] P. D. Stefanov, "Inverse scattering problem for moving obstacles", *Math. Z.* **207**:3 (1991), 461–480. MR Zbl

[Stefanov 2017] P. Stefanov, "Support theorems for the light ray transform on analytic Lorentzian manifolds", *Proc. Amer. Math. Soc.* **145**:3 (2017), 1259–1274. MR Zbl

[Stefanov and Uhlmann 1998] P. Stefanov and G. Uhlmann, "Stability estimates for the hyperbolic Dirichlet to Neumann map in anisotropic media", *J. Funct. Anal.* **154**:2 (1998), 330–358. MR Zbl

[Stefanov and Uhlmann 2005a] P. Stefanov and G. Uhlmann, "Boundary rigidity and stability for generic simple metrics", *J. Amer. Math. Soc.* **18**:4 (2005), 975–1003. MR Zbl

[Stefanov and Uhlmann 2005b] P. Stefanov and G. Uhlmann, "Stable determination of generic simple metrics from the hyperbolic Dirichlet-to-Neumann map", *Int. Math. Res. Not.* **2005**:17 (2005), 1047–1061. MR Zbl

[Stefanov et al. 2016] P. Stefanov, G. Uhlmann, and A. Vasy, "On the stable recovery of a metric from the hyperbolic DN map with incomplete data", *Inverse Probl. Imaging* **10**:4 (2016), 1141–1147. MR Zbl

[Sun 1990] Z. Q. Sun, "On continuous dependence for an inverse initial-boundary value problem for the wave equation", *J. Math. Anal. Appl.* **150**:1 (1990), 188–204. MR Zbl

[Tataru 1995] D. Tataru, "Unique continuation for solutions to PDE's: between Hörmander's theorem and Holmgren's theorem", *Comm. Partial Differential Equations* **20**:5-6 (1995), 855–884. MR Zbl

[Tataru 1999] D. Tataru, "Unique continuation for operators with partially analytic coefficients", *J. Math. Pures Appl.* (9) **78**:5 (1999), 505–521. MR Zbl

[Taylor 1981] M. E. Taylor, *Pseudodifferential operators*, Princeton Mathematical Series **34**, Princeton University Press, 1981. MR Zbl

[Waters 2014] A. Waters, "Stable determination of X-ray transforms of time dependent potentials from partial boundary data", *Comm. Partial Differential Equations* **39**:12 (2014), 2169–2197. MR Zbl

Received 26 Sep 2016. Revised 6 Jan 2018. Accepted 14 Feb 2018.

PLAMEN STEFANOV: stefanop@purdue.edu
Department of Mathematics, Purdue University, West Lafayette, IN, United States

YANG YANG: yangy5@msu.edu
Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI, United States

ANALYSIS AND PDE
Vol. 11, No. 6, 2018
dx.doi.org/10.2140/apde.2018.11.1415

msp

BLOW-UP CRITERIA FOR THE NAVIER–STOKES EQUATIONS IN NON-ENDPOINT CRITICAL BESOV SPACES

DALLAS ALBRITTON

We obtain an improved blow-up criterion for solutions of the Navier–Stokes equations in critical Besov spaces. If a mild solution u has maximal existence time $T^* < \infty$, then the non-endpoint critical Besov norms must become infinite at the blow-up time:

$$\lim_{t \uparrow T^*} \|u(\cdot, t)\|_{\dot{B}^{-1+3/p}_{p,q}(\mathbb{R}^3)} = \infty, \quad 3 < p, q < \infty.$$

In particular, we introduce a priori estimates for the solution based on elementary splittings of initial data in critical Besov spaces and energy methods. These estimates allow us to rescale around a potential singularity and apply backward uniqueness arguments. The proof does not use profile decomposition.

1. Introduction

We are interested in blow-up criteria for solutions of the incompressible Navier–Stokes equations

$$\begin{aligned} \partial_t u - \Delta u + u \cdot \nabla u + \nabla p &= 0, \\ \operatorname{div} u &= 0, \\ u(\cdot, 0) &= u_0 \end{aligned} \tag{NSE}$$

in $Q_T := \mathbb{R}^3 \times (0, T)$ with divergence-free initial data $u_0 \in C_0^\infty(\mathbb{R}^3)$. It has been known since [Leray 1934] that a unique smooth solution with sufficient decay at infinity exists locally in time. Furthermore, Leray proved that there exists a constant $c_p > 0$ with the property that if $T^* < \infty$ is the maximal time of existence of a smooth solution, then

$$\|u(\cdot, t)\|_{L^p(\mathbb{R}^3)} \geq c_p \left(\frac{1}{\sqrt{T^* - t}} \right)^{1 - \frac{3}{p}} \tag{1-1}$$

for all $3 < p \leq \infty$. Such a characterization exists because the Lebesgue norms in this range are subcritical with respect to the natural scaling symmetry of the Navier–Stokes equations,

$$u(x, t) \to \lambda u(\lambda x, \lambda^2 t), \quad p(x, t) \to \lambda^2 p(\lambda x, \lambda^2 t). \tag{1-2}$$

The behavior of the critical L^3 norm near a potential blow-up was unknown until the work of Escauriaza, Seregin, and Šverák [Escauriaza et al. 2003], who discovered an endpoint local regularity criterion in the spirit of the classical work [Serrin 1962]. In particular, they demonstrated that if u is a Leray–Hopf

MSC2010: 35Q30.
Keywords: Navier–Stokes equations, Besov spaces, blow-up criteria.

solution of the Navier–Stokes equations with maximal existence time $T^* < \infty$, then

$$\limsup_{t \uparrow T^*} \|u(\cdot, t)\|_{L^3(\mathbb{R}^3)} = \infty. \tag{1-3}$$

Their proof uses the ε-regularity criterion of Caffarelli, Kohn, and Nirenberg [Caffarelli et al. 1982] in an essential way, and moreover it introduced powerful backward uniqueness arguments for studying potential singularities of solutions to the Navier–Stokes equations. The proof is by contradiction: If a solution forms a singularity but remains in the critical space $L^\infty_t L^3_x(Q_{T^*})$, then one may zoom in on the singularity using the scaling symmetry and obtain a weak limit. The limit solution will form a singularity but also vanish identically at the blow-up time. By backwards uniqueness, the limit solution u must be identically zero in spacetime, which contradicts that it forms a singularity. This method was adapted by Phuc [2015] to cover blow-up criteria in Lorentz spaces. Interestingly, backwards uniqueness techniques have also been employed in the context of harmonic map heat flow by Wang [2008]. For a different proof of the criterion, see [Dong and Du 2009].

A few years ago, Seregin [2012a] improved the blow-up criterion of Escauriaza, Seregin and Šverák by demonstrating that the L^3 norm must become infinite at a potential blow-up:

$$\lim_{t \uparrow T^*} \|u(\cdot, t)\|_{L^3(\mathbb{R}^3)} = \infty. \tag{1-4}$$

The main new difficulty in the proof is that one no longer controls the $L^\infty_t L^3_x$ norm when zooming in on a potential singularity. Seregin addressed this difficulty by relying on certain properties of the local energy solutions introduced by Lemarié-Rieusset [2002]; see also [Kikuchi and Seregin 2007]. However, an analogous theory of local energy solutions was not known in the half-space $\mathbb{R}^3_+ := \{x \in \mathbb{R}^3 : x_3 > 0\}$.[1] In order to overcome this obstacle, Barker and Seregin [2017] introduced new a priori estimates which depend only on the norm of the initial data in the Lorentz spaces $L^{3,q}$, $3 < q < \infty$. This is accomplished by splitting the solution as

$$u = e^{t\Delta} u_0 + w, \tag{1-5}$$

where w is a correction in the energy space. The new estimates allowed Barker and Seregin to obtain an analogous blow-up criterion for Lorentz norms in the half-space. Later, Seregin and Šverák [2017] abstracted this splitting argument into the notion of a global weak L^3 solution. We direct the reader to [Barker et al. 2013] for global weak solutions with initial data in $L^{3,\infty}$.

Recently, there has been interest in adapting the "concentration compactness + rigidity" road map of [Kenig and Merle 2006] to blow-up criteria for the Navier–Stokes equations. This line of thought was initiated by Kenig and G. Koch [2011] and advanced to its current state by Gallagher, Koch, and Planchon in [Gallagher et al. 2013; 2016]. Gallagher et al. succeeded in extending a version of the blow-up criterion to the negative regularity critical Besov spaces $\dot{B}^{s_p}_{p,q}(\mathbb{R}^3)$, $3 < p, q < \infty$. Here, $s_p := -1 + \frac{3}{p}$ is the critical exponent. Specifically, it is proved in [Gallagher et al. 2016] that if $T^* < \infty$, then

$$\limsup_{t \uparrow T^*} \|u(\cdot, t)\|_{\dot{B}^{s_p}_{p,q}(\mathbb{R}^3)} = \infty. \tag{1-6}$$

[1] It appears that this theory has recently been developed in [Maekawa et al. 2017].

This proof is also by contradiction. If there is a blow-up solution in the space $L^\infty_t(\dot{B}^{s_p}_{p,q})_x$, then one may prove via profile decomposition that there is a blow-up solution in the same space and with minimal norm, made possible by small-data-global-existence results [Kato 1984]. This solution is known as a critical element. By essence of its minimality, the critical element vanishes identically at the blow-up time, so one may apply the backward uniqueness arguments of Escauriaza, Seregin, and Šverák to obtain a contradiction. The main difficulty lies in proving the existence of a profile decomposition in Besov spaces, which requires some techniques from the theory of wavelets [Koch 2010; Bahouri et al. 2011b]. A secondary difficulty is obtaining the necessary estimates near the blow-up time in order to apply the ε-regularity criterion. We note that [Kenig and Koch 2011] appears to be the first application of Kenig and Merle's road map to a parabolic equation. The nonlinear profile decomposition for the Navier–Stokes equations was first proved in [Gallagher 2001]. The paper [Bahouri et al. 2014] contains further interesting applications of profile decomposition techniques to the Navier–Stokes equations.

In this paper, we obtain the following improved blow-up criterion for the Navier–Stokes equations in critical spaces:

Theorem 1.1 (blow-up criterion). *Let $3 < p, q < \infty$ and $u_0 \in \dot{B}^{s_p}_{p,q}(\mathbb{R}^3)$ be a divergence-free vector field. Suppose u is the mild solution of the Navier–Stokes equations on $\mathbb{R}^3 \times [0, T^*)$ with initial data u_0 and maximal time of existence $T^*(u_0)$. If $T^* < \infty$, then*

$$\lim_{t \uparrow T^*} \|u(\cdot, t)\|_{\dot{B}^{s_p}_{p,q}(\mathbb{R}^3)} = \infty. \tag{1-7}$$

The local well-posedness of mild solutions, including characterizations of the maximal time of existence, are reviewed in Theorem A.2.

Let us discuss the novelty of Theorem 1.1. First, this theorem extends Seregin's L^3 criterion (1-4) to the scale of Besov spaces and replaces the lim sup condition in Gallagher, Koch and Planchon's criterion (1-6). Our proof does not rely on the profile decomposition techniques in [Gallagher et al. 2016] and may be considered to be more elementary. Rather, our methods are based on the rescaling procedure in [Seregin 2012a]. Regarding optimality, it is not clear whether Theorem 1.1 is valid for the endpoint spaces $\dot{B}^{s_p}_{p,\infty}$ and BMO^{-1}, which contain nontrivial -1-homogeneous functions, e.g., $|x|^{-1}$. Indeed, if the blow-up profile $u(\cdot, T^*)$ is locally a scale-invariant function, then rescaling around the singularity no longer provides useful information.[2] It is likely that this is an essential issue and not merely an artifact of the techniques used here. For instance, one may speculate that if type-I blow-up occurs (in the sense that the solution blows up in L^∞ at the self-similar rate), then the VMO^{-1} norm does not blow-up at the first singular time.

As in previous works on blow-up criteria for the Navier–Stokes equations, the main difficulty we encounter is in obtaining a priori estimates for solutions up to the potential blow-up time. We also require that the estimates depend only on the norm of the initial data in $\dot{B}^{s_p}_{p,q}$. The low regularity of this space creates a new difficulty because the splitting (1-5) does not appear to work in the space $\dot{B}^{s_p}_{p,q}$

[2] Since the submission of this paper, T. Barker has proven the blow-up criterion $\lim_{t \uparrow T^*} \|u(\cdot, t)\|_{\dot{B}^{s_p}_{p,\infty}} = \infty$, using Calderón-type solutions, under the extra assumption that $u(\cdot, T^*)$ vanishes in the rescaling limit [Barker 2017]. See also the forthcoming work [Albritton and Barker 2018] of T. Barker and the author.

when $\frac{2}{q}+\frac{3}{p}<1$. One problem is that when obtaining energy estimates for the correction w in (1-5), the operator

$$(U,u_0)\mapsto \int_0^T\int_{\mathbb{R}^3} e^{t\Delta}u_0\cdot\nabla U\cdot U\,dx\,dt \tag{1-8}$$

is not known to be bounded for $U\in L^\infty_t L^2_x\cap L^2_t\dot{H}^1_x$ and $u_0\in\dot{B}^{s_p}_{p,q}$. This is because $e^{t\Delta}u_0$ "just misses" the critical Lebesgue space $L^r_tL^p_x$ with $\frac{2}{r}+\frac{3}{p}=1$. Therefore, to obtain the necessary a priori estimates, we rely on a method essentially established by C. P. Calderón [1990]. The idea is as follows. We split the critical initial data $u_0\in\dot{B}^{s_p}_{p,p}$ into supercritical and subcritical parts:

$$u_0=U_0+V_0\in L^2+\dot{B}^{s_q+\varepsilon}_{q,q}. \tag{1-9}$$

When small, the data V_0 in a subcritical Besov space gives rise to a mild solution V on a prescribed time interval (not necessarily a global mild solution). The supercritical data $U_0\in L^2$ serves as initial data for a correction U in the energy space. We will refer to solutions which split in this way as Calderón solutions, see Definition 2.1, and we construct them in the sequel. Note that the unboundedness of (1-8) is similarly problematic when proving weak-strong uniqueness in Besov spaces. In recent work on weak-strong uniqueness, Barker [2018] has also dealt with this issue via the splitting (1-9). We remark that Calderón's original idea was to construct global weak solutions by splitting L^p initial data for $2<p<3$ into small data in L^3 and a correction in L^2. This idea has also been used to prove the stability of global mild solutions [Gallagher et al. 2002; Auscher et al. 2004].

Let us briefly contrast the solutions we construct via (1-9) to the global weak L^3 solutions introduced in [Seregin and Šverák 2017], which are constructed via the splitting (1-5). The correction term w in (1-5) has zero initial data, which allows one to prove that an appropriate limit of solutions also satisfies the energy inequality up to the initial time. For this reason, the global weak L^3 solutions are continuous with respect to weak convergence of initial data in L^3. Since the splitting (1-9) requires the correction to have nonzero initial condition U_0, an analogous continuity result is not as obvious for Calderón solutions. We do not seek to prove such a result here as to avoid burdening the paper technically, but we expect that it may be shown by adapting various ideas in [Seregin 2012b; Barker 2018]. Using similar ideas, we expect that one could prove that all Calderón solutions agree with the mild solution on a short time interval.

Here is the layout of the paper:

- In Section 2, we prove the existence of Calderón solutions that agree with the mild solution until the blow-up time. This is the content of Theorems 2.5 and 2.6. We also describe the properties of weak limits of Calderón solutions in Theorem 2.7. The splitting arguments for initial data in Besov spaces are contained in Lemma 2.2.

- In Section 3, we prove Theorem 1.1 using the results of Section 2.

- We conclude with an extensive Appendix that summarizes the local well-posedness theory of mild solutions in homogeneous Besov spaces and collects well-known theorems about ε-regularity and backward uniqueness. We include it for the reader's convenience and to make the paper self-contained.

Notation is reviewed in the Appendix. One important point is that we do not distinguish the notation of scalar-valued and vector-valued functions.

After completion of the present work, we learned that T. Barker and G. Koch (personal communication, 2017) obtained a different proof of the blow-up criterion (1-7). Their proof treats mild solutions by exploiting certain properties of the local energy solutions of Lemarié-Rieusset.

2. Calderón's method

In this section, we present properties of the following notion of solution:

Definition 2.1 (Calderón solution). Let $3 < p < \infty$ and $u_0 \in B^{s_p}_{p,p}(\mathbb{R}^3)$ be a divergence-free vector field. Suppose $T > 0$ is finite. We say that a distribution u on Q_T is a *Calderón solution* on Q_T with initial data u_0 if the following requirements are met:

$$u_0 = U_0 + V_0, \quad u = U + V, \tag{2-1}$$

where

$$U_0 \in L^2(\mathbb{R}^3), \quad V_0 \in \dot{B}^{s_q+\varepsilon}_{q,q}(\mathbb{R}^3), \tag{2-2}$$

$$U \in L^\infty_t L^2_x \cap L^2_t \dot{H}^1_x(Q_T), \quad V \in \mathcal{K}^{s_q+\varepsilon}_q(Q_T), \tag{2-3}$$

$$q > p, \quad 0 < \varepsilon < -s_q, \tag{2-4}$$

and V is the mild solution of the Navier–Stokes equations on Q_T with initial data V_0; see Theorem A.7. In addition, U satisfies the perturbed Navier–Stokes system

$$\partial_t U - \Delta U + \operatorname{div} U \otimes U + \operatorname{div} U \otimes V + \operatorname{div} V \otimes U = -\nabla P, \quad \operatorname{div} U = 0 \tag{2-5}$$

in the sense of distributions on Q_T, where

$$P \in L^2_t L^{\frac{3}{2}}_x(Q_T) + L^2(Q_T). \tag{2-6}$$

We require that $U(\cdot\,, t)$ is weakly continuous as an $L^2(\mathbb{R}^3)$-valued function on $[0, T]$ and that U attains its initial condition strongly in $L^2(\mathbb{R}^3)$:

$$\lim_{t\downarrow 0} \|U(\cdot\,, t) - U_0\|_{L^2(\mathbb{R}^3)} = 0. \tag{2-7}$$

Define

$$Q := (-\Delta)^{-1} \operatorname{div}\operatorname{div} V \otimes V, \quad p := P + Q. \tag{2-8}$$

We require that (u, p) is suitable for the Navier–Stokes equations,

$$\partial_t |u|^2 + 2|\nabla u|^2 \leq \Delta |u|^2 - \operatorname{div}((|u|^2 + 2p)u), \tag{2-9}$$

and that (U, P) is suitable for (2-5),

$$\partial_t |U|^2 + 2|\nabla U|^2 \leq \Delta |U|^2 - \operatorname{div}((|U|^2 + 2P)U) - \operatorname{div}(|U|^2 V) - 2U \operatorname{div}(V \otimes U). \tag{2-10}$$

The inequalities (2-9) and (2-10) are interpreted in the sense of distributions evaluated on nonnegative test functions $0 \le \varphi \in C_0^\infty(Q_T)$. Lastly, we require that U satisfies the global energy inequality

$$\int_{\mathbb{R}^3} |U(x,t_1)|^2\,dx + 2\int_{t_0}^{t_1}\int_{\mathbb{R}^3} |\nabla U(x,t)|^2\,dx\,dt \le \int_{\mathbb{R}^3} |U(x,t_0)|^2\,dx + 2\int_{t_0}^{t_1}\int_{\mathbb{R}^3} V\otimes U : \nabla U\,dx\,dt, \tag{2-11}$$

for almost every $0 \le t_0 < T$, including $t_0 = 0$, and for all $t_1 \in (t_0, T]$.

Splitting arguments. The next lemma allows us to represent critical initial data as the sum of subcritical and supercritical initial data while preserving the divergence free condition. See Proposition 2.8 in [Barker 2018] for a detailed proof.

Lemma 2.2 (splitting of critical data). *Let $3 < p < q \le \infty$ and $\theta \in (0,1)$ satisfy*

$$\frac{1}{p} = \frac{\theta}{2} + \frac{1-\theta}{q}. \tag{2-12}$$

Define $s := s_p/(1-\theta)$. For all $\lambda > 0$ and divergence-free vector fields $u \in \dot{B}^{s_p}_{p,p}(\mathbb{R}^3)$, there exist divergence-free vector fields U, V such that $u = U + V$,

$$\|U\|_{L^2(\mathbb{R}^3)} \le c\|u\|^{\frac{p}{2}}_{\dot{B}^{s_p}_{p,p}(\mathbb{R}^3)}\lambda^{1-\frac{p}{2}}, \tag{2-13}$$

$$\|V\|_{\dot{B}^{s}_{q,q}(\mathbb{R}^3)} \le c\|u\|^{\frac{p}{q}}_{\dot{B}^{s_p}_{p,p}(\mathbb{R}^3)}\lambda^{1-\frac{p}{q}}, \tag{2-14}$$

where $c > 0$ is an absolute constant.

The proof is by decomposing the Littlewood–Paley projections as

$$\dot{\Delta}_j u = (\dot{\Delta}_j u)\mathbf{1}_{\{|\dot{\Delta}_j u| > \lambda_j\}} + (\dot{\Delta}_j u)\mathbf{1}_{\{|\dot{\Delta}_j u| \le \lambda_j\}}, \quad j \in \mathbb{Z}, \tag{2-15}$$

with an appropriate choice of $\lambda_j > 0$, $j \in \mathbb{Z}$. The divergence-free condition is kept by applying the Leray projector to the resulting vector fields. Recall that the Leray projector is a Fourier multiplier with matrix-valued symbol homogeneous of degree zero and smooth away from the origin:

$$(\mathbb{P})_{ij} := \delta_{ij} + R_i R_j, \quad R_i := \frac{\partial_i}{|\nabla|}, \quad 1 \le i, j \le 3. \tag{2-16}$$

Note that $\dot{B}^{s}_{q,q}(\mathbb{R}^3)$ is indeed a subcritical space of initial data, since

$$s - \frac{3}{q} = -1 + \frac{\theta}{2(1-\theta)} > -1. \tag{2-17}$$

We will often define $\varepsilon := s - s_q > 0$.

Existence of energy solutions to NSE with lower-order terms. In this section, we will prove the existence of weak solutions to the Navier–Stokes equations with coefficients in critical Lebesgue spaces. The method of proof is well known and goes back to [Leray 1934].

Proposition 2.3 (existence of energy solutions). *Let $U_0 \in L^2(\mathbb{R}^3)$ be a divergence-free vector field, and let*

$$a, b \in L_t^l L_x^r(Q_T), \quad \frac{2}{l} + \frac{3}{r} = 1, \quad 3 < r \le \infty, \tag{2-18}$$

be vector fields for a given $T > 0$. Further assume that $\operatorname{div} b = 0$. *Then there exist a vector field U and pressure P,*

$$U \in L_t^\infty L_x^2 \cap L_t^2 \dot{H}_x^1(Q_T), \quad P \in L_t^2 L_x^{\frac{3}{2}}(Q_T) + L^2(Q_T), \tag{2-19}$$

such that the perturbed Navier–Stokes system

$$\partial_t U - \Delta U + \operatorname{div} U \otimes U + \operatorname{div} U \otimes b + \operatorname{div} a \otimes U = -\nabla P, \quad \operatorname{div} U = 0 \tag{2-20}$$

is satisfied on Q_T in the sense of distributions. In addition, $U(\cdot, t)$ is weakly continuous as an $L^2(\mathbb{R}^3)$-valued function on $[0, T]$, *and*

$$\lim_{t \downarrow 0} \|U(\cdot, t) - U_0\|_{L^2(\mathbb{R}^3)} = 0. \tag{2-21}$$

Finally, (U, P) *is suitable for* (2-20),

$$\partial_t |U|^2 + 2|\nabla U|^2 \le \Delta |U|^2 - \operatorname{div}((|U|^2 + 2P)U) - \operatorname{div}(|U|^2 b) - 2U \operatorname{div}(a \otimes U) \tag{2-22}$$

as distributions evaluated on nonnegative test functions $0 \le \varphi \in C_0^\infty(Q_T)$, and U satisfies the global energy inequality

$$\int_{\mathbb{R}^3} |U(x, t_2)|^2 \, dx \, dt + 2\int_{t_1}^{t_2} \int_{\mathbb{R}^3} |\nabla U(x, t)|^2 \, dx \, dt \le \int_{\mathbb{R}^3} |U(x, t_1)|^2 \, dx + 2\int_{t_1}^{t_2} \int_{\mathbb{R}^3} a \otimes U : \nabla U \, dx \, dt \tag{2-23}$$

for almost every $t_1 \in [0, T)$, including $t_1 = 0$, and for all $t_2 \in (t_1, T]$.

In the statement above, $U \operatorname{div}(a \otimes U)$ is the distribution

$$\langle U \operatorname{div}(a \otimes U), \varphi \rangle := -\int_0^T \int_{\mathbb{R}^3} a \otimes U : (U \otimes \nabla \varphi + \varphi \nabla U) \, dx \, dt \tag{2-24}$$

for all $\varphi \in C_0^\infty(Q_T)$.

Let us introduce some notation and basic estimates surrounding the energy space. For $0 \le t_0 < T \le \infty$, we define $Q_{t_0,T} := \mathbb{R}^3 \times (t_0, T)$. If $U \in L_t^\infty L_x^2 \cap L_t^2 \dot{H}_x^1(Q_{t_0,T})$, we define the *energy norm*

$$|U|_{2,Q_{t_0,T}}^2 := \sup_{t_0 < t < T} \int_{\mathbb{R}^3} |U(x, t)|^2 \, dx + 2\int_{t_0}^T \int_{\mathbb{R}^3} |\nabla U(x, t')|^2 \, dx \, dt'. \tag{2-25}$$

For simplicity, take $t_0 = 0$. By interpolation between $L_t^\infty L_x^2(Q_T)$ and $L_t^2 L_x^6(Q_T)$, one obtains

$$\|U\|_{L_t^m L_x^n(Q_T)} \le c|U|_{2,Q_T}, \tag{2-26}$$

$$\frac{2}{m} + \frac{3}{n} = \frac{3}{2}, \quad 2 \le m \le \infty, \quad 2 \le n \le 6. \tag{2-27}$$

For instance, it is common to take $m = n = \frac{10}{3}$. Hence, if $a \in L^l_t L^r_x(Q_T)$ is a vector field as in Proposition 2.3, we have

$$\int_0^T |U|^2 |a|^2 \, dx \, dt \le c\|a\|^2_{L^l_t L^r_x(Q_T)} |U|^2_{2,Q_T}, \tag{2-28}$$

$$\int_0^T |U|^2 |a| \, dx \, dt \le cT^{\frac{1}{2}} \|a\|_{L^l_t L^r_x(Q_T)} |U|^2_{2,Q_T}. \tag{2-29}$$

Let us now recall the tools used to estimate the time derivative $\partial_t U$ of a solution U that belongs to the energy space. To start, we will need to estimate the pointwise product of $u \in L^2(\mathbb{R}^3)$ and $v \in \dot{H}^1(\mathbb{R}^3)$:

$$\|uv\|_{\dot{H}^{-1/2}(\mathbb{R}^3)} \le c\|u\|_{L^2(\mathbb{R}^3)} \|v\|_{\dot{H}^1(\mathbb{R}^3)}; \tag{2-30}$$

see Corollary 2.55 in [Bahouri et al. 2011a]. For instance, it follows that

$$\|\operatorname{div} U \otimes U\|_{L^2_t \dot{H}^{-3/2}_x(Q_T)} \le c|U|^2_{2,Q_T}. \tag{2-31}$$

The time derivative $\partial_t U$ is typically only in $L^2_t \dot{H}_x^{-\frac{3}{2}}(Q_T)$ unless the nonlinear term is mollified. Notice also that $\dot{H}^{-\frac{3}{2}}(\mathbb{R}^3) \hookrightarrow H^{-\frac{3}{2}}(\mathbb{R}^3)$. Suppose $(U^{(n)})_{n\in\mathbb{N}}$ is a sequence of vector fields on Q_T such that

$$\sup_{n\in\mathbb{N}} |U^{(n)}|_{2,Q_T} + \|\partial_t U\|_{L^2_t H^{-3/2}_x(Q_T)} < \infty. \tag{2-32}$$

According to the Aubin–Lions lemma, see Chapter 5, Proposition 1.1 in [Seregin 2015], there exists a subsequence, still denoted by $U^{(n)}$, such that

$$U^{(n)} \to U \quad \text{in } L^2(B(R) \times (0,T)), \quad R > 0. \tag{2-33}$$

Since $\sup_{n\in\mathbb{N}} \|U^{(n)}\|_{L^{10/3}(Q_T)} < \infty$, the subsequence actually converges strongly in $L^3(B(R) \times (0,T))$. We will use this fact frequently in the sequel.

For the remainder of the paper, let us fix a nonnegative radially symmetric test function $0 \le \theta \in C^\infty_0(\mathbb{R}^3)$ with $\int_{\mathbb{R}^3} \theta \, dx = 1$. For a locally integrable function $f \in L^1_{\mathrm{loc}}(\mathbb{R}^3)$ we make the following definition:

$$\theta_\rho := \frac{1}{\rho^3}\theta\Big(\frac{\cdot}{\rho}\Big), \quad (f)_\rho := f * \theta_\rho, \quad \rho > 0. \tag{2-34}$$

The proof of the next lemma is well known, and we include it for completeness.

Lemma 2.4 (solution to mollified Navier–Stokes equation with lower-order terms). *Let $\rho > 0$. Assume the hypotheses of Proposition 2.3. Then there exists a unique U in the class*

$$U \in C([0,T]; L^2(\mathbb{R}^3)) \cap L^2_t \dot{H}^1_x(Q_T), \quad \partial_t U \in L^2_t H^{-1}_x(Q_T), \tag{2-35}$$

satisfying the mollified perturbed Navier–Stokes system in $L^2_t H^{-1}_x(Q_T)$

$$\partial_t U - \Delta U + \mathbb{P}\operatorname{div}(U \otimes (U)_\rho) + \mathbb{P}\operatorname{div}(U \otimes b + a \otimes U) = 0, \quad \operatorname{div} U = 0, \tag{2-36}$$

and such that $U(\cdot, 0) = U_0$ in $L^2(\mathbb{R}^3)$.

Proof. Assume the hypotheses of the lemma and define $T_\sharp := T$ in the statement, in order to reuse the variable T. We will consider the bilinear operator

$$B_\rho(v,w)(\cdot,t) := \int_0^t e^{(t-s)\Delta}\mathbb{P}\,\mathrm{div}\, v\otimes (w)_\rho\, ds, \tag{2-37}$$

defined formally for all vector fields v, w on spacetime, as well as the linear operator

$$L(w)(\cdot,t) := \int_0^t e^{(t-s)\Delta}\mathbb{P}\,\mathrm{div}(a\otimes w + w\otimes b)\, ds. \tag{2-38}$$

For instance, by classical energy estimates for the Stokes equations, the operators are well defined whenever $v\otimes (w)_\rho$, $a\otimes w$, $w\otimes b$ are square integrable. Specifically, due to (2-28) and the energy estimates, we have that B and L are bounded operators on the energy space, and

$$|B(v,w)|^2_{2,Q_T} \le c\|v\otimes (w)_\rho\|^2_{L^2(Q_T)} \le c(\rho)T|v|^2_{2,Q_T}|w|^2_{2,Q_T}, \tag{2-39}$$

$$|L(w)|^2_{2,Q_T} \le c\|a\otimes w\|^2_{L^2(Q_T)} \le \|a\|^2_{L^l_t L^r_x(Q_T)}|w|^2_{2,Q_T} \tag{2-40}$$

for all $T>0$. Notice that b does not enter the estimate, since $\mathrm{div}\, b = 0$. In addition, the operators take values in $C([0,T];L^2(\mathbb{R}^3)$. Moreover, notice that $\|a\|_{L^l_t L^r_x(Q_T)} \ll 1$ when $0<T\ll 1$. Hence, one may apply Lemma A.3 to solve the integral equation

$$U(\cdot,t) = e^{t\Delta}U_0 - B_\rho(U,U)(\cdot,t) - L(U)(\cdot,t) \tag{2-41}$$

up to time $0<T\ll 1$. The integral equation (2-41) is equivalent to the differential equation (2-36). The solution may be continued in the energy space up to the time $T_\sharp$ by the same method as long as the energy norm remains bounded. This will be the case, since a solution U on Q_T obeys the energy equality

$$\begin{aligned}\int_{\mathbb{R}^3}|U(x,t_2)|^2\,dx + 2\int_{t_1}^{t_2}\int_{\mathbb{R}^3}|\nabla U|^2\,dx\,dt &= \int_{\mathbb{R}^3}|U(x,t_1)|^2\,dx + 2\int_{t_1}^{t_2}\int_{\mathbb{R}^3} a\otimes U : \nabla U\,dx\,dt \\ &\le \int_{\mathbb{R}^3}|U(x,t_1)|^2\,dx + A(t_1,t_2)|U|^2_{2,Q_{t_1,t_2}}\end{aligned} \tag{2-42}$$

for all $0\le t_1<t_2\le T$, where $A(t_1,t_2) := c\|a\|_{L^l_t L^r_x(Q_{t_1,t_2})}$. Then one simply takes t_2 close enough to t_1 such that $A(t_1,t_2)\le \frac{1}{2}$ to obtain a local-in-time a priori energy bound. By repeating the argument a finite number of times, one obtains

$$|U|_{2,Q_T} \le C(\|U_0\|_{L^2(\mathbb{R}^3)}, \|a\|_{L^l_t L^r_x(Q_{T_\sharp})}). \tag{2-43}$$

Hence, the solution may be continued to the time $T_\sharp$. Finally, uniqueness follows from the construction. □

Proof of Proposition 2.3. We follow the standard procedure initiated by Leray [1934] of solving the mollified problem and taking the limit as $\rho \downarrow 0$. The arguments we present here are essentially adapted from [Seregin and Šverák 2017; Seregin 2015, Chapter 5], so we will merely summarize them.

(1) *Limits*. For each $\rho > 0$, we denote by U^ρ the unique solution from Lemma 2.4. From the proof of Lemma 2.4, recall the energy equality

$$\int_{\mathbb{R}^3} |U^\rho(x,t_2)|^2\,dx + 2\int_{t_1}^{t_2}\int_{\mathbb{R}^3} |\nabla U^\rho|^2\,dx\,dt = \int_{\mathbb{R}^3} |U^\rho(x,t_1)|^2\,dx + 2\int_{t_1}^{t_2}\int_{\mathbb{R}^3} a\otimes U^\rho : \nabla U^\rho\,dx\,dt \tag{2-44}$$

for all $0 \le t_1 < t_2 \le T$, as well as the a priori energy bound

$$|U^\rho|_{2,Q_T} \le C(\|U_0\|_{L^2(\mathbb{R}^3)}, \|a\|_{L^l_t L^r_x(Q_T)}). \tag{2-45}$$

In order to estimate the time derivative $\partial_t U \in L^2_t H_x^{-\frac{3}{2}}(Q_T)$, we rewrite (2-36) as

$$\partial_t U = \Delta U - \mathbb{P}\,\mathrm{div}(U^\rho \otimes (U^\rho)_\rho) - \mathbb{P}\,\mathrm{div}(a\otimes U^\rho + U^\rho \otimes b). \tag{2-46}$$

Then, due to the estimate (2-28), one obtains

$$\|\Delta U - \mathbb{P}\,\mathrm{div}(a\otimes U^\rho + U^\rho\otimes b)\|_{L^2_t H^{-1}_x(Q_T)} \le C(\|U_0\|_{L^2(\mathbb{R}^3)}, \|a\|_{L^l_t L^r_x(Q_T)}, \|b\|_{L^l_t L^r_x(Q_T)}), \tag{2-47}$$

and in light of (2-30), we also have

$$\begin{aligned}\|\mathbb{P}\,\mathrm{div}(U^\rho \otimes (U^\rho)_\rho)\|_{L^2_t H^{-3/2}_x(Q_T)} &\le c\|(U^\rho \otimes (U^\rho)_\rho)\|_{L^2_t H^{-1/2}_x(Q_T)}\\ &\le C(\|U_0\|_{L^2(\mathbb{R}^3)}, \|a\|_{L^l_t L^r_x(Q_T)}).\end{aligned} \tag{2-48}$$

The resulting estimate on the time derivative is

$$\|\partial_t U^\rho\|_{L^2_t H^{-3/2}_x(Q_T)} \le C(\|U_0\|_{L^2(\mathbb{R}^3)}, \|a\|_{L^l_t L^r_x(Q_T)}, \|b\|_{L^l_t L^r_x(Q_T)}). \tag{2-49}$$

By the Banach–Alaoglu theorem and (2-45), there exist $U \in L^\infty_t L^2_x \cap L^2_t \dot H^1_x(Q_T)$ and a sequence $\rho_k \downarrow 0$ such that

$$U^{\rho_k} \overset{*}{\rightharpoonup} U \quad \text{in } L^\infty_t L^2_x(Q_T), \qquad \nabla U^{\rho_k} \rightharpoonup \nabla U \quad \text{in } L^2(Q_T). \tag{2-50}$$

In addition, by (2-49) and the discussion preceding Lemma 2.4, a subsequence may be chosen such that

$$U^{\rho_k} \to U \quad \text{in } L^3_t (L^3_{\mathrm{loc}})_x(Q_T). \tag{2-51}$$

Moreover, the estimate (2-49) allows us to prove that

$$\int_{\mathbb{R}^3} U^{\rho_k}(x,\cdot)\cdot\varphi(x)\,dx \to \int_{\mathbb{R}^3} U(x,\cdot)\cdot\varphi(x)\,dx \quad \text{in } C([0,T]) \tag{2-52}$$

for all vector fields $\varphi \in L^2(\mathbb{R}^3)$. In particular,

$$U^{\rho_k}(\cdot,t) \rightharpoonup U(\cdot,t) \quad \text{in } L^2(\mathbb{R}^3), \quad t\in[0,T], \tag{2-53}$$

and the limit $U(\cdot,t)$ is weakly continuous as an $L^2(\mathbb{R}^3)$-valued function on $[0,T]$. Here is our argument for (2-52), based on Chapter 5, p. 102 of [Seregin 2015]. For each vector field $\varphi \in C_0^\infty(\mathbb{R}^3)$, consider the family

$$\mathcal{F}_\varphi := \left\{ \int_{\mathbb{R}^3} U^\rho(x,\cdot)\cdot\varphi(x)\,dx : \rho > 0 \right\} \subset C([0,T]). \tag{2-54}$$

The family $\mathcal{F}_\varphi$ is uniformly bounded, since

$$\left|\int U^\rho(x,t)\cdot\varphi(x)\,dx\right| \le \|U^\rho\|_{L^\infty_t L^2_x(Q_T)}\|\varphi\|_{L^2(\mathbb{R}^3)} \le C(\|U_0\|_{L^2(\mathbb{R}^3)},\|a\|_{L^l_t L^r_x(Q_T)})\|\varphi\|_{L^2(\mathbb{R}^3)} \tag{2-55}$$

for all $0\le t\le T$. It is also equicontinuous:

$$\begin{aligned}\left|\int U^\rho(x,t_2)\cdot\varphi(x)-U^\rho(x,t_1)\cdot\varphi(x)\,dx\right| &\le |t_2-t_1|^{\frac12}\|\partial_t U^\rho\|_{L^2_t H^{-3/2}_x(Q_T)}\|\varphi\|_{H^{3/2}(\mathbb{R}^3)}\\ &\le |t_2-t_1|^{\frac12}C(\|U_0\|_{L^2(\mathbb{R}^3)},\|a\|_{L^l_t L^r_x(Q_T)},\|b\|_{L^l_t L^r_x(Q_T)})\|\varphi\|_{H^{3/2}(\mathbb{R}^3)}\end{aligned} \tag{2-56}$$

for all $0\le t_1,t_2\le T$. Hence, we may extract a further subsequence, still denoted by ρ_k, such that (2-52). By a diagonalization argument and the density of test functions in $L^2(\mathbb{R}^3)$, we can obtain (2-52) for all vector fields $\varphi\in L^2(\mathbb{R}^3)$. This completes the summary of the convergence properties of U^{ρ_k} as $\rho_k\downarrow 0$.

Let us now analyze the behavior of U near the initial time. In the limit $\rho_k\downarrow 0$, the energy equality (2-44) gives rise to an energy *inequality*:

$$\int_{\mathbb{R}^3}|U(x,t)|^2\,dx+2\int_0^t\int_{\mathbb{R}^3}|\nabla U|^2\,dx\,dt' \le \int_{\mathbb{R}^3}|U_0(x)|^2\,dx+C(\|U_0\|_{L^2(\mathbb{R}^3)},\|a\|_{L^l_t L^r_x(Q_T)})\|a\|_{L^l_t L^r_x(Q_t)} \tag{2-57}$$

for almost every $t\in(0,T)$. This is due to the lower semicontinuity of the energy norm with respect to weak-star convergence. Since $U(\cdot,t)$ is weakly continuous as an $L^2(\mathbb{R}^3)$-valued function, the energy inequality may be extended to all $t\in(0,T]$. Finally, since $U(\cdot,0)=U_0$ and $\limsup_{t\downarrow 0}\|U(\cdot,t)\|_{L^2(\mathbb{R}^3)}\le \|U_0\|_{L^2(\mathbb{R}^3)}$ (by taking $t\downarrow 0$ in (2-57)), we obtain

$$\lim_{t\downarrow 0}\|U(\cdot,t)-U_0\|_{L^2(\mathbb{R}^3)}=0. \tag{2-58}$$

We have proven the desired properties of U except for suitability and the global energy inequality.

Let us now take the limit of the pressures. The pressure P^ρ associated to U^ρ in (2-36) may be calculated as $P^\rho := P^\rho_1+P^\rho_2$,

$$P^\rho_1 := (-\Delta)^{-1}\operatorname{div}\operatorname{div}(U^\rho\otimes(U^\rho)_\rho), \tag{2-59}$$

$$P^\rho_2 := (-\Delta)^{-1}\operatorname{div}\operatorname{div}(U^\rho\otimes b+a\otimes U^\rho). \tag{2-60}$$

By the Calderón–Zygmund estimates, we have the following bounds independent of the parameter $\rho>0$:

$$\begin{aligned}\|P^\rho_1\|_{L^2_t L^{3/2}_x(Q_T)} &\le c\|U^\rho\otimes(U^\rho)_\rho\|_{L^2_t L^{3/2}_x(Q_T)}\\ &\le c\|U^\rho\|_{L^\infty_t L^2_x(Q_T)}\|U^\rho\|_{L^2_t L^6_x(Q_T)}\\ &\le C(\|U_0\|_{L^2(\mathbb{R}^3)},\|a\|_{L^l_t L^r_x(Q_T)}),\end{aligned} \tag{2-61}$$

and similarly,

$$\begin{aligned}\|P_2^\rho\|_{L^2(Q_T)} &\le c\|a\otimes U^\rho + U^\rho\otimes b\|_{L^2(Q_T)}\\ &\le c|U^\rho|_{2,Q_T}(\|a\|_{L^l_tL^r_x(Q_T)}+\|b\|_{L^l_tL^r_x(Q_T)})\\ &\le C(\|U_0\|_{L^2(\mathbb{R}^3)},\|a\|_{L^l_tL^r_x(Q_T)},\|b\|_{L^l_tL^r_x(Q_T)}).\end{aligned} \tag{2-62}$$

In particular, we may pass to a further subsequence, still denoted by ρ_k, such that

$$P_1^{\rho_k} \rightharpoonup P_1 \quad \text{in } L^2_tL^{\frac{3}{2}}_x(Q_T), \qquad P_2^{\rho_k} \rightharpoonup P_2 \quad \text{in } L^2(Q_T), \tag{2-63}$$

$$P^{\rho_k} \rightharpoonup P := P_1 + P_2 \quad \text{in } L^2_t(L^{\frac{3}{2}}_{\mathrm{loc}})_x(Q_T). \tag{2-64}$$

Recall that $U^{\rho_k} \to U$ in $L^3_t(L^3_{\mathrm{loc}})_x$. Utilizing this fact in (2-59) and (2-60), we observe that

$$-\Delta P_1 = \operatorname{div}\operatorname{div} U\otimes U, \quad -\Delta P_2 = \operatorname{div}\operatorname{div}(a\otimes U + U\otimes b) \tag{2-65}$$

in the sense of distributions on Q_T. Finally, the Liouville theorem for entire harmonic functions in Lebesgue spaces implies

$$P_1 = (-\Delta)^{-1}\operatorname{div}\operatorname{div} U\otimes U, \quad P_2 = (-\Delta)^{-1}\operatorname{div}\operatorname{div}(a\otimes U + U\otimes b). \tag{2-66}$$

Due to the limit behavior discussed above, it is clear that (U, P) solves the system (2-20) in the sense of distributions.

(2) *Suitability*. We will now prove that U is suitable for (2-20). Specifically, we will verify the local energy inequality (2-22) following arguments in [Lemarié-Rieusset 2002] (see p. 318). We start from the mollified local energy equality

$$\partial_t|U^\rho|^2 + 2|\nabla U^\rho|^2 = \Delta|U^\rho|^2 - \operatorname{div}(|U^\rho|^2(U^\rho)_\rho + 2P^\rho U^\rho) - \operatorname{div}(|U^\rho|^2 b) - 2U^\rho\operatorname{div}(a\otimes U^\rho), \tag{2-67}$$

which is satisfied in the sense of distributions by solutions of (2-36) on Q_T. Let us analyze the convergence of each term in (2-67) as $\rho_k \downarrow 0$. Recall that

$$P^{\rho_k} \rightharpoonup P \quad \text{in } L^{\frac{3}{2}}_{\mathrm{loc}}(Q_T), \qquad U^{\rho_k} \to U \quad \text{in } L^3_{\mathrm{loc}}(Q_T). \tag{2-68}$$

This readily implies

$$|U^{\rho_k}|^2(U^{\rho_k})_{\rho_k} \to |U|^2U \quad \text{in } L^1_{\mathrm{loc}}(Q_T), \qquad P^{\rho_k}U^{\rho_k} \to PU \quad \text{in } L^1_{\mathrm{loc}}(Q_T), \tag{2-69}$$

$$|U^{\rho_k}|^2 \to |U|^2 \quad \text{in } L^{\frac{3}{2}}_{\mathrm{loc}}(Q_T). \tag{2-70}$$

Moreover, according to the estimates (2-28) and (2-29),

$$|U^{\rho_k}|^2 b \to |U|^2 b \qquad \text{in } L^1(Q_T), \tag{2-71}$$

$$U^{\rho_k}\operatorname{div}(a\otimes U^{\rho_k}) \to U\operatorname{div}(a\otimes U) \quad \text{in } \mathcal{S}'(Q_T). \tag{2-72}$$

It remains to analyze the term $|\nabla U^{\rho_k}|^2$. Upon passing to a subsequence, $|\nabla U^{\rho_k}|^2$ converges weakly-star in $\mathcal{M}(Q_T)$, the space consisting of finite Radon measures on Q_T, but its limit may not be $|\nabla U|^2$. On the other hand, recall that $\nabla U^{\rho_k} \rightharpoonup \nabla U$ in $L^2(Q_T)$. Hence, by lower semicontinuity of the L^2 norm with respect to weak convergence,

$$\liminf_{\rho_k \downarrow 0} \int_E |\nabla U^{\rho_k}|^2 - |\nabla U|^2 \, dx \geq 0 \tag{2-73}$$

for all Borel measurable sets E contained in Q_T. Therefore, upon passing to a further subsequence,

$$\mu := \lim_{\rho_k \downarrow 0} |\nabla U^{\rho_k}|^2 - |\nabla U|^2 \tag{2-74}$$

is a nonnegative finite measure on Q_T, and (2-67) becomes

$$\partial_t |U|^2 + 2|\nabla U|^2 = \Delta |U|^2 - \operatorname{div}((|U|^2 + 2P)U) - \operatorname{div}(|U|^2 b) - 2U \operatorname{div}(a \otimes U) - 2\mu. \tag{2-75}$$

(3) *Global energy inequality*. Finally, let us pass from the local energy inequality (2-22) to the global energy inequality (2-23) in the following standard way; see [Lemarié-Rieusset 2002, p. 319].

Let $0 \leq \eta \in C_0^\infty(\mathbb{R})$ be an even function such that $\eta \equiv 1$ for $|t| \leq \frac{1}{4}$, $\eta \equiv 0$ for $|t| \geq \frac{1}{2}$, and $\int_{\mathbb{R}} \eta \, dt = 1$. Define $\eta_\varepsilon(t) := \varepsilon^{-1} \eta(\varepsilon^{-1} t)$. Given $0 < t_0 < t_1 < T$, consider

$$\psi_\varepsilon(t) := \int_{-\infty}^{t} \eta_\varepsilon(t' - t_0) - \eta_\varepsilon(t' - t_1) \, dt', \quad t \in \mathbb{R}. \tag{2-76}$$

The functions ψ_ε are smooth approximations of the characteristic function $\mathbf{1}_{(t_1,t_2)}$. Now let $0 \leq \varphi \in C_0^\infty(\mathbb{R}^3)$ such that $\varphi \equiv 1$ on $B(1)$ and $\varphi \equiv 0$ outside $B(2)$. Define

$$\Phi_{\varepsilon,R}(x,t) := \psi_\varepsilon(t) \varphi\left(\frac{x}{R}\right)^2, \quad (x,t) \in \mathbb{R}^{3+1}. \tag{2-77}$$

Using $\Phi_{\varepsilon,R}$ in the local energy inequality (2-22) with $0 < \varepsilon \ll 1$, we have

$$\begin{aligned}
-\int \frac{d\psi_\varepsilon}{dt} \varphi^2\left(\frac{x}{R}\right)|U|^2 + 2\int \psi_\varepsilon \varphi^2\left(\frac{x}{R}\right)|\nabla U|^2 \\
\leq \int \psi_\varepsilon \Delta\left(\varphi^2\left(\frac{x}{R}\right)\right)|U|^2 + \frac{2}{R}\int \psi_\varepsilon \varphi\left(\frac{x}{R}\right)(\nabla\varphi)\left(\frac{x}{R}\right) \cdot |U|^2 U \\
+ \frac{4}{R}\int \psi_\varepsilon \varphi\left(\frac{x}{R}\right)(\nabla\varphi)\left(\frac{x}{R}\right) \cdot PU + \frac{2}{R}\int \psi_\varepsilon \varphi\left(\frac{x}{R}\right)(\nabla\varphi)\left(\frac{x}{R}\right) \cdot |U|^2 b \\
+ \frac{4}{R}\int \psi_\varepsilon \varphi\left(\frac{x}{R}\right) a \otimes U : U \otimes (\nabla\varphi)\left(\frac{x}{R}\right) + 2\int \psi_\varepsilon \varphi^2\left(\frac{x}{R}\right) a \otimes U : \nabla U,
\end{aligned} \tag{2-78}$$

where all the integrals are taken over Q_T. Since U is in the energy space and $P \in L^2_t L^{\frac{3}{2}}_x(Q_T) + L^2(Q_T)$, we may take the limit as $R \uparrow \infty$ to obtain

$$-\int \frac{d\psi_\varepsilon}{dt} |U|^2 + 2\int \psi_\varepsilon |\nabla U|^2 \leq 2\int \psi_\varepsilon a \otimes U : \nabla U. \tag{2-79}$$

Moreover, if t_0, t_1 are Lebesgue points of $\|U(\cdot,t)\|_{L^2(\mathbb{R}^3)}$, then in the limit as $\varepsilon \downarrow 0$,

$$\int_{\mathbb{R}^3} |U(t_1)|^2 + 2\int_{t_0}^{t_1}\int_{\mathbb{R}^3} |\nabla U|^2 \leq \int_{\mathbb{R}^3} |U(t_0)|^2 + 2\int_{t_0}^{t_1}\int_{\mathbb{R}^3} a \otimes U : \nabla U. \tag{2-80}$$

The case when the initial time is zero is recovered from (2-80) by taking the limit as $t_0 \downarrow 0$, since $\lim_{t\downarrow 0}\|U(\cdot,t)-U_0\|_{L^2(\mathbb{R}^3)} = 0$ was already demonstrated in (2-58). □

Existence of Calderón solutions. We are now ready to prove the existence of Calderón solutions.

Theorem 2.5 (existence of Calderón solutions). *Let $T > 0$ and $3 < p < \infty$. Suppose $u_0 \in \dot{B}^{s_p}_{p,p}(\mathbb{R}^3)$ is a divergence-free vector field. Then there exists a Calderón solution u on Q_T with initial data u_0.*

Proof. (1) *Splitting arguments.* Assume the hypotheses of the theorem, and let $q \in (p,\infty)$. According to Lemma 2.2, there exists $0 < \varepsilon < -s_q$ such that for all $M > 0$, we may decompose the initial data as

$$u_0 = U_0 + V_0, \tag{2-81}$$

where

$$U_0 \in L^2(\mathbb{R}^3), \quad \|U_0\|_{L^2(\mathbb{R}^3)} \le C(\|u_0\|_{\dot{B}^{s_p}_{p,p}(\mathbb{R}^3)}, M), \tag{2-82}$$

$$V_0 \in \dot{B}^{s_q+\varepsilon}_{q,q}(\mathbb{R}^3), \quad \|V_0\|_{\dot{B}^{s_q+\varepsilon}_{q,q}(\mathbb{R}^3)} < M. \tag{2-83}$$

The decomposition depends on $M > 0$. By Theorem A.7, there exists a constant $\gamma := \gamma(q,\varepsilon,T) > 0$ such that whenever

$$\|V_0\|_{\dot{B}^{s_q+\varepsilon}_{q,q}(\mathbb{R}^3)} \le \gamma, \tag{2-84}$$

there exists a unique mild solution $V \in \mathcal{K}^{s_q+\varepsilon}_q(Q_T)$ of the Navier–Stokes equations on Q_T with initial data V_0, and the mild solution obeys

$$\|V\|_{\mathcal{K}^{s_q+\varepsilon}_q(Q_T)} \le c(q,\varepsilon)\|V_0\|_{\dot{B}^{s_q+\varepsilon}_{q,q}(\mathbb{R}^3)}. \tag{2-85}$$

Let $M = \gamma$ when forming the decomposition (2-81). The corresponding mild solution V with initial data V_0 exists on Q_T and satisfies (2-85). Hence,

$$V \in L^l_t L^q_x(Q_T), \quad \frac{2}{l} + \frac{3}{q} = 1. \tag{2-86}$$

Let $U \in L^\infty_t L^2_x \cap L^2_t \dot{H}^1_x(Q_T)$ be a solution constructed in Proposition 2.3 to the perturbed Navier–Stokes equation (2-20) with initial data $U_0 \in L^2(\mathbb{R}^3)$ and coefficients $a = b = V$. If $u = U + V$ satisfies the local energy inequality (2-9), then u is a Calderón solution on Q_T with initial data u_0.

(2) *Suitability of full solution.* Let us return to the approximation procedure for U in Proposition 2.3. Define

$$u^\rho := U^\rho + V, \quad p^\rho := P^\rho + Q, \quad Q := (-\Delta)^{-1} \operatorname{div}\operatorname{div} V \otimes V. \tag{2-87}$$

We will prove that the solution $u = U + V$ with pressure $p = P + Q$ satisfies

$$\partial_t |u|^2 + 2|\nabla u|^2 = \Delta |u|^2 - \operatorname{div}((|u|^2 + 2p)u) - 2\mu \tag{2-88}$$

in the sense of distributions on Q_T, where μ is a finite nonnegative measure on Q_T. Let us start from the mollified local energy equality for u^ρ,

$$\begin{aligned}\partial_t |u^\rho|^2 + 2|\nabla u^\rho|^2 = \Delta |u^\rho|^2 - \operatorname{div}(|u^\rho|^2 (u^\rho)_\rho + 2p^\rho u^\rho) \\ - u^\rho \operatorname{div}\big(u^\rho \otimes (V - (V)_\rho) + V \otimes (U^\rho - (U^\rho)_\rho)\big).\end{aligned} \tag{2-89}$$

This is the same equality as if u^ρ solved the mollified Navier–Stokes equations, see, e.g., [Lemarié-Rieusset 2002, p. 318], except for the second line, which adjusts for the fact that V solves the actual Navier–Stokes equations instead of the mollified equations. The distribution on the second line converges to zero as $\rho_k \downarrow 0$, and all the other convergence arguments are as in Step 2 of Proposition 2.3. $\square$

We now demonstrate that there exists a Calderón solution which agrees with the mild solution until its maximal time of existence.

Theorem 2.6 (mild Calderón solutions). *Under the assumptions of Theorem 2.5, there exists a Calderón solution u on Q_T with initial data u_0 such that u agrees with the mild solution* $\mathrm{NS}(u_0)$ *until the time* $\min(T, T^*(u_0))$.

The mild solution $\mathrm{NS}(u_0)$ under consideration is constructed in Theorem A.2.

Proof. (1) *Introducing the integral equation.* Let U_0, V_0, V be as in Step 1 of Theorem 2.5. From now on, we will define $T_\sharp := T$ in the statement of the theorem, so that the variable T can be reused.

The set-up is as follows. Recall that for all $0 < T < \min(T_\sharp, T^*)$,

$$\mathrm{NS}(u_0) \in \mathring{\mathcal{K}}_p(Q_T) \cap \mathring{\mathcal{K}}_\infty(Q_T), \tag{2-90}$$

$$V \in \mathcal{K}_q^{s_q+\varepsilon}(Q_T) \cap \mathcal{K}_\infty^{-1+\varepsilon}(Q_T). \tag{2-91}$$

Let us define $\widetilde{U} := \mathrm{NS}(u_0) - V$. Then $\widetilde{U} \in \mathring{\mathcal{K}}_q(Q_T) \cap \mathring{\mathcal{K}}_\infty(Q_T)$ for all $0 < T < \min(T_\sharp, T^*)$ is a mild solution of the integral equation

$$W(\cdot, t) = e^{t\Delta} U_0 - B(W, W)(\cdot, t) - L(W)(\cdot, t), \tag{2-92}$$

where we formally define

$$B(v, w)(\cdot, t) := \int_0^t e^{(t-s)\Delta} \mathbb{P} \operatorname{div} v \otimes w \, ds, \tag{2-93}$$

$$L(w)(\cdot, t) := B(w, V)(\cdot, t) + B(V, w)(\cdot, t) \tag{2-94}$$

for all vector fields v, w on spacetime. The initial data U_0 belongs to the class

$$U_0 \in [\dot{B}^{s_p}_{p,p}(\mathbb{R}^3) + \dot{B}^{s_q+\varepsilon}_{q,q}(\mathbb{R}^3)] \cap L^2(\mathbb{R}^3). \tag{2-95}$$

First, we will demonstrate that the mild solution of the integral equation (2-92) in $\mathring{\mathcal{K}}_q(Q_T)$ is unique, where $0 < T < \min(T_\sharp, T^*)$. Notice that in the decomposition $u = U + V$ of a Calderón solution, the vector field U formally solves (2-92). We will show that in the proof of Theorem 2.5, it is possible to construct U in the space $\mathring{\mathcal{K}}_q(Q_T)$. Therefore, U will satisfy $U \equiv \widetilde{U}$ on Q_T, and the proof will be complete.

(2) *Existence of mild solutions to integral equation.* Let us summarize the local well-posedness of mild solutions $W \in \mathring{\mathcal{K}}_q(Q_T)$ to the integral equation (2-92). Our main goal is to establish estimates on the operators B and L. Then the local existence of mild solutions in $\mathring{\mathcal{K}}_q(Q_S)$ for some $0 < S \ll 1$ will follow from Lemma A.3.

Let $0 < T < \min(T_\sharp, T^*)$ and $v, w \in \mathring{\mathcal{K}}_q(Q_T)$. First, observe that $L : \mathring{\mathcal{K}}_q(Q_T) \to \mathring{\mathcal{K}}_q(Q_T)$ is bounded with operator norm satisfying $\|L\|_{\mathring{\mathcal{K}}_q(Q_T)} < 1$ when $0 < T \ll 1$. Indeed, according to the Kato estimates in Lemma A.1,

$$\|L(v)\|_{\mathcal{K}_q(Q_T)} = \|B(v, V) + B(V, v)\|_{\mathcal{K}_q(Q_T)} \le cT^{\frac{\varepsilon}{2}} \|v\|_{\mathcal{K}_q(Q_T)} \|V\|_{\mathcal{K}_q^{s_q+\varepsilon}(Q_T)}. \tag{2-96}$$

That L preserves the decay properties near the initial time follows from examining the limit $T \downarrow 0$ in the estimates above. One may also show that B is bounded on $\mathring{\mathcal{K}}_q(Q_T)$ with norm independent of T. Finally, due to (2-95), we have the property

$$\lim_{T\downarrow 0} \|e^{t\Delta} U_0\|_{\mathcal{K}_q(Q_T)} = 0. \tag{2-97}$$

Hence, Lemma A.3 implies the existence of a mild solution $W \in X_S$ to the integral equation (2-92) for a time $0 < S \ll 1$. The solution $W(\cdot, t)$ is also continuous in $L^q(\mathbb{R}^3)$ after the initial time. As for uniqueness, note that each mild solution $W \in \mathring{\mathcal{K}}_q(Q_T)$ of (2-92) obeys the property that $W + V \in \mathring{\mathcal{K}}_q(Q_T)$ is a mild solution of the Navier–Stokes equations, so $W + V \equiv \mathrm{NS}(u_0)$ and $W \equiv \widetilde{U}$ on Q_T.

(3) *U is in $\mathring{\mathcal{K}}_q(Q_S)$ for $0 < S \ll 1$.* Let us return to the approximations U^ρ constructed in the proofs of Theorem 2.5 and Proposition 2.3. Recall that U^ρ solves the system

$$\begin{gathered} \partial_t U^\rho - \Delta U^\rho + \operatorname{div} U^\rho \otimes (U^\rho)_\rho + \operatorname{div} U^\rho \otimes V + \operatorname{div} V \otimes U^\rho = -\nabla P^\rho, \\ \operatorname{div} U^\rho = 0, \end{gathered} \tag{2-98}$$

with initial condition $U^\rho(\cdot, 0) = U_0$. By repeating the arguments in Step 2, there exists a time $S > 0$ independent of the parameter $\rho > 0$ and a unique mild solution $W^\rho \in \mathring{\mathcal{K}}_q(Q_S)$ of the integral equation

$$W^\rho(\cdot, t) = e^{t\Delta} U_0 - B_\rho(W, W)(\cdot, t) - L(W)(\cdot, t), \tag{2-99}$$

where the operator B_ρ is defined by

$$B_\rho(v, w)(t) := \int_0^t e^{(t-s)\Delta} \mathbb{P} \operatorname{div} v \otimes (w)_\rho \, ds. \tag{2-100}$$

Since $U_0 \in L^2(\mathbb{R}^3)$, we may shorten the time $S > 0$ so that $W^\rho \in L^\infty_t L^2_x(Q_S)$. This assertion follows from applying Lemma A.3 in the intersection space $\mathring{\mathcal{K}}_q(Q_S) \cap L^\infty_t L^2_x(Q_S)$ with $0 < S \ll 1$, once we have the necessary estimates

$$\|L(v)\|_{L^\infty_t L^2_x(Q_T)} \le cT^{\frac{\varepsilon}{2}} \|v\|_{L^\infty_t L^2_x(Q_T)} \|V\|_{\mathcal{K}_q^{s_q+\varepsilon}(Q_T)}, \tag{2-101}$$

$$\|B_\rho(v, w)\|_{L^\infty_t L^2_x(Q_T)} \le c\|v\|_{L^\infty_t L^2_x(Q_T)} \|w\|_{\mathcal{K}_q(Q_T)}. \tag{2-102}$$

In fact, we obtain that $W^\rho \in C([0, S]; L^2(\mathbb{R}^3) \cap L^2_t \dot{H}^1_x(Q_S)$. This follows from viewing W^ρ as a solution of the mollified Navier–Stokes equations (no lower-order terms) with initial data $U_0 \in L^2(\mathbb{R}^3)$ and forcing term $\operatorname{div} F$, where

$$F = -W^\rho \otimes V - V \otimes W^\rho \in L^2(Q_S), \tag{2-103}$$

since $V \in \mathcal{K}_\infty^{-1+\varepsilon}(Q_S)$ and $W^\rho \in L^\infty_t L^2_x(Q_S)$.

Finally, we pass to the limit as $\rho \downarrow 0$. The only possible issue is that $\mathring{\mathcal{K}}_q(Q_S)$ is not closed under weak-star limits, but this will not be a problem because Lemma A.3 actually gave the uniform bound

$$\|U^\rho\|_{\mathcal{K}_q(Q_T)} \le \kappa(T), \quad 0 < T \le T_0, \tag{2-104}$$

where $\kappa(T)$ is a nonnegative continuous function on the interval $[0, S]$ and $\kappa(0) = 0$. Now by taking the weak-star limit along a subsequence $\rho_k \downarrow 0$, we obtain

$$\|U\|_{\mathcal{K}_q(Q_T)} \le \kappa(T), \quad 0 < T \le S. \tag{2-105}$$

Together with the convergence arguments in Proposition 2.3, this implies that $U \in \mathring{\mathcal{K}}_q(Q_S)$ is a mild solution of the integral equation (2-92). By the uniqueness argument in Step 2, $U \equiv \widetilde{U}$ on Q_S.

(4) *Agreement until* $T^*(u_0)$. So far, we have only shown that $U \equiv \widetilde{U}$ on a short time interval $[0, S]$. After the initial time, the mild solutions $\mathrm{NS}(u_0)$ and V of the Navier–Stokes equations are in subcritical spaces:

$$\mathrm{NS}(u_0), V \in C([t_0, \min(T_\sharp, T^*)); L^q(\mathbb{R}^3)) \tag{2-106}$$

for all $0 < t_0 < \min(T_\sharp, T^*)$. We will prove that $\widetilde{U}$ is in the energy space and then argue via weak-strong uniqueness. To do so, one may develop the existence theory for mild solutions $W \in C([t_0, T]; L^q(\mathbb{R}^3))$ of the integral equation

$$W(\cdot, t) = e^{(t-t_0)\Delta}\widetilde{U}(\cdot, t_0) - \int_{t_0}^t e^{(t-s)\Delta}\mathbb{P}\,\mathrm{div}\, W \otimes W\, ds - \int_{t_0}^t e^{(t-s)\Delta}\mathbb{P}\,\mathrm{div}(W \otimes V + V \otimes W)\, ds, \tag{2-107}$$

where $t_0 \in [T_0, \min(T_\sharp, T^*))$. Note that whenever $\widetilde{U}(\cdot, t_0) \in L^2(\mathbb{R}^3)$, the unique mild solution W will also be in the energy space. The proof is similar to Steps 2–3. In this way, one obtains that

$$\widetilde{U} \in C([0, T); L^2(\mathbb{R}^3)) \cap L^2_t \dot{H}^1_x(Q_T) \tag{2-108}$$

for all $0 < T < \min(T_\sharp, T^*)$ and satisfies the energy equality

$$\int_{\mathbb{R}^3} |\widetilde{U}(x,t_2)|^2\, dx + 2\int_{t_1}^{t_2}\int_{\mathbb{R}^3} |\nabla\widetilde{U}|^2\, dx\, dt = \int_{\mathbb{R}^3} |\widetilde{U}(x,t_1)|^2\, dx + 2\int_{t_1}^{t_2}\int_{\mathbb{R}^3} V\otimes\widetilde{U} : \nabla\widetilde{U}\, dx\, dt \tag{2-109}$$

for all $0 \le t_1 < t_2 < \min(T_\sharp, T^*)$. In fact, by the Gronwall-type argument in Lemma 2.4, one may actually show that $\widetilde{U} \in L^2_t \dot{H}^1_x(Q_{\min(T_\sharp, T^*)})$. In addition, we know that U obeys the global energy inequality (2-23) for almost every $0 \le t_1 < T_\sharp$, including $t_1 = 0$, and for every $t_2 \in (t_1, T_\sharp]$. Let us write $D := U - \widetilde{U}$. Then D obeys the energy inequality

$$\int_{\mathbb{R}^3} |D(x,t_2)|^2\, dx + 2\int_{t_1}^{t_2}\int_{\mathbb{R}^3} |\nabla D|^2\, dx\, dt \le 2\int_{t_1}^{t_2}\int_{\mathbb{R}^3} (\widetilde{U} + V) \otimes D : \nabla D\, dx\, dt, \tag{2-110}$$

where $t_1 := \frac{1}{2}T_0$. To obtain the energy inequality for D, one must write $|D|^2 = |U|^2 + |\widetilde{U}|^2 - 2U \cdot \widetilde{U}$, $|\nabla D|^2 = |\nabla U|^2 + |\nabla\widetilde{U}|^2 - 2\nabla U : \nabla\widetilde{U}$, and utilize the identity

$$\int_{\mathbb{R}^3} U(x,t) \cdot \widetilde{U}(x,t)\, dx + 2\int_0^t\int_{\mathbb{R}^3} \nabla U : \nabla\widetilde{U}\, dx\, dt' - \int_{\mathbb{R}^3} U(x,0) \cdot W(x,0)\, dx = \int_0^t\int_{\mathbb{R}^3} (\partial_t U - \Delta U) \cdot \widetilde{U} + U \cdot (\partial_t\widetilde{U} - \Delta\widetilde{U})\, dx\, dt'. \tag{2-111}$$

This argument is typical in weak-strong uniqueness proofs. The identity (2-111) is clear for smooth vector fields with compact support, but it may be applied in more general situations by approximation. As in Lemma 2.4, the energy inequality gives

$$|D|^2_{2,Q_{t_1,t_2}} \leq B(t_1,t_2)|D|^2_{2,Q_{t_1,t_2}}, \tag{2-112}$$

where we have defined

$$B(t_1,t_2) := c(\|\widetilde{U}\|_{L^l_t L^q_x(Q_{t_1,t_2})} + \|V\|_{L^l_t L^q_x(Q_{t_1,t_2})}). \tag{2-113}$$

One then takes t_2 close to t_1 such that $B(t_1,t_2) \leq \frac{1}{2}$ to obtain $D \equiv 0$ on Q_{t_1,t_2}. The equality $U \equiv \widetilde{U}$ may be propagated forward until $\min(T_\sharp, T^*)$ by repeating the argument (even if $\|\widetilde{U}\|_{L^l_t L^q_x(Q_{t_1,T^*})} = \infty$). □

Finally, here is a result that contains the limiting arguments we will use in Theorem 1.1. Namely, we demonstrate that a weakly converging sequence of initial data has a corresponding subsequence of Calderón solutions that converges locally strongly to a solution of the Navier–Stokes equations.

As a reminder, we do not prove that the limit solution is a Calderón solution, though we expect such a result to be true. The issue is that the limit solution does not evidently satisfy the energy inequality starting from the initial time.

Theorem 2.7 (weak convergence of Calderón solutions). *Let* $3 < p < \infty$ *and* $T > 0$. *Suppose* $(u_0^{(n)})_{n\in\mathbb{N}}$ *is a sequence of divergence-free vector fields such that*

$$u_0^{(n)} \rightharpoonup u_0 \quad \text{in } \dot{B}^{s_p}_{p,p}(\mathbb{R}^3). \tag{2-114}$$

Then for each $n \in \mathbb{N}$, *there exists a Calderón solution* $u^{(n)}$ *on* Q_T *with initial data* $u_0^{(n)}$ *and associated pressure* $p^{(n)}$ *such that the solution* $u^{(n)}$ *agrees with the mild solution* $\mathrm{NS}(u_0^{(n)})$ *until time* $\min(T, T^*(u_0^{(n)}))$.

Furthermore, there exists a distributional solution (u, p) *of the Navier–Stokes equations on* Q_T *with the following properties*:

$$u_0 = U_0 + V_0, \quad u = U + V, \tag{2-115}$$

where

$$U_0 \in L^2(\mathbb{R}^3), \quad V_0 \in \dot{B}^{s_q+\varepsilon}_{q,q}(\mathbb{R}^3), \tag{2-116}$$

$$U \in L^\infty_t L^2_x \cap L^2_t \dot{H}^1_x(Q_T), \quad V \in \mathcal{K}^{s_q+\varepsilon}_q(Q_T), \tag{2-117}$$

$$q > p, \quad 0 < \varepsilon < -s_q, \tag{2-118}$$

V *is a mild solution of the Navier–Stokes equations on* Q_T *with initial data* V_0, *and* U *solves the perturbed Navier–Stokes system* (2-5) *in* Q_T. *The vector field* $U(\cdot,t)$ *is weakly continuous as an* $L^2(\mathbb{R}^3)$-*valued function on* $[0,T]$ *and satisfies* $U(\cdot,0) = U_0 \in L^2(\mathbb{R}^3)$. *The solution* (u,p) *is suitable in* Q_T *in the sense of* (2-9), *and* U *satisfies the local energy inequality* (2-10) *in* Q_T *and the global energy inequality* (2-11) *for almost every* $0 < t_1 < T$ *and for all* $t_2 \in (t_1, T]$.

Finally, there exists a subsequence, still denoted by n, *such that the Calderón solutions* $u^{(n)}$ *converge to* u *in the following senses*:

$$u^{(n)} \to u \quad \text{in } L^3_{\mathrm{loc}}(\mathbb{R}^3\times]0,T]), \qquad p^{(n)} \rightharpoonup p \quad \text{in } L^{\frac{3}{2}}_{\mathrm{loc}}(\mathbb{R}^3\times]0,T]),$$

$$u^{(n)} \to u \quad \text{in } C([0,T];\mathcal{S}'(\mathbb{R}^3)). \tag{2-119}$$

In particular,

$$u^{(n)}(\cdot,t) \overset{*}{\rightharpoonup} u(\cdot,t) \quad \text{in } \mathcal{S}'(\mathbb{R}^3), \quad 0 \le t \le T. \tag{2-120}$$

Proof. (1) *Splitting*. Let us assume the hypotheses of the theorem. There exists a constant $A > 0$ such that

$$\|u_0^{(n)}\|_{\dot{B}^{s_p}_{p,p}(\mathbb{R}^3)} \le A. \tag{2-121}$$

We will follow the proof of Theorem 2.5. Let $q \in (p,\infty)$. The constants below are independent of n but may depend on p, q, T. According to Lemma 2.2, there exists $0 < \varepsilon < -s_q$ such that for all $n \in \mathbb{N}$,

$$u_0^{(n)} = U_0^{(n)} + V_0^{(n)}, \tag{2-122}$$

where

$$U_0^{(n)} \in L^2(\mathbb{R}^3), \qquad \|U_0^{(n)}\|_{L^2(\mathbb{R}^3)} \le C(A, M), \tag{2-123}$$

$$V_0^{(n)} \in \dot{B}^{s_q+\varepsilon}_{q,q}(\mathbb{R}^3), \quad \|V_0^{(n)}\|_{\dot{B}^{s_q+\varepsilon}_{q,q}(\mathbb{R}^3)} < M, \tag{2-124}$$

According to Theorem A.7, we may choose $0 < M \ll 1$ such that the mild solutions $V^{(n)}$ with initial data $V_0^{(n)}$ exist on Q_T and satisfy

$$\|t^{k+\frac{l}{2}} \partial_t^k \nabla^l V^{(n)}\|_{\mathcal{K}^{s_q+\varepsilon}_q(Q_T)} \le c(k,l) \tag{2-125}$$

for all integers $k, l \ge 0$.

(2) *Convergence of* $V^{(n)}$. Due to the estimate (2-125) and the Ascoli–Arzelà theorem, we may pass to a subsequence, still denoted by n, such that

$$V^{(n)} \overset{*}{\rightharpoonup} V \quad \text{in } \mathcal{K}^{s_q+\varepsilon}_q(Q_T), \qquad \partial_t^k \nabla^l V^{(n)} \to \partial_t^k \nabla^l V \quad \text{in } C(K) \tag{2-126}$$

for all $K \subset \mathbb{R}^3 \times (0,T]$ compact and integers $k, l \ge 0$. The Navier–Stokes equations imply

$$\partial_t V^{(n)} = -\Delta V^{(n)} - \mathbb{P} \operatorname{div} V^{(n)} \otimes V^{(n)} \tag{2-127}$$

in the sense of tempered distributions on Q_T, and one may estimate the time derivative $\partial_t V^{(n)}$ by the right-hand side of (2-127):

$$\sup_{n\in\mathbb{N}} \|\Delta V^{(n)}\|_{L^l_t W^{-2,q}_x(Q_T)} + \|\mathbb{P} \operatorname{div} V^{(n)} \otimes V^{(n)}\|_{L^{l/2}_t W^{-1,q/2}_x(Q_T)} < \infty, \tag{2-128}$$

where $\frac{2}{l} + \frac{3}{q} = 1$. Hence, there exists a subsequence such that

$$\int_{\mathbb{R}^3} V^{(n)}(x,\cdot)\cdot\varphi\,dx \to \int_{\mathbb{R}^3} V(x,\cdot)\cdot\varphi\,dx \quad \text{in } C([0,T]), \quad \varphi \in \mathcal{S}(\mathbb{R}^3). \tag{2-129}$$

By the Calderón–Zygmund estimates and (2-125), the associated pressures satisfy

$$Q^{(n)} := (-\Delta)^{-1} \operatorname{div}\operatorname{div} V^{(n)} \otimes V^{(n)} \rightharpoonup Q \quad \text{in } L^{\frac{l}{2}}_t L^{\frac{q}{2}}_x(Q_T). \tag{2-130}$$

In particular, the convergence occurs weakly in $L^{\frac{3}{2}}_{\text{loc}}(Q_T)$. By similar arguments to those in Step 1 of Proposition 2.3, the pressure satisfies $Q = (-\Delta)^{-1} \operatorname{div}\operatorname{div} V \otimes V$, and (V, Q) solves the Navier–Stokes equations on Q_T in the sense of distributions. Therefore, since $V \in \mathcal{K}^{s_q+\varepsilon}_q(Q_T)$ and $V(\cdot,t) \overset{*}{\rightharpoonup} V_0$ in tempered distributions as $t \downarrow 0$, we conclude that V is the mild solution of the Navier–Stokes equations on Q_T with initial data V_0 as in Theorem A.7. See [Lemarié-Rieusset 2002, p. 122] for further remarks on the equivalence between differential and integral forms of the Navier–Stokes equations.

(3) *Convergence of* $U^{(n)}$. For each $n \in \mathbb{N}$, let $U^{(n)} \in L^\infty_t L^2_x \cap L^2_t \dot{H}^1_x(Q_T)$, whose existence is guaranteed by Theorem 2.6, so that the Calderón solution

$$u^{(n)} = U^{(n)} + V^{(n)} \tag{2-131}$$

agrees with the mild solution $\mathrm{NS}(u_0^{(n)})$ up to $\min(T, T^*(u_0^{(n)}))$. It remains to consider the limit of $U^{(n)}$. The proof is very similar to the proofs of Proposition 2.3 and Theorem 2.5, so we will merely summarize what must be done. Recall from Step 1 that $\|U_0^{(n)}\|_{L^2(\mathbb{R}^3)} \leq C(A, M)$. We use the energy inequality and a Gronwall-type argument to obtain

$$|U^{(n)}|^2_{2,Q_T} \leq C(A, M). \tag{2-132}$$

Hence, we may take weak-star limits in the energy space upon passing to a subsequence. As before, we may estimate the time derivative using the perturbed Navier–Stokes equations. The result is that

$$\|\partial_t U^{(n)}\|_{L^2_t H^{-3/2}_x(Q_T)} \leq C(A, M). \tag{2-133}$$

Consequently, we may extract a further subsequence such that $U^{(n)} \to U$ in $L^3_t (L^3_{\mathrm{loc}})_x(Q_T)$, the limit $U(\cdot, t)$ is weakly continuous on $[0, T]$ as an $L^2(\mathbb{R}^3)$-valued function, and

$$\int_{\mathbb{R}^3} U^{(n)}(x, \cdot) \cdot \varphi(x)\, dx \to \int_{\mathbb{R}^3} U(x, \cdot) \cdot \varphi(x)\, dx \quad \text{in } C([0, T]), \quad \varphi \in L^2(\mathbb{R}^3). \tag{2-134}$$

Now recall the associated pressures $P^{(n)} := P_1^{(n)} + P_2^{(n)}$,

$$P_1^{(n)} := (-\Delta)^{-1} \operatorname{div} \operatorname{div} U^{(n)} \otimes U^{(n)}, \tag{2-135}$$

$$P_2^{(n)} := (-\Delta)^{-1} \operatorname{div} \operatorname{div} (U^{(n)} \otimes V^{(n)} + V^{(n)} \otimes U^{(n)}). \tag{2-136}$$

By the Calderón–Zygmund estimates, we pass to a subsequence to obtain

$$P_1^{(n)} \rightharpoonup P_1 \quad \text{in } L^2_t L^{\frac{3}{2}}_x(Q_T), \qquad P_2^{(n)} \rightharpoonup P_2 \quad \text{in } L^2(Q_T), \tag{2-137}$$

$$P^{(n)} \rightharpoonup P := P_1 + P_2 \quad \text{in } L^{\frac{3}{2}}_{\mathrm{loc}}(Q_T). \tag{2-138}$$

Since $-\Delta P_1 = \operatorname{div} \operatorname{div} U \otimes U$ and $-\Delta P_2 = \operatorname{div} \operatorname{div}(U \otimes V + V \otimes U)$ in the sense of distributions on Q_T, we obtain from the Liouville theorem that

$$P_1 = (-\Delta)^{-1} \operatorname{div} \operatorname{div} U \otimes U, \quad P_2 = (-\Delta)^{-1} \operatorname{div} \operatorname{div}(U \otimes V + V \otimes U). \tag{2-139}$$

Define $p := P + Q$. It is clear from the limiting procedure that (u, p) satisfies the Navier–Stokes equations on Q_T and (U, P) satisfies the perturbed Navier–Stokes equations on Q_T. Each is satisfied in the sense of distributions.

As in the proofs of Step 2 in Proposition 2.3 and Step 2 in Theorem 2.5, respectively, we may show that the limit (u, p) is a suitable weak solution of the Navier–Stokes equations on Q_T in the sense of (2-9) and that (U, P) satisfies the local energy inequality (2-10) on Q_T. Moreover, as in Step 3 of Proposition 2.3, (2-10) implies the global energy inequality (2-11) for almost every $0 < t_1 < T$ and for all $t_2 \in (t_1, T]$. $\square$

3. Proof of main result

We are ready to prove Theorem 1.1. The proof follows the scheme set forth in [Seregin 2012a] except that we use Calderón solutions to take the limit of the rescaled solutions.

Proof of Theorem 1.1. Let us assume the hypotheses of the theorem. According to the chain of embeddings (A-48), we have

$$u_0 \in \dot{B}^{s_p}_{p,q}(\mathbb{R}^3) \hookrightarrow \dot{B}^{s_m}_{m,m}(\mathbb{R}^3), \quad m := \max(p, q). \tag{3-1}$$

By the uniqueness results in Theorem A.2, the notion of mild solution and maximal time of existence are unchanged by considering the larger homogeneous Besov space. Thus, without loss of generality, we will assume $p = q = m$.

(1) *Rescaling*. Let u be the mild solution of the Navier–Stokes equations with divergence-free initial data $u_0 \in \dot{B}^{s_p}_{p,p}(\mathbb{R}^3)$ and $T^*(u_0) < \infty$ from the statement of the theorem. In Corollary A.6, we proved that u must form a singularity at time $T^*(u_0)$. By the translation and scaling symmetries of the Navier–Stokes equations, we may assume that the singularity occurs at the spatial origin and time $T^*(u_0) = 1$.

Suppose for contradiction that there exists a sequence $t_n \uparrow 1$ and constant $M > 0$ such that

$$\|u(\cdot, t_n)\|_{\dot{B}^{s_p}_{p,p}(\mathbb{R}^3)} \le M. \tag{3-2}$$

The solution $u(\cdot, t)$ is continuous on $[0, 1]$ in the sense of tempered distributions (for instance, because u agrees with a Calderón solution on Q_T). We must have

$$\|u(\cdot, 1)\|_{\dot{B}^{s_p}_{p,p}(\mathbb{R}^3)} \le M. \tag{3-3}$$

Let us zoom in around the singularity. For each $n \in \mathbb{N}$, we define

$$u^{(n)}(x, t) := \lambda_n u(\lambda_n x, t_n + \lambda_n^2 t), \quad (x, t) \in Q_1, \tag{3-4}$$

where $\lambda_n := (1 - t_n)^{\frac{1}{2}}$. Then $u^{(n)}$ is the mild solution of the Navier–Stokes equations on Q_1 with divergence-free initial data $u_0^{(n)} = \lambda_n u(\lambda_n x, t_n)$, and

$$\|u_0^{(n)}\|_{\dot{B}^{s_p}_{p,p}(\mathbb{R}^3)} \le M. \tag{3-5}$$

Let us pass to a subsequence, still denoted by n, such that

$$u_0^{(n)} \rightharpoonup v_0 \quad \text{in } \dot{B}^{s_p}_{p,p}(\mathbb{R}^3). \tag{3-6}$$

(2) *Limiting procedure*. We now apply Theorem 2.7 to the weakly converging sequence $(u_0^{(n)})_{n \in \mathbb{N}}$. For each $n \in \mathbb{N}$, there exists a Calderón solution $u^{(n)}$ on Q_1 with initial data $u_0^{(n)}$ such that $u^{(n)}$ agrees with

the mild solution $\mathrm{NS}(u_0^{(n)})$ on Q_1. Furthermore, by passing to a subsequence, still denoted by n, we have

$$u^{(n)} \to v \quad \text{in } L^3_{\mathrm{loc}}(Q_1), \qquad p^{(n)} \rightharpoonup q \quad \text{in } L^{\frac{3}{2}}_{\mathrm{loc}}(Q_1), \tag{3-7}$$

where (v, q) solves the Navier–Stokes equations in the sense of distributions on Q_1 and satisfies the many additional properties listed in Theorem 2.7. In particular, $v \in C([0,1]; \mathcal{S}'(\mathbb{R}^3))$. According to Proposition A.9 concerning persistence of the singularity, the solution v also has a singularity at the spatial origin and time $T = 1$.

Next, we observe that the solution v vanishes identically at time $T = 1$:

$$v(\cdot, 1) = 0. \tag{3-8}$$

Indeed, Theorem 2.7 implies that

$$u^{(n)}(\cdot, 1) \overset{*}{\rightharpoonup} v(\cdot, 1) \quad \text{in } \mathcal{S}'(\mathbb{R}^3), \tag{3-9}$$

and by the scaling property of $u(\cdot, 1) \in \dot{B}^{s_p}_{p,p}(\mathbb{R}^3)$,

$$\langle u^{(n)}(\cdot, 1), \varphi \rangle = \langle u(\cdot, 1), \lambda_n^{-2} \varphi(\cdot / \lambda_n) \rangle \to 0, \quad \varphi \in \mathcal{S}(\mathbb{R}^3). \tag{3-10}$$

The property (3-10) is a consequence of the density of Schwartz functions in $\dot{B}^{s_p}_{p,p}(\mathbb{R}^3)$. It is certainly true with $u(\cdot, 1)$ replaced by a Schwartz function $\psi \in \mathcal{S}(\mathbb{R}^3)$, and therefore,

$$\begin{aligned} |\langle u(\cdot, 1), \lambda_n^{-2} \varphi(\cdot / \lambda_n) \rangle| &\le |\langle \psi, \lambda_n^{-2} \varphi(\cdot / \lambda_n) \rangle| + |\langle u(\cdot, 1) - \psi, \lambda_n^{-2} \varphi(\cdot / \lambda_n) \rangle| \\ &\le o(1) + c \|u(\cdot, 1) - \psi\|_{\dot{B}^{s_p}_{p,p}(\mathbb{R}^3)} \|\varphi\|_{\dot{B}^{-s_p}_{p',p'}(\mathbb{R}^3)} \quad \text{as } n \to \infty, \end{aligned} \tag{3-11}$$

where $\frac{1}{p} + \frac{1}{p'} = 1$, for all $\varphi \in \mathcal{S}(\mathbb{R}^3)$.

(3) *Backward uniqueness*. Our goal is to demonstrate that

$$\omega := \operatorname{curl} v \equiv 0 \quad \text{on } Q_{\frac{1}{2},1}. \tag{3-12}$$

Suppose temporarily that (3-12) is satisfied. From the well-known vector identity $\Delta v = \nabla \operatorname{div} v - \operatorname{curl} \operatorname{curl} v$, we obtain

$$\Delta v = 0 \quad \text{on } Q_{\frac{1}{2},1}. \tag{3-13}$$

Now the Liouville theorem for entire harmonic functions and the decomposition

$$v = U + V \in L^3(Q_1) + L^\infty_t L^q_x(Q_{\delta,1}), \quad 0 < \delta < 1, \tag{3-14}$$

from Theorem 2.7 imply that $v \equiv 0$ on $Q_{\frac{1}{2},1}$. This contradicts that v is singular at time $T = 1$ and finishes the proof.

We will now prove (3-12). Based on (3-14) and

$$q = P + Q \in L^{\frac{3}{2}}(Q_1) + L^\infty_t L^{\frac{q}{2}}_x(Q_{\delta,1}), \quad 0 < \delta < 1, \tag{3-15}$$

we have

$$\lim_{|x| \to \infty} \int_{\frac{1}{4}}^1 \int_{B(x,1)} |v|^3 + |q|^{\frac{3}{2}} \, dx \, dt = 0. \tag{3-16}$$

Recall from Theorem 2.7 that (v, q) obeys the local energy inequality (2-9), so it is a suitable weak solution on Q_1. The ε-regularity criterion in Theorem A.8 implies that there exist constants $R, \kappa > 0$,

$K := (\mathbb{R}^3 \setminus B(R)) \times \left(\frac{1}{2}, 1\right)$, such that

$$\sup_K |v| + |\nabla v| + |\nabla^2 v| < \kappa. \tag{3-17}$$

Recall the equation satisfied by the vorticity: $\partial_t \omega - \Delta\omega = -\operatorname{curl}(v\nabla v)$. This implies

$$|\partial_t \omega - \Delta\omega| \le c(|\nabla\omega| + |\omega|) \quad \text{in } K \tag{3-18}$$

for a constant $c > 0$ depending on κ. Also, $w(\cdot, 1) = 0$ due to (3-8). Now, according to Theorem A.10 concerning backward uniqueness, $\omega \equiv 0$ in K.

It remains to demonstrate that $\omega \equiv 0$ in $\overline{B(R)} \times \left(\frac{1}{2}, 1\right)$. We claim that there exists a dense open set $G \subset (0, 1)$ such that v is smooth on $\Omega := \mathbb{R}^3 \times G$. With the claim in hand, let $z_0 = (x_0, t_0) \in \Omega \cap K$ such that $|x_0| = 2R$. Note that $\omega \equiv 0$ in a neighborhood of z_0. In addition, by the smoothness of v, there exist $0 < \varepsilon \ll 1$ and $c > 0$ depending on z_0 such that

$$|\partial_t \omega - \Delta\omega| \le c(|\nabla\omega| + |\omega|) \quad \text{in } Q := B(x_0, 4R) \times (t_0 - \varepsilon, t_0 + \varepsilon) \subset \Omega. \tag{3-19}$$

Hence, the assumptions of Theorem A.11 concerning unique continuation are satisfied in Q, and $\omega \equiv 0$ in Q. This implies that $\omega \equiv 0$ in $\mathbb{R}^3 \times (t_0 - \varepsilon, t_0 + \varepsilon)$. Since $z_0 \in \Omega \cap K$ was arbitrary, we obtain that $\omega \equiv 0$ in Ω. Now the density of G and weak continuity $v \in C([0, 1]; \mathcal{S}'(\mathbb{R}^3))$ imply $\omega \equiv 0$ on $Q_{\frac{1}{2},1}$.

Finally, we prove the claim that there exists a dense open set $G \subset (0, 1)$ such that v is smooth on $\mathbb{R}^3 \times G$. Recall from Theorem 2.7 that $v = U + V$, where V is the smooth mild solution in Q_1 from Theorem A.7 with initial data V_0. To treat U, consider the set Π of times $t_1 \in (0, 1)$ such that $U(\cdot, t_1) \in H^1(\mathbb{R}^3)$ and U satisfies the global energy inequality (2-11) for all $t_2 \in (t_1, 1]$. The latter condition ensures that

$$\lim_{t \downarrow t_1} \|U(\cdot, t) - U(\cdot, t_1)\|_{L^2(\mathbb{R}^3)} = 0. \tag{3-20}$$

The set Π has full measure in $(0, 1)$. For each $t_1 \in \Pi$, there exist $\varepsilon := \varepsilon(t_1) > 0$, a vector field $\widetilde{U} \in C([t_1, t_1 + \varepsilon]; H^1(\mathbb{R}^3))$, and pressure $\widetilde{P} \in L^\infty_t L^3_x(Q_{t_1, t_1+\varepsilon})$ such that

$$\partial_t \widetilde{U} - \Delta\widetilde{U} + \operatorname{div} \widetilde{U} \otimes \widetilde{U} + \operatorname{div} V \otimes \widetilde{U} + \operatorname{div} \widetilde{U} \otimes V = -\nabla\widetilde{P}, \quad \operatorname{div} \widetilde{U} = 0, \tag{3-21}$$

in the sense of distributions on $Q_{t_1, t_1+\varepsilon}$. Furthermore, $\widetilde{U}$ is smooth on $Q_{t_1, t_1+\varepsilon}$. This may be proven by developing the local well-posedness theory for the integral equation

$$\widetilde{U}(\cdot, t) = e^{(t-t_1)\Delta} U(\cdot, t_1) - \int_{t_1}^{t} e^{(t-s)\Delta} \mathbb{P} \operatorname{div}(\widetilde{U} \otimes \widetilde{U} + V \otimes \widetilde{U} + \widetilde{U} \otimes V)\, ds \tag{3-22}$$

with $U(\cdot, t_1) \in H^1(\mathbb{R}^3)$. One must use that $\partial_t^k \nabla^l V \in C([\delta, 1]; L^\infty(\mathbb{R}^3))$ for all $0 < \delta < 1$ and integers $k, l \ge 0$. By weak-strong uniqueness for (3-21), as in Step 4 in Theorem 2.6, we must have that $U \equiv \widetilde{U}$ on $Q_{t_1, t_1+\varepsilon}$. Hence, U is smooth on $Q_{t_1, t_1+\varepsilon}$. We define the dense open set $G \subset (0, 1)$ as follows:

$$G := \bigcup_{t_1 \in \Pi} (t_1, t_1 + \varepsilon(t_1)), \quad \overline{G} \supseteq \overline{\Pi} = [0, 1], \tag{3-23}$$

This completes the proof of the claim and Theorem 1.1. □

Appendix

This appendix has been written with two goals in mind.

The first goal is to unify two well-known approaches to mild solutions with initial data in the critical homogeneous Besov spaces. Mild solutions in the time-space homogeneous Besov spaces were considered in the blow-up criterion of Gallagher, Koch and Planchon [Gallagher et al. 2016], whereas mild solutions in Kato spaces are better suited to our own needs. In Theorem A.2, we prove that the two notions of solution coincide and have the same maximal time of existence. The third subsection contains the proof, while the first two are devoted to background material on Littlewood–Paley theory and Besov spaces. We also address the subcritical theory in Kato spaces in Theorem A.7. Our hope is that the details provided herein will equip the reader to fill in any fixed-point arguments we have omitted in Sections 2–3.

The second goal is to collect various results related to the theory of suitable weak solutions. While the blow-up arguments we have presented are not overly complicated, they do rely on technical machinery developed by a number of authors starting in the 1970s. We summarize the theory in the fourth subsection and refer to [Escauriaza et al. 2003] for additional details.

Littlewood–Paley theory and homogeneous Besov spaces. We now summarize the basics of Littlewood–Paley theory and homogeneous Besov spaces. Our treatment is based on the presentation in [Bahouri et al. 2011a, Chapter 2]. The situation is as follows. There exist smooth functions φ and χ on $\mathbb{R}^3$ with the properties

$$\operatorname{supp}(\varphi) \subset \left\{\xi \in \mathbb{R}^3 : \tfrac{3}{4} \le |\xi| \le \tfrac{8}{3}\right\}, \qquad \sum_{j\in\mathbb{Z}} \varphi(2^{-j}\xi) = 1, \quad \xi \in \mathbb{R}^3 \setminus \{0\}, \tag{A-1}$$

$$\operatorname{supp}(\chi) \subset \left\{\xi \in \mathbb{R}^3 : |\xi| \le \tfrac{4}{3}\right\}, \qquad \chi(\xi) + \sum_{j\ge 0} \varphi(2^{-j}\xi) = 1, \quad \xi \in \mathbb{R}^3. \tag{A-2}$$

For all $j \in \mathbb{Z}$, we define the homogeneous dyadic block $\dot{\Delta}_j$ and the homogeneous low-frequency cutoff $\dot{S}_j$ to be the following Fourier multipliers:

$$\dot{\Delta}_j := \varphi(2^{-j}D), \quad \dot{S}_j := \chi(2^{-j}D). \tag{A-3}$$

For tempered distributions u_0 on $\mathbb{R}^3$, the convergence of the sum $\sum_{j\le 0} \dot{\Delta}_j u_0$ typically occurs only in the sense of tempered distributions modulo polynomials; see pp. 28–30 in [Lemarié-Rieusset 2002]. To remove ambiguity, we will consider the following subspace of tempered distributions on $\mathbb{R}^3$:

$$\mathcal{S}'_h := \left\{u_0 \in \mathcal{S}' : \lim_{\lambda\to\infty} \|\theta(\lambda D)u_0\|_{L^\infty(\mathbb{R}^3)} = 0 \ \text{ for all } \theta \in \mathcal{D}(\mathbb{R}^3)\right\}. \tag{A-4}$$

The subspace $\mathcal{S}'_h$ is not closed in the standard topology on tempered distributions. We will often refer to the condition defining (A-4) as the "realization condition." Let us recall the family of homogeneous Besov seminorms, defined for all tempered distributions u_0 on $\mathbb{R}^3$:

$$\|u_0\|_{\dot{B}^s_{p,q}(\mathbb{R}^3)} := \left\|2^{js}\|\dot{\Delta}_j u_0\|_{L^p(\mathbb{R}^3)}\right\|_{\ell^q(\mathbb{Z})}, \quad s \in \mathbb{R}, \quad 1 \le p,q \le \infty. \tag{A-5}$$

These are norms when restricted to tempered distributions in the class $u_0 \in \mathcal{S}'_h$. We now introduce the family of homogeneous Besov spaces,

$$\dot{B}^s_{p,q}(\mathbb{R}^3) := \{u_0 \in \mathcal{S}'_h : \|u_0\|_{\dot{B}^s_{p,q}(\mathbb{R}^3)} < \infty\}, \quad s \in \mathbb{R}, \quad 1 \le p, q \le \infty. \tag{A-6}$$

This is a family of normed vector spaces. As long as the condition

$$s < \frac{3}{p} \quad \text{or} \quad s = \frac{3}{p},\ q = 1 \tag{A-7}$$

is satisfied, $\dot{B}^s_{p,q} \cap \dot{B}^{s_1}_{p_1,q_1}(\mathbb{R}^3)$ is a Banach space for all $s_1 \in \mathbb{R}$ and $1 \le p_1, q_1 \le \infty$, and there is no ambiguity modulo polynomials.

Let us now recall a particularly useful characterization of homogeneous Besov spaces in terms of Kato-type norms. Our reference is [Bahouri et al. 2011a, Theorem 2.34]. Let $0 < T \le \infty$. The following family of norms is defined for locally integrable functions $u \in L^1_{\mathrm{loc}}(Q_T)$:

$$\|u\|_{\mathcal{K}^s_{p,q}(Q_T)} := \big\| t^{-\frac{s}{2}} \|u(\cdot,t)\|_{L^p(\mathbb{R}^3)} \big\|_{L^q((0,T),\frac{dt}{t})}, \quad s \in \mathbb{R}, \quad 1 \le p, q \le \infty. \tag{A-8}$$

We now define the Kato spaces with the above norms:

$$\mathcal{K}^s_{p,q}(Q_T) := \{u \in L^1_{\mathrm{loc}}(Q_T) : \|u\|_{\mathcal{K}^s_{p,q}(Q_T)} < \infty\}. \tag{A-9}$$

To simplify notation, we will define

$$\mathcal{K}^s_p(Q_T) := \mathcal{K}^s_{p,\infty}(Q_T), \quad \mathcal{K}_p(Q_T) := \mathcal{K}^{s_p}_p(Q_T). \tag{A-10}$$

The caloric characterization of homogeneous Besov spaces is as follows. For all $s < 0$, there exists a constant $c := c(s) > 0$ such that

$$c^{-1} \|e^{t\Delta} u_0\|_{\mathcal{K}^s_{p,q}(Q_\infty)} \le \|u_0\|_{\dot{B}^s_{p,q}(\mathbb{R}^3)} \le c \|e^{t\Delta} u_0\|_{\mathcal{K}^s_{p,q}(Q_\infty)} \tag{A-11}$$

for all tempered distributions u_0 on $\mathbb{R}^3$.

We now introduce the time-space homogeneous Besov spaces that appear naturally when solving the Navier–Stokes equations with initial data in homogeneous Besov spaces. Our presentation follows [Bahouri et al. 2011a, Section 2.6.3]. Let $0 < T \le \infty$. We have the following family of seminorms on tempered distributions $u \in \mathcal{S}'(Q_T)$:

$$\|u\|_{\widetilde{L}^r_T \dot{B}^s_{p,q}} := \big\| 2^{js} \|\Delta_j u_0\|_{L^r_t L^p_x(Q_T)} \big\|_{\ell^q(\mathbb{Z})}, \quad s \in \mathbb{R}, \quad 1 \le p, q, r \le \infty. \tag{A-12}$$

The time-space homogeneous Besov spaces on Q_T are defined below:

$$\widetilde{L}^r_T \dot{B}^s_{p,q} := \Big\{ u \in \mathcal{S}'(Q_T) : \|u\|_{\widetilde{L}^r_T \dot{B}^s_{p,q}} < \infty, \\ \lim_{j \downarrow -\infty} \|\dot{S}_j u\|_{L^1_t L^\infty_x(Q_{T_1,T_2})} = 0 \ \text{ for all } 0 < T_1 < T_2 < T \Big\}, \\ s \in \mathbb{R}, \quad 1 \le p, q, r \le \infty. \tag{A-13}$$

The second condition in (A-13) is analogous to the realization condition (A-4). We have that $\widetilde{L}^r_T \dot{B}^s_{p,q} \cap \widetilde{L}^{r_1}_T \dot{B}^{s_1}_{p_1,q_1}$ is a Banach space for all $s_1 \in \mathbb{R}$ and $1 \le r_1, p_1, q_1 \le \infty$ provided that (A-7) is satisfied. To simplify notation, we will omit the reference to T in the norm when $T = \infty$. We also sometimes employ spaces $\widetilde{L}^r_{\delta,T}\dot{B}^s_{p,q}$ on the spacetime domain $Q_{\delta,T}$ that are defined in the same way.

We now review the Bony paraproduct decomposition as described in [Bahouri et al. 2011a, Section 2.6]. Consider the operators

$$\dot{T}_u v := \sum_{j\in\mathbb{Z}} \dot{S}_{j-1}u\,\dot{\Delta}_j v, \qquad \dot{R}(u,v) := \sum_{|j-j'|\le 1} \dot{\Delta}_j u\,\dot{\Delta}_{j'} v, \tag{A-14}$$

defined formally for all tempered distributions u, v on $\mathbb{R}^3$. These operators represent low-high and high-high interactions in the formal product

$$uv = \dot{T}_u v + \dot{T}_v u + \dot{R}(u,v). \tag{A-15}$$

If the sums defining the paraproduct operators converge, one may use (A-15) to extend the notion of product to a wider class. Consider

$$\begin{gathered} 1 \le p, p_1, p_2, q, q_1, q_2 \le \infty, \qquad s, s_1, s_2 \in \mathbb{R}, \\ \frac{1}{p} = \frac{1}{p_1} + \frac{1}{p_2}, \qquad \frac{1}{q} = \frac{1}{q_1} + \frac{1}{q_2}, \qquad s = s_1 + s_2. \end{gathered} \tag{A-16}$$

Then one has the estimates

$$\|\dot{T}_u v\|_{\dot{B}^s_{p,q}(\mathbb{R}^3)} \le c(s,s_1)\|u\|_{\dot{B}^{s_1}_{p_1,q_1}(\mathbb{R}^3)}\|v\|_{\dot{B}^{s_2}_{p_2,q_2}(\mathbb{R}^3)} \quad \text{(provided } s_1 < 0\text{)}, \tag{A-17}$$

$$\|\dot{R}(u,v)\|_{\dot{B}^s_{p,q}(\mathbb{R}^3)} \le c(s)\|u\|_{\dot{B}^{s_1}_{p_1,q_1}(\mathbb{R}^3)}\|v\|_{\dot{B}^{s_2}_{p_2,q_2}(\mathbb{R}^3)} \quad \text{(provided } s > 0\text{)}. \tag{A-18}$$

The additional condition (A-7) will imply that $\dot{T}_u v,\ \dot{R}(u,v) \in \mathcal{S}'_h$. The analogous estimates in time-space homogeneous Besov norms are

$$\|\dot{T}_u v\|_{\widetilde{L}^r_T \dot{B}^s_{p,q}} \le c(s,s_1)\|u\|_{\widetilde{L}^{r_1}_T \dot{B}^{s_1}_{p_1,q_1}}\|v\|_{\widetilde{L}^{r_2}_T \dot{B}^{s_2}_{p_2,q_2}} \quad \text{(provided } s_1 < 0\text{)}, \tag{A-19}$$

$$\|\dot{R}(u,v)\|_{\widetilde{L}^r_T \dot{B}^s_{p,q}} \le c(s)\|u\|_{\widetilde{L}^{r_1}_T \dot{B}^{s_1}_{p_1,q_1}}\|v\|_{\widetilde{L}^{r_2}_T \dot{B}^{s_2}_{p_2,q_2}} \quad \text{(provided } s > 0\text{)}, \tag{A-20}$$

where

$$1 \le r, r_1, r_2 \le \infty, \qquad \frac{1}{r} = \frac{1}{r_1} + \frac{1}{r_2}, \qquad 0 < T \le \infty. \tag{A-21}$$

The paraproduct decomposition will play a crucial role in proving estimates on the nonlinear term.

Heat estimates in homogeneous Besov spaces. In this subsection, we recall estimates for the heat equation in time-space homogeneous Besov spaces and Kato spaces.

Regarding heat estimates of frequency-localized data, the primary observation is the following; see [Bahouri et al. 2011a, Section 2.1.2] as well as the appendix of [Gallagher et al. 2016]. Let $\mathcal{C} \subset \mathbb{R}^3$ be an annulus and $\lambda > 0$. There exist constants $C, c > 0$ depending only on the annulus $\mathcal{C}$ such that for all tempered distributions u_0 satisfying $\operatorname{supp}(\hat{u}_0) \subset \lambda\mathcal{C}$,

$$\|e^{t\Delta}u_0\|_{L^p(\mathbb{R}^3)} \le Ce^{-ct\lambda^2}\|u_0\|_{L^p(\mathbb{R}^3)}, \quad 1 \le p \le \infty, \tag{A-22}$$

Let $0 < T \le \infty$ and f be a tempered distribution on Q_T with spatial Fourier transform satisfying $\operatorname{supp} \hat{f} \subset \lambda\mathcal{C} \times [0,T]$. There exists a constant $C > 0$, depending only on the annulus $\mathcal{C}$, such that

$$\|e^{t\Delta}u_0\|_{L_t^{r_2}L_x^{p_2}(Q_\infty)} \le C\lambda^{-\frac{2}{r_2}}\lambda^{3(\frac{1}{p_1}-\frac{1}{p_2})}\|u_0\|_{L^{p_1}(\mathbb{R}^3)}, \tag{A-23}$$

$$\left\|\int_0^t e^{(t-s)\Delta} f(\cdot,s)\,ds\right\|_{L_t^{r_2}L_x^{p_2}(Q_T)} \le C\lambda^{2(\frac{1}{r_1}-\frac{1}{r_2}-1)}\lambda^{3(\frac{1}{p_1}-\frac{1}{p_2})}\|f\|_{L_t^{r_1}L_x^{p_1}(Q_T)}, \tag{A-24}$$

$$1 \le p_1 \le p_2 \le \infty, \quad 1 \le r_1 \le r_2 \le \infty. \tag{A-25}$$

The estimates follow from (A-22) by Bernstein's inequality and Young's convolution inequality. As an application, we obtain

$$\|e^{t\Delta}u_0\|_{\widetilde{L}^{r_2}\dot{B}^{s_{p_2}+2/r_2}_{p_2,q_2}} \le C\|u_0\|_{\dot{B}^{s_{p_1}}_{p_1,q_1}(\mathbb{R}^3)}, \tag{A-26}$$

$$\left\|\int_0^t e^{(t-s)\Delta}\mathbb{P}\operatorname{div} F(\cdot,s)\,ds\right\|_{\widetilde{L}^{r_2}_T\dot{B}^{s_{p_2}+2/r_2}_{p_2,q_2}} \le C\|F\|_{\widetilde{L}^{r_1}_T\dot{B}^{s_{p_1}+2/r_1-1}_{p_1,q_1}}, \tag{A-27}$$

$$1 \le p_1 \le p_2 \le \infty, \quad 1 \le q_1 \le q_2 \le \infty, \quad 1 \le r_1 \le r_2 \le \infty. \tag{A-28}$$

Here, we have employed that $\mathbb{P}$ is a homogeneous Fourier multiplier of degree zero smooth away from the origin; see Proposition 2.40 in [Bahouri et al. 2011a]. Let us now comment on continuity in time. Regarding the estimates (A-26) and (A-27), the solutions belong to the class $C([0,T];\dot{B}^{s_{p_2}}_{p_2,q_2}(\mathbb{R}^3))$ as long as $r_1, q_1 < \infty$ and the realization condition in (A-13) is met. For example, the realization condition is met whenever $(s_{p_2}+2/r_2, p_2, q_2)$ satisfies (A-7). Because the mild solutions we seek exist in spaces $\widetilde{L}^\infty_T\dot{B}^{s_p}_{p,q}$ with $3 < p < \infty$, realization will often be automatic.

The second set of estimates we will discuss are the estimates in Kato spaces that arise naturally from the caloric characterization (A-11) of homogeneous Besov spaces. We summarize them in a single lemma:

Lemma A.1 (estimates in Kato spaces). *Let $0 < T \le \infty$ and $1 \le p_1 \le p_2 \le \infty$ such that*

$$s_2 - \frac{3}{p_2} = 1 + s_1 - \frac{3}{p_1}. \tag{A-29}$$

In addition, assume the conditions

$$s_1 > -2, \quad \frac{3}{p_1} - \frac{3}{p_2} < 1. \tag{A-30}$$

(*For instance, if $p_2 = \infty$, then the latter condition is satisfied when $p_1 > 3$. If $p_1 = 2$, then the latter condition is satisfied when $p_2 < 6$.*) *Then*

$$\left\|\int_0^t e^{(t-\tau)\Delta}\mathbb{P}\operatorname{div} F(\cdot,\tau)\,d\tau\right\|_{\mathcal{K}^{s_2}_{p_2}(Q_T)} \le C(s_1,p_1,p_2)\|F\|_{\mathcal{K}^{s_1}_{p_1}(Q_T)} \tag{A-31}$$

for all distributions $F \in \mathcal{K}^{s_1}_{p_1}(Q_T)$, and the solution u to the corresponding heat equation belongs to $C((0,T];L^{p_2}(\mathbb{R}^3))$. Let $k, l \ge 0$ be integers. If we further require that

$$t^{\alpha+\frac{|\beta|}{2}}\partial_t^\alpha \nabla^\beta F \in \mathcal{K}^{s_1}_{p_1}(Q_T), \tag{A-32}$$

for all integers $0 \le \alpha \le k$ and multi-indices $\beta \in (\mathbb{N}_0)^3$ with $|\beta| \le l$, then we have

$$\left\| t^{k+\frac{l}{2}} \partial_t^k \nabla^l \int_0^t e^{(t-\tau)\Delta} \mathbb{P} \operatorname{div} F(\cdot,\tau)\, d\tau \right\|_{\mathcal{K}^{s_2}_{p_2}(Q_T)} \le C(k,l,s_1,p_1,p_2) \Big(\sum_{\alpha=0}^{k} \sum_{\beta=0}^{l} \| t^{\alpha+\frac{|\beta|}{2}} F \|_{\mathcal{K}^{s_1}_{p_1}(Q_T)} \Big), \tag{A-33}$$

and the spacetime derivatives $\partial_t^k \nabla^l u$ of the solution u belong to $C((0,T]; L^{p_2}(\mathbb{R}^3))$.

Proof. Recall the following estimates on the Oseen kernel [Lemarié-Rieusset 2002, Chapter 11]. Let $\alpha \ge 0$ be an integer and $\beta \in (\mathbb{N}_0)^3$ be a multi-index. Then

$$\| \partial_t^\alpha \nabla^\beta e^{t\Delta} \mathbb{P} \operatorname{div} u_0 \|_{L^{p_2}(\mathbb{R}^3)} \le c(\alpha,\beta) t^{\frac{1}{2}(\frac{3}{p_2} - \frac{3}{p_1} - 1 - |\beta| - 2\alpha)} \| u_0 \|_{L^{p_1}(\mathbb{R}^3)} \tag{A-34}$$

for all $t > 0$ and $1 \le p_1 \le p_2 \le \infty$. In addition, the semigroup commutes with partial derivatives in the space variables.

Let us consider the case when α, β are zero. Suppose that s_1, s_2, p_1, p_2, and F obey the hypotheses of the lemma. Then

$$\begin{aligned}
&\left\| \int_0^t e^{(t-\tau)\Delta} \mathbb{P} \operatorname{div} F(\cdot,\tau)\, d\tau \right\|_{L^{p_2}(\mathbb{R}^3)} \\
&\qquad \le \int_0^t \| e^{(t-\tau)\Delta} \mathbb{P} \operatorname{div} F(\cdot,\tau) \|_{L^{p_2}(\mathbb{R}^3)}\, d\tau \\
&\qquad \overset{\text{(A-34)}}{\le} c \int_0^t (t-\tau)^{\frac{1}{2}(\frac{3}{p_2} - \frac{3}{p_1} - 1)} \| F(\cdot,\tau) \|_{L^{p_1}(\mathbb{R}^3)}\, d\tau \\
&\qquad \le c \int_0^t (t-\tau)^{\frac{1}{2}(\frac{3}{p_2} - \frac{3}{p_1} - 1)} \tau^{\frac{s_1}{2}}\, d\tau \times \sup_{0<\tau<T} \tau^{-\frac{s_1}{2}} \| F(\cdot,\tau) \|_{L^{p_1}(\mathbb{R}^3)} \\
&\qquad \overset{\text{(A-30)}}{\le} c \Big[\Big(\frac{s_1}{2} + 1 \Big)^{-1} - 2 \Big(\frac{3}{p_2} - \frac{3}{p_1} + 1 \Big)^{-1} \Big] t^{\frac{1}{2}(\frac{3}{p_2} - \frac{3}{p_1} + s_1 + 1)} \times \| F \|_{\mathcal{K}^{s_1}_{p}(Q_T)} \\
&\qquad \overset{\text{(A-29)}}{\le} c \Big[\Big(\frac{s_1}{2} + 1 \Big)^{-1} - 2 \Big(\frac{3}{p_2} - \frac{3}{p_1} + 1 \Big)^{-1} \Big] t^{\frac{s_2}{2}} \times \| F \|_{\mathcal{K}^{s_1}_{p}(Q_T)}.
\end{aligned} \tag{A-35}$$

This completes the proof of the first estimate. Now let us define

$$u(\cdot,t) := \int_0^t e^{(t-s)} \mathbb{P} \operatorname{div} F(s)\, ds \tag{A-36}$$

for all $0 < t \le T$ and observe the identity

$$u(\cdot,t) = e^{(t-s)\Delta} u(\cdot,\tau) + \int_s^t e^{(t-\tau)\Delta} \mathbb{P} \operatorname{div} F(\cdot,\tau)\, d\tau \tag{A-37}$$

for all $0 < s < t$. To prove that $u \in C((0,T]; L^{p_2}(\mathbb{R}^3))$, one merely estimates

$$\begin{aligned}
\| u(\cdot,t) - u(\cdot,s) \|_{L^{p_2}(\mathbb{R}^3)} &\le \| e^{(t-s)\Delta} u(\cdot,s) - u(\cdot,s) \|_{L^{p_2}(\mathbb{R}^3)} + \int_s^t \| e^{(t-\tau)\Delta} \mathbb{P} \operatorname{div} F(\cdot,\tau) \|_{L^{p_2}(\mathbb{R}^3)}\, d\tau \\
&\le o(1) + c \int_s^t (t-\tau)^{\frac{1}{2}(\frac{3}{p_2} - \frac{3}{p_1} - 1)} \tau^{\frac{s_1}{2}}\, d\tau \times \| F \|_{\mathcal{K}^{s_1}_{p_1}(Q_T)} = o(1)
\end{aligned} \tag{A-38}$$

as $|t-s| \to 0$, according to the assumption (A-30) on the exponents.

Let us now demonstrate how to prove the estimates on spatial derivatives. One estimates the integral in two parts,

$$\begin{aligned}
&\Big\| \int_0^t \nabla^l e^{(t-\tau)\Delta} \mathbb{P}\,\mathrm{div}\, F(\cdot,\tau)\,d\tau \Big\|_{L^{p_2}(\mathbb{R}^3)} \\
&\leq c(l) \int_0^{\frac{t}{2}} (t-\tau)^{\frac{1}{2}(\frac{3}{p_2}-\frac{3}{p_1}-1-l)} \|F(\cdot,\tau)\|_{L^{p_1}(\mathbb{R}^3)}\,d\tau + c\int_{\frac{t}{2}}^t (t-\tau)^{\frac{1}{2}(\frac{3}{p_2}-\frac{3}{p_1}-1)} \|\nabla^l F(\cdot,\tau)\|_{L^{p_1}(\mathbb{R}^3)}\,d\tau \\
&\leq c(l) \int_0^{\frac{t}{2}} (t-\tau)^{\frac{1}{2}(\frac{3}{p_2}-\frac{3}{p_1}-1-l)} \tau^{\frac{s_1}{2}}\,d\tau \times \|F\|_{\mathcal{K}^{s_1}_{p_1}(Q_T)} \\
&\qquad\qquad + c\int_{\frac{t}{2}}^t (t-\tau)^{\frac{1}{2}(\frac{3}{p_2}-\frac{3}{p_1}-1)} \tau^{\frac{1}{2}(s_1-l)}\,d\tau \times \|\tau^{\frac{l}{2}} \nabla^l F(\cdot,\tau)\|_{\mathcal{K}^{s_1}_{p_1}(Q_T)} \\
&\leq c(l,s_1,p_1,p_2) t^{\frac{1}{2}(s_2-l)} (\|F\|_{\mathcal{K}^{s_1}_{p_1}(Q_T)} + \|\tau^{\frac{l}{2}} \nabla^l F(\cdot,\tau)\|_{\mathcal{K}^{s_1}_{p_1}(Q_T)}).
\end{aligned} \tag{A-39}$$

The proof of continuity in $L^{p_2}(\mathbb{R}^3)$ is similar to (A-38) except with spatial derivatives in the identity (A-37).

The proof of estimates on the temporal derivatives is slightly more cumbersome due to the weighted spaces under consideration and that the temporal derivatives do not preserve the form of the equation. By differentiating the identity (A-37) in time, one obtains

$$\partial_t u(\cdot,t) = \partial_t e^{(t-s)\Delta} u(\cdot,s) + e^{(t-s)\Delta}\mathbb{P}\,\mathrm{div}\, F(\cdot,s) + \int_s^t e^{(t-\tau)\Delta}\mathbb{P}\,\mathrm{div}\,\partial_\tau F(\cdot,\tau)\,d\tau, \tag{A-40}$$

and more generally,

$$\partial_t^k u(\cdot,t) = \partial_t^k e^{(t-s)\Delta} u(\cdot,\tau) + \sum_{\alpha=1}^k \partial_t^{k-\alpha} e^{(t-s)\Delta}\mathbb{P}\,\mathrm{div}\,\partial_s^{\alpha-1} F(\cdot,s) + \int_s^t e^{(t-\tau)\Delta}\mathbb{P}\,\mathrm{div}\,\partial_\tau^k F(\cdot,\tau)\,d\tau. \tag{A-41}$$

(In obtaining the identities, it is beneficial to compare with the differential form of the equation.) Now set $s := \frac{1}{2}t$ and denote the terms by I, II, and III, respectively. We estimate

$$\begin{aligned}
\|I\|_{L^{p_2}} &\leq c(k) t^{-k} \big\|u\big(\cdot,\tfrac{1}{2}t\big)\big\|_{L^{p_2}} \\
&\leq c(k,p_2) t^{-k+\frac{s_2}{2}} \|u\|_{\mathcal{K}^{s_2}_{p_2}(Q_T)} \\
&\leq c(k,s_1,p_1,p_2) t^{-k+\frac{s_2}{2}} \|F\|_{\mathcal{K}^{s_1}_{p_1}(Q_T)},
\end{aligned} \tag{A-42}$$

according to our original estimate. Furthermore,

$$\begin{aligned}
\|II\|_{L^{p_2}} &\leq c(k) \sum_{\alpha=1}^k t^{\alpha-k+\frac{1}{2}(\frac{3}{p_2}-\frac{3}{p_1}-1)} \big\|(\partial_t^{\alpha-1} F)\big(\tfrac{1}{2}t\big)\big\|_{L^{p_1}} \\
&\leq c(k,s_1,p_1,p_2) t^{-k+\frac{s_2}{2}} \sum_{\alpha=1}^k \|\tau^{\alpha-1} F\|_{\mathcal{K}^{s_1}_{p_1}(Q_T)},
\end{aligned} \tag{A-43}$$

and finally,

$$\begin{aligned}\|III\|_{L^{p_2}} &\leq c\int_{\frac{t}{2}}^{t}(t-\tau)^{\frac{1}{2}(\frac{3}{p_2}-\frac{3}{p_1}-1)}\|\partial_\tau^k F(\cdot,\tau)\|_{L^{p_1}}\,ds\\ &\leq c\int_{\frac{t}{2}}^{t}(t-\tau)^{\frac{1}{2}(\frac{3}{p_2}-\frac{3}{p_1}-1)}\tau^{-k+\frac{s_1}{2}}\,ds\times\|\tau^k\partial_\tau^k F\|_{\mathcal{K}^{s_1}_{p_2}(Q_T)}\\ &\leq c(k,s_1,p_1,p_2)t^{-k+\frac{s_2}{2}}\|\tau^k\partial_\tau^k F\|_{\mathcal{K}^{s_1}_{p_2}(Q_T)}.\end{aligned}\tag{A-44}$$

This completes the proof of the time-derivative estimates. The proof of continuity is similar to (A-38) except that one must use the identity (A-41).

Regularity in spacetime may be obtained by applying the temporal estimates to the spatial derivatives, since the spatial derivatives preserve the form of the equation. □

Mild solutions in homogeneous Besov spaces. The goal of this subsection is to review the well-posedness theory of the Navier–Stokes equations with initial data in Besov spaces.

Let us recall the notion of a mild solution to the Navier–Stokes equations, i.e., a tempered distribution u on spacetime that satisfies the integral equation

$$u(\cdot,t)=e^{t\Delta}u_0-\int_0^t e^{(t-s)\Delta}\mathbb{P}\operatorname{div}u\otimes u\,ds \tag{A-45}$$

in a suitable function space. The operator $e^{t\Delta}\mathbb{P}\operatorname{div}$ is defined by convolution with the gradient of the Oseen kernel; see Chapter 11 of [Lemarié-Rieusset 2002]. We will often simply write

$$u(\cdot,t)=e^{t\Delta}u_0-B(u,u)(\cdot,t), \tag{A-46}$$

$$B(v,w)(\cdot,t):=\int_0^t e^{(t-s)\Delta}\mathbb{P}\operatorname{div}v\otimes w\,ds. \tag{A-47}$$

Distributional solutions to the Navier–Stokes equations are mild under rather general hypotheses, as discussed in Chapter 14 of [Lemarié-Rieusset 2002]. Small-data-global-existence in the spirit of the seminal work [Kato 1984] is known for divergence-free initial data in the spaces

$$\dot{H}^{\frac{1}{2}}(\mathbb{R}^3)\hookrightarrow L^3(\mathbb{R}^3)\hookrightarrow \dot{B}^{s_{p_1}}_{p_1,q_1}(\mathbb{R}^3)\hookrightarrow B^{s_{p_2}}_{p_2,q_2}(\mathbb{R}^3)\hookrightarrow \mathrm{BMO}^{-1}(\mathbb{R}^3), \tag{A-48}$$

where $3<p_1\leq p_2<\infty$ and $3<q_1\leq q_2\leq\infty$. The case $\mathrm{BMO}^{-1}(\mathbb{R}^3)$ was treated in [Koch and Tataru 2001] and appears to be optimal. Ill-posedness has been demonstrated in [Bourgain and Pavlović 2008] in the maximal critical space $\dot{B}^{-1}_{\infty,\infty}(\mathbb{R}^3)$. Local-in-time existence is known for initial data in the spaces

$$\dot{H}^{\frac{1}{2}}(\mathbb{R}^3)\hookrightarrow L^3(\mathbb{R}^3)\hookrightarrow \dot{B}^{s_p}_{p,q}(\mathbb{R}^3)\hookrightarrow \mathrm{VMO}^{-1}(\mathbb{R}^3) \tag{A-49}$$

as long as $3<p,q<\infty$, where

$$\mathrm{VMO}^{-1}(\mathbb{R}^3):=\overline{\mathcal{S}(\mathbb{R}^3)}^{\mathrm{BMO}^{-1}(\mathbb{R}^3)}. \tag{A-50}$$

The existence theory for initial data in homogeneous Besov spaces is summarized in the following theorem.

Theorem A.2 (mild solutions in critical Besov spaces). *Let $3 < p, q < \infty$ and $u_0 \in \dot{B}^{s_p}_{p,q}(\mathbb{R}^3)$ be a divergence-free vector field. Then there exists a time $0 < T^*(u_0) \le \infty$ and a mild solution u of the Navier–Stokes equations such that*

$$u \in C([0,T]; \dot{B}^{s_p}_{p,q}(\mathbb{R}^3)) \cap \widetilde{L}^1_T \dot{B}^{s_p+2}_{p,q} \cap \widetilde{L}^\infty_T \dot{B}^{s_p}_{p,q}, \tag{A-51}$$

$$u \in \mathring{\mathcal{K}}_p(Q_T) \cap \mathring{\mathcal{K}}_\infty(Q_T) \cap C((0,T]; L^p \cap L^\infty(\mathbb{R}^3)) \tag{A-52}$$

for all $0 < T < T^*(u_0)$. In addition, u is the unique mild solution satisfying (A-51) and the unique mild solution satisfying (A-52). Lastly, if $T^*(u_0) < \infty$, then the following two conditions are satisfied:

(i)
$$\lim_{t \uparrow T^*} \|u(\cdot,t)\|_{L^{p_0}(\mathbb{R}^3)} = \infty. \tag{A-53}$$

(ii) *For all $p_0 \in [p,\infty)$, $q_0 \in [q,\infty)$, and $1 \le r_0 \le \infty$ such that $s_{p_0} + 2/r_0 \in (0, 3/p_0)$,*

$$\lim_{T \uparrow T^*} \|u\|_{\widetilde{L}^{r_0}_T \dot{B}^{s_{p_0}+2/r_0}_{p_0,q_0}} = \infty. \tag{A-54}$$

In the statement, we have utilized the Banach space

$$\mathring{\mathcal{K}}^s_p(Q_T) := \{u \in \mathcal{K}^s_p(Q_T) : t^{-\frac{s}{2}} \|u(\cdot,t)\|_{L^p(\mathbb{R}^3)} \to 0 \text{ as } t \downarrow 0\}. \tag{A-55}$$

We will discuss the proof after reviewing two lemmas concerning quadratic equations in Banach spaces based on Lemmas A.1 and A.2 in [Gallagher et al. 2002]. See also Lemma 5 in [Auscher et al. 2004].

Lemma A.3 (abstract Picard lemma). *Let X be a Banach space, $L : X \to X$ a bounded linear operator such that $I - L : X \to X$ is invertible, and B a continuous bilinear operator on X satisfying*

$$\|B(x,y)\|_X \le \gamma \|x\|_X \|y\|_X \tag{A-56}$$

for some $\gamma > 0$ and all $x, y \in X$. Then for all $a \in X$ satisfying

$$\|(I-L)^{-1}a\|_X < \frac{1}{4\|(I-L)^{-1}\|_X \gamma}, \tag{A-57}$$

the Picard iterates $P_k(a)$, defined recursively by

$$P_0(a) := a, \qquad P_{k+1}(a) := a + L(P_k) + B(P_k, P_k), \quad k \ge 0, \tag{A-58}$$

converge in X to the unique solution $x \in X$ of the equation

$$x = a + L(x) + B(x,x) \tag{A-59}$$

such that

$$\|x\|_X < \frac{1}{2\|(I-L)^{-1}\|_X \gamma}. \tag{A-60}$$

Regarding the hypothesis on L, the operator $I - L : X \to X$ is invertible with norm

$$\|(I-L)^{-1}\|_X \le \frac{1}{1 - \|L\|_X} \tag{A-61}$$

whenever $\|L\|_X < 1$. We use this fact in Lemma 2.4 and Theorem 2.6.

Often one applies Lemma A.3 to an intersection of spaces. For instance, to solve the Navier–Stokes equations with divergence-free initial data $u_0 \in L^2 \cap L^3(\mathbb{R}^3)$, the space X may be chosen as $X := L^5_{t,x} \cap L^\infty_t L^2_x \cap L^2_t \dot{H}^1_x(Q_T)$. Similarly, one may choose X to include higher derivatives in order to prove higher regularity. Technically, when one includes more derivatives in the space X, one may need to shorten the time interval on which Lemma A.3 is applied and argue that the additional regularity is propagated forward in time. This is cleverly avoided in [Germain et al. 2007].

Lemma A.4 (propagation of regularity). *In the notation of Lemma A.3, let $E \hookrightarrow X$ be a Banach space. Suppose that L is bounded on E such that $I - L : E \to E$ is invertible and B maps $E \times X \to E$ and $X \times E \to E$ with*

$$\max(\|B(y,z)\|_E, \|B(z,y)\|_E) \leq \eta \|y\|_E \, \|z\|_X \tag{A-62}$$

for some $\eta > 0$ and all $y \in E$, $z \in X$. Finally, suppose that

$$\|(I-L)^{-1}\|_E \eta \leq \|(I-L)^{-1}\|_X \gamma. \tag{A-63}$$

For all $a \in E$ satisfying (A-57), the solution x from Lemma A.3 belongs to E and satisfies

$$\|x\|_E \leq 2\|(I-L)^{-1} a\|_E. \tag{A-64}$$

Lemma A.4 does not require the quantity $\|a\|_E$ to be small, but one may have to increase $\gamma > 0$ in order to meet the condition (A-63).

Proof of Theorem A.2. Let us assume the hypotheses of Theorem A.2. Let $r > 2$ such that $s_p + \frac{2}{r} \in \left(0, \frac{3}{p}\right)$.

(1) *Constructing a local-in-time mild solution.* To obtain local-in-time solutions, we will apply Lemma A.3 to the integral formulation (A-45) of the Navier–Stokes equations in the Banach space

$$X_T := \widetilde{L}^r_T \dot{B}^{s_p + \frac{2}{r}}_{p,q} \cap \mathring{\mathcal{K}}_p(Q_T). \tag{A-65}$$

Note that the realization condition in (A-13) is satisfied, so the time-space homogeneous Besov space in (A-65) is complete.

Let us prove that the bilinear operator B is bounded on X_T. In fact, it is bounded separately on the two spaces in the intersection. To prove that B is bounded on $\mathcal{K}_p(Q_T)$, we use Hölder's inequality,

$$\|u \otimes v\|_{\mathcal{K}^{2s_p}_{p/2}(Q_T)} \leq \|u\|_{\mathcal{K}_p(Q_T)} \|v\|_{\mathcal{K}_p(Q_T)}, \tag{A-66}$$

and conclude with the heat estimates in Lemma A.1. That the subspace $\mathring{\mathcal{K}}_p(Q_T)$ is stabilized follows from taking the limit $T \downarrow 0$. To prove boundedness on $\widetilde{L}^r_T \dot{B}^{s_p+2/r}_{p,q}$, we use the Bony decomposition (A-15). First, apply the low-high paraproduct estimate (A-19) and Sobolev embedding to obtain

$$\|\dot{T}_u v\|_{\widetilde{L}^{r/2}_T \dot{B}^{-1+s_p+4/r}_{p,q/2}} \leq c\|u\|_{\widetilde{L}^r_T \dot{B}^{-1+2/r}_{\infty,q}} \|v\|_{\widetilde{L}^r_T \dot{B}^{s_p+2/r}_{p,q}} \leq c\|u\|_{\widetilde{L}^r_T \dot{B}^{s_p+2/r}_{p,q}} \|v\|_{\widetilde{L}^r_T \dot{B}^{s_p+2/r}_{p,q}}. \tag{A-67}$$

The analogous estimate is valid for $\dot{T}_v u$. Now we combine the heat estimate (A-27), substituting $F := \dot{T}_u v + \dot{T}_v u$, with the low-high paraproduct estimate:

$$\left\| \int_0^t e^{(t-s)\Delta} \mathbb{P} \operatorname{div}(\dot{T}_u v + \dot{T}_v u)\, ds \right\|_{\widetilde{L}^r_T \dot{B}^{s_p+2/r}_{p,q}} \leq c\|u\|_{\widetilde{L}^r_T \dot{B}^{s_p+2/r}_{p,q}} \|v\|_{\widetilde{L}^r_T \dot{B}^{s_p+2/r}_{p,q}}. \tag{A-68}$$

To estimate the high-high contribution, we apply the property (A-20) to obtain

$$\|\dot{R}(u,v)\|_{\widetilde{L}^{r/2}_T \dot{B}^{2(s_p+2/r)}_{p/2,q/2}} \le c\|u\|_{\widetilde{L}^r_T \dot{B}^{s_p+2/r}_{p,q}} \|v\|_{\widetilde{L}^r_T \dot{B}^{s_p+2/r}_{p,q}}. \tag{A-69}$$

Notice that $2\big(s_p+\frac{2}{r}\big) \in \big(0, \frac{6}{p}\big)$, so the realization condition in (A-13) is satisfied. We then apply the heat estimate (A-27) with $F := \dot{R}(u,v)$ to obtain

$$\left\| \int_0^t e^{(t-s)\Delta} \mathbb{P} \operatorname{div} \dot{R}(u,v)\, ds \right\|_{\widetilde{L}^r_T \dot{B}^{s_p+2/r}_{p,q}} \le c\|u\|_{\widetilde{L}^r_T \dot{B}^{s_p+2/r}_{p,q}} \|v\|_{\widetilde{L}^r_T \dot{B}^{s_p+2/r}_{p,q}}. \tag{A-70}$$

This completes the proof of boundedness in $\widetilde{L}^r_T \dot{B}^{s_p+2/r}_{p,q}$.

The last step in applying Lemma A.3 is to obtain the smallness condition (A-57). Since $u_0 \in \dot{B}^{s_p}_{p,q}(\mathbb{R}^3)$, we have from (A-11) and (A-26) that

$$e^{t\Delta}u_0 \in X_\infty, \qquad \|e^{t\Delta}u_0\|_{X_T} \to 0 \quad \text{as } T \downarrow 0. \tag{A-71}$$

Hence, (A-57) will be satisfied as long as $0 < T \ll 1$ depending on u_0. This completes Step 1.

(2) *Further properties of mild solutions*. Our next goal is to prove that a given mild solution, $u \in X_T$ belongs to the full range of function spaces stated in the theorem. It is clear from (A-11) and (A-26) that $e^{t\Delta}u_0$ is in the desired spaces, so we will focus on the mapping properties of the nonlinear term.

First, suppose u is a mild solution in $\mathring{\mathcal{K}}_p(Q_T)$. The integral equation (A-45) combined with the estimates in Lemma A.1 allow one to bootstrap u into the space $\mathring{\mathcal{K}}_\infty(Q_T)$. The same estimates in Lemma A.1 also give $u \in C((0,T]; L^p \cap L^\infty(\mathbb{R}^3))$. We conclude that u belongs to (A-52).

Now suppose u is a mild solution in $\widetilde{L}^r_T \dot{B}^{s_p+2/r}_{p,q}$. We may combine the paraproduct estimates (A-67) and (A-69) in Step 1 with the heat estimates (A-27) to obtain the mapping property

$$B : \widetilde{L}^r_T \dot{B}^{s_p+2/r}_{p,q} \times \widetilde{L}^r_T \dot{B}^{s_p+2/r}_{p,q} \to C([0,T]; \dot{B}^{s_p}_{p,q}(\mathbb{R}^3)) \cap \widetilde{L}^\infty_T \dot{B}^{s_p}_{p,q}. \tag{A-72}$$

Demonstrating $u \in \widetilde{L}^1_T \dot{B}^{s_p+2}_{p,q}$ requires a bootstrapping argument that we borrow from Remark A.2 of [Gallagher et al. 2002]. Specifically, suppose that $u \in \widetilde{L}^{r_0}_T \dot{B}^{s_p+2/r_0}_{p,q}$ for some $r_0 > 1$ such that $s_p + 2/r_0 > 0$. Let us define the exponents

$$\frac{1}{l(\varepsilon)} := \frac{1-\varepsilon}{2} - \frac{3}{2p}, \qquad \frac{1}{r_1(\varepsilon)} := \frac{1}{r_0} + \frac{1}{l(\varepsilon)}, \qquad 0 \le \varepsilon < s_p + \frac{2}{r_0}. \tag{A-73}$$

From interpolation, it is clear that $u \in \widetilde{L}^{l(\varepsilon)}_T \dot{B}^{s_p+2/l(\varepsilon)}_{p,q}$. Now consider the additional restrictions $\varepsilon > 0$ and $r_1(\varepsilon) \ge 1$. One may verify that

$$s_p + \frac{2}{l(\varepsilon)} = -\varepsilon, \quad s_1(\varepsilon) := 2s_p + \frac{2}{r_0} + \frac{2}{l(\varepsilon)} = s_p + \frac{2}{r_0} - \varepsilon > 0. \tag{A-74}$$

According to the paraproduct laws (A-19) and (A-20), we have

$$\|\dot{T}_u u\|_{\widetilde{L}^{r_1(\varepsilon)}_T \dot{B}^{s_1(\varepsilon)}_{p/2,q/2}} + \|\dot{R}(u,u)\|_{\widetilde{L}^{r_1(\varepsilon)}_T \dot{B}^{s_1(\varepsilon)}_{p/2,q/2}} \le c\|u\|_{\widetilde{L}^{l(\varepsilon)}_T \dot{B}^{-\varepsilon}_{p,q}} \|u\|_{\widetilde{L}^{r_0}_T \dot{B}^{s_p+2/r_0}_{p,q}}. \tag{A-75}$$

We then apply the heat estimate (A-27) to obtain

$$B(u,u) \in \widetilde{L}^{r_1(\varepsilon)}_T \dot{B}^{s_p+\frac{2}{r_1(\varepsilon)}}_{p,q}, \tag{A-76}$$

provided that $0 < \varepsilon < s_p + 2/r_0$ and $r_1(\varepsilon) \geq 1$. In fact, by repeating the arguments with a new value $1 < \tilde{r}_0 < r_0$, one may treat the case $\varepsilon = 0$. The final result is that

$$B(u,u) \in \widetilde{L}_T^{\max(1,r_1)} \dot{B}_{p,q}^{s_p + \min(2, \frac{2}{r_1})}, \quad \frac{1}{r_0} - \frac{s_p}{2} = \frac{1}{r_1}. \tag{A-77}$$

Hence, one may improve $1/r_0$ by the fixed amount $-s_p/2$ at each iteration. This completes the bootstrapping argument, so u belongs to the class (A-51).

(3) *Uniqueness*. Suppose that u_1, u_2 are two mild solutions on Q_T in the class (A-51) with the same initial data u_0. For contradiction, assume there exists a time $0 < T_0 < T$ such that $u_1 \equiv u_2$ on Q_{T_0} but u_1 and u_2 are not identical on $Q_{T_0,T_0+\delta}$ for all $0 < \delta < T - T_0$. Because the solutions are continuous with values in $\dot{B}_{p,q}^{s_p}(\mathbb{R}^3)$, we must have $\tilde{u}_0 := u_1(\cdot, T_0) = u_2(\cdot, T_0)$. When $0 < \delta \ll 1$, depending on u_1 and u_2, the two solutions on $Q_{T_0,T_0+\delta}$ with the same initial data $\tilde{u}_0$ fall into the perturbative regime of Lemma A.3 in the critical time-space homogeneous Besov spaces. Hence, the solutions coincide on $Q_{T_0,T_0+\delta}$, which is a contradiction.

Similarly, suppose that u_1, u_2 are two mild solutions on Q_T in the class (A-52). For $0 < \delta \ll 1$ depending on u_1 and u_2, the solutions on Q_δ fall into the perturbative regime of Lemma A.3 in the Kato space $\mathring{\mathcal{K}}_p(Q_\delta)$. Hence, the solutions coincide on a small time interval. Uniqueness may be propagated forward in time according to the subcritical theory in $L^p(\mathbb{R}^3)$.

(4) *Characterizing the maximal time of existence*. We now return to the mild solution u that we constructed in Step 1. The solution may be continued according to the subcritical theory in $L^p(\mathbb{R}^3)$ and the critical theory in time-space homogeneous Besov spaces. The result is the following. There exists a time $0 < T^*(u_0) \leq \infty$ such that for all $0 < T < T^*(u_0)$, u is the unique mild solution in X_T with initial data u_0, and for all $T > T^*(u_0)$, the solution u cannot be extended in X_T. The time $T^*(u_0)$ is the *maximal time of existence* of the mild solution u with initial data u_0. In the proof, we will denote it simply by T^*.

Suppose that $T^* < \infty$. Then $\lim_{T \uparrow T^*} \|u\|_{X_T} = \infty$, since otherwise the solution can be continued past T^*. A priori, we know that

$$\text{(i}'\text{)} \lim_{t \uparrow T^*} \|u(\cdot, t)\|_{L^p(\mathbb{R}^3)} = \infty \quad \text{or} \quad \text{(ii}'\text{)} \lim_{T \uparrow T^*} \|u\|_{\widetilde{L}_T^r \dot{B}_{p,q}^{s_p + 2/r}} = \infty. \tag{A-78}$$

Note that we may avoid writing $\limsup_{t \uparrow T^*} \|u(\cdot, t)\|_{L^p(\mathbb{R}^3)}$ in (i′) because the $L^p(\mathbb{R}^3)$ norm is subcritical. It is immediate that (i) implies (i′) and (ii) implies (ii′). We will now demonstrate the implications in the reverse direction.

Suppose that (i) does not hold. In other words, there exists $p \leq p_0 \leq \infty$ such that $u \in \mathcal{K}_{p_0}(Q_{T^*})$. By the bootstrapping in Kato spaces mentioned in Step 2, we obtain that

$$u \in \mathcal{K}_\infty(Q_{T^*}) \cap C((0, T^*]; L^\infty(\mathbb{R}^3)). \tag{A-79}$$

Then, we may apply Lemma A.4 concerning propagation of regularity in the spaces $X = L^\infty(Q_T)$ and $E = X \cap L_t^\infty L_x^p(Q_T)$ with initial data $u(\cdot, t_0) \in L^p \cap L^\infty(\mathbb{R}^3)$ and initial time t_0 close to T^*. The lemma prevents $\|u(\cdot, t)\|_{L^p(\mathbb{R}^3)}$ from blowing up as $t \uparrow T^*$, so (i′) does not hold. This is the same argument as in the last step of Proposition A.5.

Now suppose that (ii) does not hold. Then there exists $p \le p_0 < \infty$, $q \le q_0 < \infty$, and $r_0 > 2$ such that $s_{p_0} + 2/r_0 \in (0, 3/p_0)$ and $u \in \widetilde{L}^{r_0}_{T^*} \dot{B}^{s_{p_0}+2/r_0}_{p_0,q_0}$. By the arguments in Step 2, we must have that

$$u \in C((0,T^*]; \dot{B}^{s_{p_0}}_{p_0,q_0}) \cap \widetilde{L}^1_{T^*} \dot{B}^{s_{p_0}+2}_{p_0,q_0} \cap \widetilde{L}^\infty_{T^*} \dot{B}^{s_{p_0}}_{p_0,q_0}. \tag{A-80}$$

As in the previous paragraph, we may apply Lemma A.4 in the spaces $X = \widetilde{L}^{r_0}_T \dot{B}^{s_{p_0}+2/r_0}_{p_0,q_0}$, $E = X \cap \widetilde{L}^r_T \dot{B}^{s_p+2/r}_{p,q}$, with initial data $u(\cdot,t_0) \in \dot{B}^{s_p}_{p,q} \cap \dot{B}^{s_{p_0}}_{p_0,q_0}(\mathbb{R}^3)$ and initial time t_0 close to T^*. This proves that the norm of u in $\widetilde{L}^r_{T^*} \dot{B}^{s_p+2/r}_{p,q}$ stays bounded, so (ii′) fails. Here it is crucial that $u(\cdot,t)$ is continuous on $[0,T^*]$ with values in $\dot{B}^{s_{p_0}}_{p_0,q_0}(\mathbb{R}^3)$, so that the existence time of the solution with initial data $u(\cdot,t_0)$ has a uniform lower bound. This was not an issue in the subcritical setting. Also, we do not record here the bilinear estimates necessary to apply Lemma A.4. They are similar to the estimates in Step 1.

(5) *More characterizing*. It remains to prove that (i′) is equivalent to (ii′). To begin, we will show that (ii) implies (i′) by arguing the contrapositive. Suppose

$$\sup_{\frac14 T^* \le t \le T^*} \|u(\cdot,t)\|_{L^p(\mathbb{R}^3)} < \infty. \tag{A-81}$$

By the subcritical theory in $L^p(\mathbb{R}^3)$, it is not difficult to show that

$$\sup_{\frac12 T^* \le t \le T^*} \|\nabla u(\cdot,t)\|_{L^p(\mathbb{R}^3)} < \infty. \tag{A-82}$$

One may interpolate between (A-81) and (A-82) to obtain for all $0 < s < 1$,

$$\sup_{\frac12 T^* \le t \le T^*} \|u(\cdot,t)\|_{\dot{B}^s_{p,1}(\mathbb{R}^3)} \le c(s) \sup_{\frac12 T^* \le t \le T^*} \|u(\cdot,t)\|^s_{\dot{B}^1_{p,\infty}(\mathbb{R}^3)} \|u(\cdot,t)\|^{1-s}_{\dot{B}^0_{p,\infty}(\mathbb{R}^3)} < \infty; \tag{A-83}$$

see Proposition 2.22 in [Bahouri et al. 2011a]. Now, let $m := \max(q,r)$. By Minkowski's and Hölder's inequalities,

$$\|u\|_{\widetilde{L}^r_{T^*/2,T^*} \dot{B}^{s_p+2/r}_{p,m}} \le \|u\|_{L^r_t(\dot{B}^{s_p+2/r}_{p,m})_x(Q_{T^*/2,T^*})} \le c\,(T^*)^{\frac1r} \sup_{\frac12 T^* \le t \le T^*} \|u(\cdot,t)\|_{\dot{B}^{s_p+2/r}_{p,1}(\mathbb{R}^3)} < \infty, \tag{A-84}$$

and one concludes that (ii) fails.

Now we will demonstrate that (i′) implies (ii′), again by arguing the contrapositive. Let us assume that $\|u\|_{\widetilde{L}^r_{T^*} \dot{B}^{s_p+2/r}_{p,q}} < \infty$, so that by the mapping properties of the nonlinear term in Step 2,

$$u \in C([0,T^*]; \dot{B}^{s_p}_{p,q}(\mathbb{R}^3)). \tag{A-85}$$

Our goal is to prove that the following quantity is bounded:

$$\|u\|_{\widetilde{L}^\infty_{T_0,T^*} \dot{B}^{3/p}_{p,q}(\mathbb{R}^3)} < \infty \tag{A-86}$$

for some $0 < T_0 < T^*$. Indeed, interpolating between (A-85) and (A-86), one may demonstrate that

$$\sup_{T_0 \le t \le T^*} \|u(\cdot,t)\|_{\dot{B}^0_{p,1}(\mathbb{R}^3)} \le c \sup_{T_0 \le t \le T^*} \|u(\cdot,t)\|^{s_p+1}_{\dot{B}^{s_p}_{p,\infty}(\mathbb{R}^3)} \|u(\cdot,t)\|^{-s_p}_{\dot{B}^{s_p+1}_{p,\infty}(\mathbb{R}^3)} < \infty. \tag{A-87}$$

In light of the embedding $\dot{B}^0_{p,1}(\mathbb{R}^3) \hookrightarrow L^p(\mathbb{R}^3)$, this will complete the proof that (i′) fails.

To prove (A-86), we will need more estimates for the heat equation. Let $\mathcal{C} \subset \mathbb{R}^3$ be an annulus and $\lambda > 0$. Let $0 < T \leq \infty$ and f be a tempered distribution on Q_T with spatial Fourier transform satisfying $\operatorname{supp} \hat{f} \subset \lambda\mathcal{C} \times [0, T]$. Then

$$\lambda \left\| t^{\frac{1}{2}} \int_0^t e^{(t-\tau)\Delta} f(\cdot,\tau)\,d\tau \right\|_{L_t^{r_2} L_x^{p_2}(Q_T)} \leq C\lambda^{2(\frac{1}{r_1}-\frac{1}{r_2}-1)} \lambda^{3(\frac{1}{p_1}-\frac{1}{p_2})} (\|f\|_{L_t^{r_1} L_x^{p_1}(Q_T)} + \lambda \|t^{\frac{1}{2}} f\|_{L_t^{r_1} L_x^{p_1}(Q_T)}), \quad 1 \leq p_1 \leq p_2 \leq \infty, \quad 1 \leq r_1 \leq r_2 \leq \infty. \tag{A-88}$$

Recall the smoothing effect (A-22) of the heat flow. When $p_1 = p_2$, one may write

$$\begin{aligned}
\lambda \left\| \int_0^t e^{(t-\tau)\Delta} f(\cdot,\tau)\,d\tau \right\|_{L^{p_1}(\mathbb{R}^3)}
&\leq C \int_0^{\frac{t}{2}} \|\nabla e^{(t-\tau)\Delta} f(\cdot,\tau)\|_{L^{p_1}(\mathbb{R}^3)}\,d\tau + \lambda \int_{\frac{t}{2}}^t \|e^{(t-\tau)\Delta} f(\cdot,\tau)\|_{L^{p_1}(\mathbb{R}^3)}\,d\tau \\
&\leq C \int_0^{\frac{t}{2}} (t-\tau)^{-\frac{1}{2}} e^{-c(t-\tau)\lambda^2} \|f(\cdot,\tau)\|_{L^{p_1}(\mathbb{R}^3)}\,d\tau + C\lambda \int_{\frac{t}{2}}^t e^{-c(t-\tau)\lambda^2} \tau^{-\frac{1}{2}} \|\tau^{\frac{1}{2}} f(\cdot,\tau)\|_{L^{p_1}(\mathbb{R}^3)}\,d\tau \\
&\leq C t^{-\frac{1}{2}} \left(\int_0^{\frac{t}{2}} e^{-c(t-\tau)\lambda^2} \|f(\cdot,\tau)\|_{L^{p_1}(\mathbb{R}^3)}\,ds + \lambda \int_{\frac{t}{2}}^t e^{-c(t-\tau)\lambda^2} \|\tau^{\frac{1}{2}} f(\cdot,\tau)\|_{L^{p_1}(\mathbb{R}^3)}\,d\tau \right)
\end{aligned} \tag{A-89}$$

and treat the subsequent integration in time by Young's convolution inequality. The case $p_2 > p_1$ follows from Bernstein's inequality. This leads us to a higher regularity estimate in Besov spaces analogous to Lemma A.1:

$$\left\| t^{\frac{1}{2}} \int_0^t e^{(t-s)\Delta} \mathbb{P} \operatorname{div} F(\cdot,s)\,ds \right\|_{\widetilde{L}_T^{r_2} \dot{B}_{p_2,q_2}^{3/p_2}} \leq c(\|F\|_{\widetilde{L}_T^{r_1} \dot{B}_{p_1,q_1}^{s_{p_1}+2/r_1-1}} + \|t^{1/2} F\|_{\widetilde{L}_T^{r_1} \dot{B}_{p_1,q_1}^{s_{p_1}+2/r}}), \tag{A-90}$$

when also $1 \leq q_1 \leq q_2 \leq \infty$.

Let us return to the proof. We will apply Lemma A.3 in the following Banach spaces to obtain (A-86):

$$Y_T := \{ v \in \widetilde{L}_T^r \dot{B}_{p,q}^{s_p+\frac{2}{r}} : t^{\frac{1}{2}} v \in \widetilde{L}_T^r \dot{B}_{p,q}^{\frac{3}{p}+\frac{2}{r}} \}, \tag{A-91}$$

$$\|v\|_{Y_T} := \max(\|v\|_{\widetilde{L}_T^r \dot{B}_{p,q}^{s_p+2/r}}, \|t^{\frac{1}{2}} v\|_{\widetilde{L}_T^r \dot{B}_{p,q}^{3/p+2/r}}). \tag{A-92}$$

Let us show that B is a bounded operator on Y_T with norm κ independent of the time $T > 0$. We need only prove boundedness of the upper part of the norm. Now, according to the low-high paraproduct estimate (A-19) and Sobolev embedding,

$$\begin{aligned}
\|t^{\frac{1}{2}} \dot{T}_v w\|_{\widetilde{L}_T^{r/2} \dot{B}_{p,q/2}^{s_p+4/r}} &\leq c\|v\|_{\widetilde{L}_T^r \dot{B}_{\infty,q}^{-1+2/r}} \|t^{\frac{1}{2}} w\|_{\widetilde{L}_T^r \dot{B}_{p,q}^{3/p+2/r}} \\
&\leq c\|v\|_{\widetilde{L}_T^r \dot{B}_{p,q}^{s_p+2/r}} \|t^{\frac{1}{2}} w\|_{\widetilde{L}_T^r \dot{B}_{p,q}^{3/p+2/r}} \leq c\|v\|_{Y_T} \|w\|_{Y_T},
\end{aligned} \tag{A-93}$$

and similarly for $\dot{T}_w v$. According to the high-high estimate (A-20), we have

$$\|t^{\frac{1}{2}}\dot{R}(v,w)\|_{\widetilde{L}^{r/2}_T\dot{B}^{-1+6/p+4/r}_{p/2,q/2}} \le c\|v\|_{\widetilde{L}^r_T\dot{B}^{s_p+2/r}_{p,q}}\|t^{\frac{1}{2}}w\|_{\widetilde{L}^r_T\dot{B}^{3/p+2/r}_{p,q}} \le c\|v\|_{Y_T}\|w\|_{Y_T}. \quad \text{(A-94)}$$

Since $v, w \in X_T$, there is no ambiguity modulo polynomials in forming the paraproduct operators. The bilinear estimate is then completed by combining (A-90) with (A-67), (A-93) for the low-high terms and (A-69), (A-94) for the high-high terms. By the same paraproduct estimates and applying (A-90) with $r_2 = \infty$, one may show that

$$B : Y_T \times Y_T \to C((0,T]; \dot{B}^{\frac{3}{p}}_{p,q}(\mathbb{R}^3)). \quad \text{(A-95)}$$

This is the property we will use below.

We are ready to conclude. Recall that $u \in C([0,T^*]; \dot{B}^{s_p}_{p,q}(\mathbb{R}^3))$ by (A-85). By continuity, there exists a time $T_1 > 0$ such that $\|e^{t\Delta}u(\cdot,t_0)\|_{Y_{T_1}} < (4\kappa)^{-1}$ as $t_0 \uparrow T^*$. Now we may apply Lemma A.3 in the space Y_{T_1} with $u(\cdot, T^*-T_1)$ as initial data, and the result agrees with the mild solution u on the time interval $[T^*-T_1, T^*]$. Therefore, we may take $T_0 := T^* - \frac{1}{2}T_1$ in (A-86) to complete the proof. (In fact, by similar arguments of resolving the equation, one may show that (A-86) holds everywhere away from the initial time.) □

Let us also provide the following characterization of the maximal time of existence.

Proposition A.5 (formation of singularity at blow-up time). *Let $3 < p < \infty$ and $u_0 \in L^p(\mathbb{R}^3)$ be a divergence-free vector field. Let $T^*(u_0) > 0$ be the maximal time of existence of the unique mild solution u such that $u(\cdot,t)$ is continuous in $L^p(\mathbb{R}^3)$. If $T^*(u_0) < \infty$, then u must have a singular point at time $T^*(u_0)$.*

Proof. This argument is based on similar arguments in [Rusin and Šverák 2011; Jia and Šverák 2013]. Let u be as in the hypothesis of the proposition and assume that $T^*(u_0) < \infty$. Following [Calderón 1990], for all $\delta > 0$, there exist divergence-free vector fields $U_0 \in L^2(\mathbb{R}^3) \cap L^p(\mathbb{R}^3)$ and $V_0 \in L^p(\mathbb{R}^3)$ such that

$$u_0 = U_0 + V_0, \quad \|V_0\|_{L^p(\mathbb{R}^3)} < \delta. \quad \text{(A-96)}$$

Let V denote the unique mild solution associated to the initial data V_0 such that $V(\cdot,t)$ is continuous in $L^p(\mathbb{R}^3)$. Furthermore, one may choose $0 < \delta \ll 1$ such that $T^*(V_0) \ge 2T^*(u_0)$. We will abbreviate $T^*(u_0)$ as T^*. The remainder U solves the perturbed Navier–Stokes equations on Q_{T^*},

$$\partial_t U - \Delta U + \operatorname{div} U \otimes U + \operatorname{div} U \otimes V + \operatorname{div} V \otimes U = -\nabla P, \quad \operatorname{div} U = 0, \quad \text{(A-97)}$$

with the initial condition $U(\cdot,0) = U_0$. By the well-posedness theory for (A-97) as well as the energy inequality, one may prove that U is in the energy space up to the blow-up time:

$$U \in L^\infty_t L^2_x \cap L^2_t \dot{H}^1_x(Q_{T^*}). \quad \text{(A-98)}$$

Based on the decomposition of u, we obtain

$$u \in L^3(Q_{T^*}) + L^\infty_t L^p_x(Q_{T^*}), \quad \text{(A-99)}$$

$$p \in L^{\frac{3}{2}}(Q_{T^*}) + L^\infty_t L^{\frac{p}{2}}_x(Q_{T^*}), \quad \text{(A-100)}$$

where $p := (-\Delta)^{-1} \operatorname{div}\operatorname{div} u \otimes u$ is the pressure associated to u. Recall now that u is in subcritical spaces, so one may justify the local energy inequality for (u, p) up to the blow-up time T^*. Moreover, (A-99) implies

$$\lim_{|x|\to\infty} \int_0^{T^*}\int_{B(x,1)} |u|^3 + |p|^{\frac{3}{2}}\,dx\,dt = 0. \tag{A-101}$$

Therefore, by the ε-regularity criterion in Theorem A.8, there exist constants $R, \kappa > 0$ and the set $K := (\mathbb{R}^3 \setminus B(R)) \times \left(\frac{1}{2}T^*, T^*\right)$ such that

$$\sup_K |u(x,t)| < \kappa. \tag{A-102}$$

Finally, suppose that u has no singular points at time T^*. This assumption, paired with the estimate (A-102), implies that $u \in L^\infty(Q_{\varepsilon,T^*})$ for some $\varepsilon \in \left(\frac{1}{2}T^*, T^*\right)$. Consider the bilinear estimates

$$\begin{aligned} \|B(v,w)\|_{L_t^\infty L_x^p(Q_T)} &\le cT^{\frac{1}{2}}\|v\|_{L_t^\infty L_x^p(Q_T)}\|w\|_{L^\infty(Q_T)},\\ \|B(v,w)\|_{L_t^\infty L_x^p(Q_T)} &\le cT^{\frac{1}{2}}\|v\|_{L^\infty(Q_T)}\|w\|_{L_t^\infty L_x^p(Q_T)} \end{aligned} \tag{A-103}$$

for all $T > 0$. Now one applies Lemma A.4 with the bilinear estimates (A-103) and the choice of spaces

$$X := L^\infty(Q_{t_0,T^*}), \quad E := X \cap L_t^\infty L_x^p(Q_{t_0,T^*}) \tag{A-104}$$

for each $\varepsilon < t_0 < T^*$. This prevents $\|u(\cdot,t)\|_{L^p(\mathbb{R}^3)}$ from becoming unbounded as $t \uparrow T^*$ and completes the proof. □

Corollary A.6. *Let u be the mild solution of Theorem A.2 with initial data u_0. If $T^*(u_0) < \infty$, then u has a singular point at time $T^*(u_0)$.*

Lastly, we record an existence theorem for mild solutions with initial data in subcritical Besov spaces. Similar results concerning smoothness of solutions belonging to the critical space $L^5(Q_T)$ were proven in [Dong and Du 2007].

Theorem A.7 (mild solutions in subcritical Besov spaces). *Let $3 < p, q \le \infty$, $0 < \varepsilon < -s_p$, $s := s_p + \varepsilon$, and $M > 0$. There exists a positive constant $\kappa := \kappa(s, p)$ such that for all divergence-free vector fields $u_0 \in \dot{B}^s_{p,q}(\mathbb{R}^3)$ with $\|u_0\|_{\dot{B}^s_{p,q}(\mathbb{R}^3)} \le M$, there exists a unique mild solution $u \in \mathcal{K}^s_p(Q_T)$ of the Navier–Stokes equations on Q_T with initial data u_0 and $T^{\frac{\varepsilon}{2}} := \kappa/M$. Moreover, the solution satisfies*

$$\partial_t^k \nabla^l u \in C((0,T]; L^p(\mathbb{R}^3)) \cap C((0,T]; L^\infty(\mathbb{R}^3)), \tag{A-105}$$

$$\|t^{k+\frac{l}{2}}\partial_t^k \nabla^l u\|_{\mathcal{K}^s_p(Q_T)} + \|t^{k+\frac{l}{2}}\partial_t^k \nabla^l u\|_{\mathcal{K}^{-1+\varepsilon}_\infty(Q_T)} \le c(s,p,k,l)M \tag{A-106}$$

for all integers $k, l \ge 0$. Hence, u is smooth in Q_T.

***ε*-regularity and backward uniqueness.** In this section, we record a number of important theorems relevant for the blow-up procedure in Theorem 1.1. Our primary reference is the seminal paper of Escauriaza, Seregin, and Šverák [Escauriaza et al. 2003].

We will now introduce the relevant notation and definitions. For $R > 0$ and $z_0 = (x_0, t_0) \in \mathbb{R}^{3+1}$, we define the Euclidean balls

$$B(x_0, R) := \{y \in \mathbb{R}^3 : |x_0 - y| < R\}, \quad B(R) := B(0, R), \tag{A-107}$$

and the parabolic balls

$$Q(z_0, R) := B(x_0, R) \times (t_0 - R^2, t_0), \quad Q(R) := Q(0, R). \tag{A-108}$$

Suppose that $u : Q(z_0, R) \to \mathbb{R}^3$ is a measurable function. We will say that z_0 is a *singular point* of u if for all $0 < r < R$, we have $u \notin L^\infty(Q(z_0, r))$. In this case, we will say that u is *singular* at z_0. Otherwise, we will say that z_0 is a *regular point* of u.

We say that (v, q) is a *suitable weak solution* of the Navier–Stokes equations in $Q(z_0, R)$ if the following requirements are satisfied:

(i) $v \in L^\infty_t L^2_x \cap L^2_t H^1_x(Q(z_0, R))$ and $q \in L^{\frac{3}{2}}_{t,x}(Q(z_0, R))$.

(ii) (v, q) solves the Navier–Stokes equations in the sense of distributions on $Q(z_0, R)$.

(iii) For all $0 \le \varphi \in C_0^\infty(B(R) \times (t_0 - R^2, t_0])$, the pair (v, q) satisfies the local energy inequality

$$\int_{B(x_0,R)} \varphi |v(x,t)|^2 \, dx + 2 \int_{t_0 - R^2}^{t} \int_{B(x_0,R)} \varphi |\nabla v|^2 \, dx \, dt' \\ \le \int_{t_0 - R^2}^{t} \int_{B(x_0,R)} (|v|^2 (\Delta \varphi + \partial_t \varphi) + v \cdot \nabla \varphi (|v|^2 + 2q)) \, dx \, dt', \tag{A-109}$$

for every $t \in (t_0 - R^2, t_0]$.

We say that (v, q) is a suitable weak solution of the Navier–Stokes equations in an open set $\Omega \subset \mathbb{R}^{3+1}$ if (v, q) is a suitable weak solution in each parabolic ball $Q \subset \Omega$.

Note that we permit the test functions φ in the local energy inequality to be supported near t_0, but this is not strictly necessary. Also, one may use the Navier–Stokes equations to demonstrate that a suitable weak solution $u(\cdot, t)$ is weakly continuous on $[t_0 - R^2, t_0]$ as a function with values in $L^2(B(x, R))$.

Let us show that the distributional local energy inequality (2-9) in the definition of Calderón solution implies the local energy inequality (A-109) in the definition of suitable weak solution. Let $0 \le \eta \in C_0^\infty(\mathbb{R})$ such that $\eta \equiv 1$ when $|\tau| \le \frac{1}{4}$, $\eta \equiv 0$ when $|\tau| \ge \frac{1}{2}$, and $\int_{\mathbb{R}} \eta \, d\tau = 1$. For all $\varepsilon > 0$, define $\eta_\varepsilon(\tau) := \varepsilon^{-1} \eta(\varepsilon^{-1} \tau)$. Now let $0 \le \varphi \in C_0^\infty(B(R) \times (t_0 - R^2, t_0])$. Define

$$\Phi_{t,\varepsilon}(x, t') := \varphi(x, t') \left(1 - \int_\infty^{t'} \eta_\varepsilon(\tau - t) \, d\tau\right), \quad t \in (t_0 - R^2, t_0), \quad \varepsilon > 0. \tag{A-110}$$

Using $\Phi_{t,\varepsilon}$ as a test function in (2-10) and passing to the limit $\varepsilon \downarrow 0$ proves (A-109) for almost every $t \in (t_0 - R^2, t_0)$. That the inequality is true for all $t \in (t_0 - R^2, t_0]$ follows from the weak continuity of $u(\cdot, t) \in L^2(B(x, R))$.

We will now state the Caffarelli–Kohn–Nirenberg ε-regularity criterion for suitable weak solutions; see [Caffarelli et al. 1982; Lin 1998; Ladyzhenskaya and Seregin 1999; Escauriaza et al. 2003].

Theorem A.8 (ε-regularity criterion [Escauriaza et al. 2003]). *There exist absolute constants $\varepsilon_0 > 0$ and $c_{0,k} > 0$, $k \in \mathbb{N}$, with the following property. Assume (v, q) is a suitable weak solution on $Q(1)$ satisfying*

$$\int_{Q(1)} |v|^3 + |q|^{\frac{3}{2}} \, dx \, dt < \varepsilon_0. \tag{A-111}$$

Then, for each $k \in \mathbb{N}$, we know $\nabla^{k-1} v$ is Hölder continuous on $\overline{Q(\frac{1}{2})}$ and satisfies

$$\sup_{Q(\frac{1}{2})} |\nabla^{k-1} v(x,t)| < c_{0,k}. \tag{A-112}$$

The ε-regularity criterion may be used to prove that singularities of suitable weak solutions persist under locally strong limits.

Proposition A.9 (persistence of singularity [Rusin and Šverák 2011]). *Let (v_k, q_k) be a sequence of suitable weak solutions on $Q(1)$ such that $v_k \to v$ in $L^3(Q(1))$, $q_k \rightharpoonup q$ in $L^{\frac{3}{2}}(Q(1))$. Assume v_k is singular at $(x_k, t_k) \in \overline{Q(1)}$ and $(x_k, t_k) \to 0$. Then v is singular at the spacetime origin.*

Finally, we recall two theorems concerning backward uniqueness and unique continuation of solutions to differential inequalities.

Theorem A.10 (backward uniqueness [Escauriaza et al. 2003]). *Let $Q_+ := \mathbb{R}^3_+ \times (0,1)$. Suppose $u : Q_+ \to \mathbb{R}^3$ satisfies the following conditions*:

(i) $|u_t + \Delta u| \leq c(|\nabla u| + |u|)$ *a.e. in Q_+ for some $c > 0$.*

(ii) $u(\cdot, 0) = 0$.

(iii) $|u(x,t)| \leq e^{M|x|^2}$ *in Q_+.*

(iv) $u, u_t, \nabla^2 u \in L^2_t (L^2_{\text{loc}})_x (Q_+)$.

Then $u \equiv 0$ on Q_+.

To make sense of condition (ii), one may use that $u \in C([0,1]; \mathcal{D}'(\mathbb{R}^3_+))$, due to condition (iv).

Theorem A.11 (unique continuation [Escauriaza et al. 2003]). *Let R, $T > 0$ and $Q(R,T) := B(R) \times (0,T) \subset \mathbb{R}^{3+1}$. Suppose $u : Q(R,T) \to \mathbb{R}^3$ satisfies the following conditions*:

(i) $u, u_t, \nabla^2 u \in L^2(Q(R,T))$.

(ii) $|u_t + \Delta u| \leq c(|\nabla u| + |u|)$ *a.e. in $Q(R,T)$ for some $c > 0$.*

(iii) $|u(x,t)| \leq C_k(|x| + \sqrt{t})^k$ *in $Q(R,T)$ for some $C_k > 0$, all $k \geq 0$.*

Then $u(x,0) = 0$ for all $x \in B(R)$.

Acknowledgments

The author would like to thank his advisor, Vladimír Šverák, as well as Daniel Spirn, Alex Gutierrez, Laurel Ohm, and Tobias Barker for valuable suggestions. The author also thanks the reviewer for a careful reading of the manuscript.

References

[Albritton and Barker 2018] D. Albritton and T. Barker, "Global weak Besov solutions of the Navier–Stokes equations and applications", preprint, 2018. arXiv

[Auscher et al. 2004] P. Auscher, S. Dubois, and P. Tchamitchian, "On the stability of global solutions to Navier–Stokes equations in the space", *J. Math. Pures Appl.* (9) **83**:6 (2004), 673–697. MR Zbl

[Bahouri et al. 2011a] H. Bahouri, J.-Y. Chemin, and R. Danchin, *Fourier analysis and nonlinear partial differential equations*, Grundlehren der Mathematischen Wissenschaften **343**, Springer, 2011. MR Zbl

[Bahouri et al. 2011b] H. Bahouri, A. Cohen, and G. Koch, "A general wavelet-based profile decomposition in the critical embedding of function spaces", *Confluentes Math.* **3**:3 (2011), 387–411. MR Zbl

[Bahouri et al. 2014] H. Bahouri, J.-Y. Chemin, and I. Gallagher, "Stability by rescaled weak convergence for the Navier–Stokes equations", *C. R. Math. Acad. Sci. Paris* **352**:4 (2014), 305–310. MR Zbl

[Barker 2017] T. Barker, "Existence and weak* stability for the Navier–Stokes system with initial values in critical Besov spaces", preprint, 2017. arXiv

[Barker 2018] T. Barker, "Uniqueness results for weak Leray–Hopf solutions of the Navier–Stokes system with initial values in critical spaces", *J. Math. Fluid Mech.* **20**:1 (2018), 133–160.

[Barker and Seregin 2017] T. Barker and G. Seregin, "A necessary condition of potential blowup for the Navier–Stokes system in half-space", *Math. Ann.* **369**:3-4 (2017), 1327–1352. MR Zbl

[Barker et al. 2013] T. Barker, G. Seregin, and V. Šverák, "On stability of weak Navier–Stokes solutions with large $L^{3,\infty}$ initial data", preprint, 2013.

[Bourgain and Pavlović 2008] J. Bourgain and N. Pavlović, "Ill-posedness of the Navier–Stokes equations in a critical space in 3D", *J. Funct. Anal.* **255**:9 (2008), 2233–2247. MR Zbl

[Caffarelli et al. 1982] L. Caffarelli, R. Kohn, and L. Nirenberg, "Partial regularity of suitable weak solutions of the Navier–Stokes equations", *Comm. Pure Appl. Math.* **35**:6 (1982), 771–831. MR Zbl

[Calderón 1990] C. P. Calderón, "Existence of weak solutions for the Navier–Stokes equations with initial data in L^p", *Trans. Amer. Math. Soc.* **318**:1 (1990), 179–200. MR Zbl

[Dong and Du 2007] H. Dong and D. Du, "On the local smoothness of solutions of the Navier–Stokes equations", *J. Math. Fluid Mech.* **9**:2 (2007), 139–152. MR Zbl

[Dong and Du 2009] H. Dong and D. Du, "The Navier–Stokes equations in the critical Lebesgue space", *Comm. Math. Phys.* **292**:3 (2009), 811–827. MR Zbl

[Escauriaza et al. 2003] L. Escauriaza, G. A. Seregin, and V. Šverák, "$L_{3,\infty}$-solutions of Navier–Stokes equations and backward uniqueness", *Uspekhi Mat. Nauk* **58**:2(350) (2003), 3–44. In Russian; translated in *Russian Math. Surveys* **58**:2 (2003), 211–250. MR Zbl

[Gallagher 2001] I. Gallagher, "Profile decomposition for solutions of the Navier–Stokes equations", *Bull. Soc. Math. France* **129**:2 (2001), 285–316. MR Zbl

[Gallagher et al. 2002] I. Gallagher, D. Iftimie, and F. Planchon, "Asymptotics and stability for global solutions to the Navier–Stokes equations", exposé VI in *Journées "Équations aux Dérivées Partielles"* (Forges-les-Eaux, 2002), Univ. Nantes, 2002. MR Zbl

[Gallagher et al. 2013] I. Gallagher, G. S. Koch, and F. Planchon, "A profile decomposition approach to the $L_t^\infty(L_x^3)$ Navier–Stokes regularity criterion", *Math. Ann.* **355**:4 (2013), 1527–1559. MR Zbl

[Gallagher et al. 2016] I. Gallagher, G. S. Koch, and F. Planchon, "Blow-up of critical Besov norms at a potential Navier–Stokes singularity", *Comm. Math. Phys.* **343**:1 (2016), 39–82. MR Zbl

[Germain et al. 2007] P. Germain, N. Pavlović, and G. Staffilani, "Regularity of solutions to the Navier–Stokes equations evolving from small data in BMO^{-1}", *Int. Math. Res. Not.* **2007**:21 (2007), art. id. rnm087. MR Zbl

[Jia and Šverák 2013] H. Jia and V. Šverák, "Minimal L^3-initial data for potential Navier–Stokes singularities", *SIAM J. Math. Anal.* **45**:3 (2013), 1448–1459. MR Zbl

[Kato 1984] T. Kato, "Strong L^p-solutions of the Navier–Stokes equation in $\mathbb{R}^m$, with applications to weak solutions", *Math. Z.* **187**:4 (1984), 471–480. MR Zbl

[Kenig and Koch 2011] C. E. Kenig and G. S. Koch, "An alternative approach to regularity for the Navier–Stokes equations in critical spaces", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **28**:2 (2011), 159–187. MR Zbl

[Kenig and Merle 2006] C. E. Kenig and F. Merle, "Global well-posedness, scattering and blow-up for the energy-critical, focusing, non-linear Schrödinger equation in the radial case", *Invent. Math.* **166**:3 (2006), 645–675. MR Zbl

[Kikuchi and Seregin 2007] N. Kikuchi and G. Seregin, "Weak solutions to the Cauchy problem for the Navier–Stokes equations satisfying the local energy inequality", pp. 141–164 in *Nonlinear equations and spectral theory*, edited by M. S. Birman and N. N. Uraltseva, Amer. Math. Soc. Transl. Ser. 2 **220**, Amer. Math. Soc., Providence, RI, 2007. MR Zbl

[Koch 2010] G. S. Koch, "Profile decompositions for critical Lebesgue and Besov space embeddings", *Indiana Univ. Math. J.* **59**:5 (2010), 1801–1830. MR Zbl

[Koch and Tataru 2001] H. Koch and D. Tataru, "Well-posedness for the Navier–Stokes equations", *Adv. Math.* **157**:1 (2001), 22–35. MR Zbl

[Ladyzhenskaya and Seregin 1999] O. A. Ladyzhenskaya and G. A. Seregin, "On partial regularity of suitable weak solutions to the three-dimensional Navier–Stokes equations", *J. Math. Fluid Mech.* **1**:4 (1999), 356–387. MR Zbl

[Lemarié-Rieusset 2002] P. G. Lemarié-Rieusset, *Recent developments in the Navier–Stokes problem*, Research Notes in Mathematics **431**, CRC, Boca Raton, FL, 2002. MR Zbl

[Leray 1934] J. Leray, "Sur le mouvement d'un liquide visqueux emplissant l'espace", *Acta Math.* **63**:1 (1934), 193–248. MR Zbl

[Lin 1998] F. Lin, "A new proof of the Caffarelli–Kohn–Nirenberg theorem", *Comm. Pure Appl. Math.* **51**:3 (1998), 241–257. MR Zbl

[Maekawa et al. 2017] Y. Maekawa, H. Miura, and C. Prange, "Local energy weak solutions for the Navier–Stokes equations in the half-space", preprint, 2017. arXiv

[Phuc 2015] N. C. Phuc, "The Navier–Stokes equations in nonendpoint borderline Lorentz spaces", *J. Math. Fluid Mech.* **17**:4 (2015), 741–760. MR Zbl

[Rusin and Šverák 2011] W. Rusin and V. Šverák, "Minimal initial data for potential Navier–Stokes singularities", *J. Funct. Anal.* **260**:3 (2011), 879–891. MR Zbl

[Seregin 2012a] G. Seregin, "A certain necessary condition of potential blow up for Navier–Stokes equations", *Comm. Math. Phys.* **312**:3 (2012), 833–845. MR Zbl

[Seregin 2012b] G. Seregin, "Leray–Hopf solutions to Navier–Stokes equations with weakly converging initial data", pp. 251–258 in *Mathematical aspects of fluid mechanics*, edited by J. C. Robinson et al., London Math. Soc. Lecture Note Ser. **402**, Cambridge Univ. Press, 2012. MR Zbl

[Seregin 2015] G. Seregin, *Lecture notes on regularity theory for the Navier–Stokes equations*, World Scientific, Hackensack, NJ, 2015. MR Zbl

[Seregin and Šverák 2017] G. Seregin and V. Šverák, "On global weak solutions to the Cauchy problem for the Navier–Stokes equations with large L_3-initial data", *Nonlinear Anal.* **154** (2017), 269–296. MR Zbl

[Serrin 1962] J. Serrin, "On the interior regularity of weak solutions of the Navier–Stokes equations", *Arch. Rational Mech. Anal.* **9** (1962), 187–195. MR Zbl

[Wang 2008] C. Wang, "Heat flow of harmonic maps whose gradients belong to $L^n_x L^\infty_t$", *Arch. Ration. Mech. Anal.* **188**:2 (2008), 351–369. MR Zbl

Received 20 Feb 2017. Revised 3 Dec 2017. Accepted 14 Feb 2018.

DALLAS ALBRITTON: albri050@umn.edu
School of Mathematics, University of Minnesota, Minneapolis, MN, United States

ANALYSIS AND PDE
Vol. 11, No. 6, 2018
dx.doi.org/10.2140/apde.2018.11.1457

DOLGOPYAT'S METHOD AND THE FRACTAL UNCERTAINTY PRINCIPLE

SEMYON DYATLOV AND LONG JIN

We show a fractal uncertainty principle with exponent $\frac{1}{2}-\delta+\varepsilon,\ \varepsilon>0$, for Ahlfors–David regular subsets of $\mathbb{R}$ of dimension $\delta\in(0,1)$. This is an improvement over the volume bound $\frac{1}{2}-\delta$, and ε is estimated explicitly in terms of the regularity constant of the set. The proof uses a version of techniques originating in the works of Dolgopyat, Naud, and Stoyanov on spectral radii of transfer operators. Here the group invariance of the set is replaced by its fractal structure. As an application, we quantify the result of Naud on spectral gaps for convex cocompact hyperbolic surfaces and obtain a new spectral gap for open quantum baker maps.

1. Introduction

A *fractal uncertainty principle* (FUP) states that no function can be localized close to a fractal set in both position and frequency. Its most basic form is

$$\|\mathbb{1}_{\Lambda(h)}\mathcal{F}_h\mathbb{1}_{\Lambda(h)}\|_{L^2(\mathbb{R})\to L^2(\mathbb{R})}=\mathcal{O}(h^\beta)\quad \text{as } h\to 0, \tag{1-1}$$

where $\Lambda(h)$ is the h-neighborhood of a bounded set $\Lambda\subset\mathbb{R}$, β is called the *exponent* of the uncertainty principle, and $\mathcal{F}_h$ is the semiclassical Fourier transform:

$$\mathcal{F}_h u(\xi)=(2\pi h)^{-\frac{1}{2}}\int_{\mathbb{R}} e^{-\frac{ix\xi}{h}}u(x)\,dx. \tag{1-2}$$

We additionally assume that Λ is an Ahlfors–David regular set (see Definition 1.1) of dimension $\delta\in(0,1)$ with some regularity constant $C_R>1$. Using the bounds $\|\mathcal{F}_h\|_{L^2\to L^2}=1,\ \|\mathcal{F}_h\|_{L^1\to L^\infty}\le h^{-\frac{1}{2}}$, the Lebesgue volume bound $\mu_L(\Lambda(h))\le Ch^{1-\delta}$, and Hölder's inequality, it is easy to obtain (1-1) with $\beta=\max\big(0,\frac{1}{2}-\delta\big)$.

Fractal uncertainty principles were applied by Dyatlov and Zahl [2016], Dyatlov and Jin [2017], and Bourgain and Dyatlov [2016] to the problem of essential spectral gap in quantum chaos: *which open quantum chaotic systems have exponential decay of local energy at high frequency*? A fractal uncertainty principle can be used to show local energy decay $\mathcal{O}(e^{-\beta t})$, as was done for convex cocompact hyperbolic quotients in [Dyatlov and Zahl 2016] and for open quantum baker's maps in [Dyatlov and Jin 2017]. Here Λ is related to the set of all trapped classical trajectories of the system and (1-1) needs to be replaced by a more general statement, in particular allowing for a different phase in (1-2). The volume bound $\beta=\frac{1}{2}-\delta$ corresponds to the Patterson–Sullivan gap or more generally, the *pressure gap*. See Sections 4–5 below for a more detailed discussion.

MSC2010: 28A80, 35B34, 81Q50.
Keywords: resonances, fractal uncertainty principle.

A natural question is: can one obtain (1-1) with $\beta > \max\big(0, \frac{1}{2}-\delta\big)$, and if so, how does the size of the improvement depend on δ and C_R? Partial answers to this question have been obtained in the papers mentioned above:

- Dyatlov and Zahl [2016] obtained FUP with $\beta > 0$ when $\big|\delta - \frac{1}{2}\big|$ is small depending on C_R, and gave the bound $\beta > \exp(-\boldsymbol{K}(1+\log^{14} C_R))$, where $\boldsymbol{K}$ is a global constant.
- Bourgain and Dyatlov [2016] proved FUP with $\beta > 0$ in the entire range $\delta \in (0,1)$, with no explicit bounds on the dependence of β on δ, C_R.
- Dyatlov and Jin [2017] showed that discrete Cantor sets satisfy FUP with $\beta > \max\big(0, \frac{1}{2}-\delta\big)$ in the entire range $\delta \in (0,1)$ and obtained quantitative lower bounds on the size of the improvement — see Section 5 below.

Our main result, Theorem 1, shows that FUP holds with $\beta > \frac{1}{2}-\delta$ in the case $\delta \in (0,1)$, and gives bounds on $\beta - \frac{1}{2} + \delta$ which are polynomial in C_R and thus stronger than the ones in [Dyatlov and Zahl 2016]. Applications include

- an essential spectral gap for convex cocompact hyperbolic surfaces of size $\beta > \frac{1}{2}-\delta$, recovering and making quantitative the result of [Naud 2005], see Section 4;
- an essential spectral gap of size $\beta > \max\big(0, \frac{1}{2}-\delta\big)$ for open quantum baker's maps, extending the result of [Dyatlov and Jin 2017] to matrices whose sizes are not powers of the base, see Section 5. (For the case $\delta > \frac{1}{2}$ we use the results of [Bourgain and Dyatlov 2016] rather than Theorem 1.)

1A. *Statement of the result.* We recall the following definition of Ahlfors–David regularity, which requires that a set (or a measure) has the same dimension δ at all points and on a range of scales:

Definition 1.1. Let $X \subset \mathbb{R}$ be compact, μ_X be a finite measure supported on X, and $\delta \in [0,1]$. We say that (X, μ_X) is δ-*regular* up to scale $h \in [0,1)$ with regularity constant $C_R \ge 1$ if

- for each interval I of size $|I| \ge h$, we have $\mu_X(I) \le C_R |I|^\delta$;
- if additionally $|I| \le 1$ and the center of I lies in X, then $\mu_X(I) \ge C_R^{-1}|I|^\delta$.

Our fractal uncertainty principle has a general form which allows for two different sets X, Y of different dimensions in (1-1), replaces the Lebesgue measure by the fractal measures μ_X, μ_Y, and allows a general nondegenerate phase and amplitude in (1-2):

Theorem 1. *Assume that (X, μ_X) is δ-regular, and (Y, μ_Y) is δ'-regular, up to scale $h \in (0,1)$ with constant C_R, where $0 < \delta, \delta' < 1$, and $X \subset I_0$, $Y \subset J_0$ for some intervals I_0, J_0. Consider an operator $\mathcal{B}_h : L^1(Y, \mu_Y) \to L^\infty(X, \mu_X)$ of the form*

$$\mathcal{B}_h f(x) = \int_Y \exp\left(\frac{i\Phi(x,y)}{h}\right) G(x,y) f(y)\, d\mu_Y(y), \tag{1-3}$$

where $\Phi(x,y) \in C^2(I_0 \times J_0; \mathbb{R})$ satisfies $\partial^2_{xy}\Phi \ne 0$ and $G(x,y) \in C^1(I_0 \times J_0; \mathbb{C})$.

Then there exist constants $C, \varepsilon_0 > 0$ such that

$$\|\mathcal{B}_h\|_{L^2(Y,\mu_Y) \to L^2(X,\mu_X)} \le C h^{\varepsilon_0}. \tag{1-4}$$

Here ε_0 depends only on δ, δ', C_R,

$$\varepsilon_0 = (5C_R)^{-80\left(\frac{1}{\delta(1-\delta)}+\frac{1}{\delta'(1-\delta')}\right)}, \tag{1-5}$$

and C additionally depends on I_0, J_0, Φ, G.

Remarks. (1) Theorem 1 implies the Lebesgue-measure version of FUP, (1-1), with exponent $\beta = \frac{1}{2} - \delta + \varepsilon_0$. Indeed, assume that (Λ, μ_Λ) is δ-regular up to scale h with constant C_R. Put $X := \Lambda(h)$ and let μ_X be $h^{\delta-1}$ times the restriction of the Lebesgue measure to X. Then (X, μ_X) is δ-regular up to scale h with constant $30C_R^2$; see Lemma 2.2. We apply Theorem 1 with $(Y, \mu_Y) := (X, \mu_X)$, $G \equiv 1$, and $\Phi(x, y) = -xy$; then

$$\|\mathbb{1}_{\Lambda(h)}\mathcal{F}_h\mathbb{1}_{\Lambda(h)}\|_{L^2(\mathbb{R})\to L^2(\mathbb{R})} = \frac{h^{\frac{1}{2}-\delta}}{\sqrt{2\pi}}\|\mathcal{B}_h\|_{L^2(X,\mu_X)\to L^2(X,\mu_X)} \le Ch^{\frac{1}{2}-\delta+\varepsilon_0}.$$

(2) Definition 1.1 is slightly stronger than [Bourgain and Dyatlov 2016, Definition 1.1] (where "up to scale h" should be interpreted as "on scales h to 1") because it imposes an upper bound on $\mu_L(I)$ when $|I| > 1$. However, this difference is insignificant as long as X is compact. Indeed, if $X \subset [-R, R]$ for some integer $R > 0$, then using upper bounds on μ_L on intervals of size 1 we get $\mu_L(I) \le \mu_L(X) \le 2RC_R \le 2RC_R|I|^\delta$ for each interval I of size $|I| > 1$.

(3) The restriction $\delta, \delta' > 0$ is essential. Indeed, if $\delta' = 0$, $Y = \{0\}$, μ_Y is the delta measure, and $f \equiv 1$, $G \equiv 1$, then

$$\|\mathcal{B}_h f\|_{L^2(X,\mu_X)} = \sqrt{\mu_X(X)}.$$

The restriction $\delta, \delta' < 1$ is technical; however, in the application to Lebesgue-measure FUP this restriction is not important since $\beta = \frac{1}{2} - \delta + \varepsilon_0 < 0$ when δ is close to 1.

(4) The constants in (1-5) are far from sharp. However, the dependence of ε_0 on C_R cannot be removed entirely. Indeed, [Dyatlov and Jin 2017] gives examples of Cantor sets for which the best exponent ε_0 in (1-4) decays polynomially as $C_R \to \infty$; see Proposition 3.17 of that paper. See also Sections 5B–5C.

1B. ***Ideas of the proof.*** The proof of Theorem 1 is inspired by the method originally developed by Dolgopyat [1998] and its application to essential spectral gaps for convex cocompact hyperbolic surfaces by Naud [2005]. In fact, Theorem 1 implies a quantitative version of Naud's result; see Section 4. More recently, Dolgopyat's method has been applied to the spectral-gap problem by Petkov and Stoyanov [2010], Stoyanov [2011; 2012], Oh and Winter [2016], and Magee, Oh and Winter [Magee et al. 2017].

We give a sketch of the proof, assuming for simplicity that $G \equiv 1$. For $f \in L^2(Y, \mu_Y)$, we have

$$\|\mathcal{B}_h f\|_{L^2(X,\mu_X)} \le \sqrt{\mu_X(X)\mu_Y(Y)} \cdot \|f\|_{L^2(Y,\mu_Y)}, \tag{1-6}$$

applying Hölder's inequality and the bound $\|\mathcal{B}_h\|_{L^1(X,\mu_X)\to L^\infty(Y,\mu_Y)} \le 1$. However, under a mild assumption on the differences between the phases $\Phi(x, y)$ for different x, y, the resulting estimate is not sharp, as illustrated by the following example where $X = Y = \{1, 2\}$, $\mu_X(j) = \mu_Y(j) = \frac{1}{2}$ for $j = 1, 2$, and $\omega_{j\ell} := \Phi(j, \ell)/h$:

Lemma 1.2. *Assume that* $\omega_{j\ell} \in \mathbb{R}$, $j, \ell = 1, 2$, *satisfy*

$$\tau := \omega_{11} + \omega_{22} - \omega_{12} - \omega_{21} \notin 2\pi\mathbb{Z}. \tag{1-7}$$

For $f_1, f_2 \in \mathbb{C}$, *put*

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} := \frac{1}{2} \begin{pmatrix} \exp(i\omega_{11}) & \exp(i\omega_{12}) \\ \exp(i\omega_{21}) & \exp(i\omega_{22}) \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}.$$

Assume that $(f_1, f_2) \neq 0$. *Then*

$$|u_1|^2 + |u_2|^2 < |f_1|^2 + |f_2|^2. \tag{1-8}$$

Remark. Note that (1-8) cannot be replaced by either of the statements

$$|u_1| + |u_2| < |f_1| + |f_2|, \quad \max(|u_1|, |u_2|) < \max(|f_1|, |f_2|).$$

Indeed, the first statement fails when $f_1 = 0$, $f_2 = 1$. The second one fails if $\omega_{11} = \omega_{12}$ and $f_1 = f_2 = 1$. This explains why we use L^2 norms in the iteration step, Lemma 3.2.

Proof. We have

$$\tfrac{1}{2}(|u_1|^2 + |u_2|^2) \leq \max(|u_1|^2, |u_2|^2) \leq \left(\tfrac{1}{2}(|f_1| + |f_2|)\right)^2 \leq \tfrac{1}{2}(|f_1|^2 + |f_2|^2). \tag{1-9}$$

Assume that (1-8) does not hold. Then the inequalities in (1-9) have to be equalities, which implies that $|u_1| = |u_2|$, $|f_1| = |f_2| > 0$, and for $a = 1, 2$,

$$\exp(i(\omega_{a1} - \omega_{a2})) f_1 \bar{f_2} \geq 0.$$

The latter statement contradicts (1-7). □

To get the improvement h^{ε_0} in (1-4), we use the nonsharpness of (1-6) on many scales:

- We fix a large integer $L > 1$ depending on δ, C_R and discretize X and Y on scales $1, L^{-1}, \dots, L^{-K}$, where $h \sim L^{-K}$. This results in two trees of intervals V_X, V_Y, with vertices of height k corresponding to intervals of length $\sim L^{-k}$.

- For each interval J in the tree V_Y, we consider the function

$$F_J(x) = \frac{1}{\mu_Y(J)} \exp\left(-\frac{i\Phi(x, y_J)}{h}\right) \mathcal{B}_h(\mathbb{1}_J f)(x),$$

where y_J is the center of J. The function F_J oscillates on scale $h/|J|$. Thus both F_J and the rescaled derivative $h|J|^{-1}F_J'$ are controlled in uniform norm by $\|f\|_{L^1(Y,\mu_Y)}$. We express this fact using the spaces $\mathcal{C}_\theta$ introduced in Section 2B.

- If $J_1, \dots, J_B \in V_Y$ are the children of J, then F_J can be written as a convex combination of $F_{J_1}, \dots, F_{J_B}$ multiplied by some phase factors $e^{i\Psi_b}$; see (3-12). We then employ an iterative procedure which estimates a carefully chosen norm of F_J via the norms of $F_{J_1}, \dots, F_{J_B}$. Each step in this procedure gives a gain $1 - \varepsilon_1 < 1$ in the norm, and after K steps we obtain a gain polynomial in h.

- To obtain a gain at each step, we consider two intervals $I \in V_X$, $J \in V_Y$ such that $|I| \cdot |J| \sim Lh$, take their children $I_1, \dots, I_A$ and $J_1, \dots, J_B$, and argue similarly to Lemma 1.2 to show that the triangle inequality for $e^{i\Psi_1}J_1, \dots, e^{i\Psi_B}J_B$ cannot be sharp on all the intervals $I_1, \dots, I_A$.

• To do the latter, we take two pairs of children I_a, I'_a (with generic points in I_a, I'_a denoted by $x_a, x_{a'}$) and J_b, J'_b. Due to the control on the derivatives of F_{J_b}, the differences $|F_{J_b}(x_a) - F_{J_b}(x_{a'})|$ and $|F_{J_{b'}}(x_a) - F_{J_{b'}}(x_{a'})|$ are bounded by $(Lh)^{-1}|J| \cdot |x_a - x_{a'}|$. On the other hand, the phase shift τ from (1-7) equals

$$\tau = \Psi_b(x_a) + \Psi_{b'}(x_{a'}) - \Psi_{b'}(x_a) - \Psi_b(x_{a'}) \sim h^{-1}(x_a - x_{a'})(y_b - y_{b'}).$$

Choosing a, a', b, b' such that $|x_a - x_{a'}| \sim L^{-\frac{2}{3}}|I|$, $|y_b - y_{b'}| \sim L^{-\frac{2}{3}}|J|$, and recalling that $|I| \cdot |J| \sim Lh$, we see that $\tau \sim L^{-\frac{1}{3}}$ does not lie in $2\pi\mathbb{Z}$ and it is larger than $(Lh)^{-1}|J| \cdot |x_a - x_{a'}| \sim L^{-\frac{2}{3}}$. This gives the necessary improvement on each step. Keeping track of the parameters in the argument, we obtain the bound (1-5) on ε_0.

This argument has many similarities with the method of Dolgopyat mentioned above. In particular, an inductive argument using L^2 norms appears for instance in [Naud 2005, Lemma 5.4], which also features the spaces $\mathcal{C}_\theta$. The choice of children $I_a, I_{a'}, J_b, J_{b'}$ in the last step above is similar to the nonlocal integrability condition (NLIC); see for instance [Naud 2005, Sections 2 and 5.3]. However, our inductive Lemma 3.2 avoids the use of Dolgopyat operators and dense subsets, see for instance [Naud 2005, p.138], instead relying on strict convexity of balls in Hilbert spaces, see Lemma 2.7.

Moreover, the strategy of obtaining an essential spectral gap for hyperbolic surfaces in the present paper is significantly different from that of [Naud 2005]. The latter uses zeta-function techniques to reduce the spectral-gap question to a spectral radius bound of a Ruelle transfer operator of the Bowen–Series map associated to the surface. The present paper instead relies on microlocal analysis of the scattering resolvent in [Dyatlov and Zahl 2016] to reduce the gap problem to a fractal uncertainty principle, thus decoupling the dynamical aspects of the problem from the combinatorial ones. The role of the group invariance of the limit set, used in [Naud 2005], is played here by its δ-regularity, proved by Sullivan [1979], and words in the group are replaced by vertices in the discretizing tree.

1C. *Structure of the paper.*

- In Section 2, we establish basic properties of Ahlfors–David regular sets (Section 2A), introduce the functional spaces used (Section 2B), and show several basic identities and inequalities (Section 2C).
- In Section 3, we prove Theorem 1.
- In Section 4, we apply Theorem 1 and the results of [Dyatlov and Zahl 2016] to establish an essential spectral gap for convex cocompact hyperbolic surfaces.
- In Section 5, we apply Theorem 1 and the results of [Dyatlov and Jin 2017; Bourgain and Dyatlov 2016] to establish an essential spectral gap for open quantum baker's maps.

2. Preliminaries

2A. ***Regular sets and discretization.*** An *interval* in $\mathbb{R}$ is a subset of the form $I = [c, d]$, where $c < d$. Define the center of I by $\frac{1}{2}(c + d)$ and the size of I by $|I| = d - c$.

Let μ be a finite measure on $\mathbb{R}$ with compact support. Fix an integer $L \geq 2$. Following [Dyatlov and Zahl 2016, Section 6.4], we describe the *discretization of μ with base L*. For each $k \in \mathbb{Z}$, let V_k be the set of all intervals $I = [c, d]$ which satisfy the following conditions:

- $c, d \in L^{-k}\mathbb{Z}$.
- For each $q \in L^{-k}\mathbb{Z}$ with $c \leq q < d$, we have $\mu([q, q + L^{-k}]) > 0$.
- $\mu_X([c - L^{-k}, c]) = \mu([d, d + L^{-k}]) = 0$.

In other words, V_k is obtained by partitioning $\mathbb{R}$ into intervals of size L^{-k}, throwing out intervals of zero measure μ, and merging consecutive intervals.

We define the set of vertices of the discretization as

$$V := \bigsqcup_{k \in \mathbb{Z}} V_k,$$

and define the *height function* by putting $H(I) := k$ if $I \in V_k$. (It is possible that V_k intersect for different k, so formally speaking, a vertex is a pair (k, I), where $I \in V_k$.) We say that $I \in V_k$ is a *parent* of $I' \in V_{k+1}$, and I' is a *child* of I, if $I' \subset I$. It is easy to check that the resulting structure has the following properties:

- Any two distinct intervals $I, I' \in V_k$ are at least L^{-k} apart.
- $\mu(\mathbb{R} \setminus \bigsqcup_{I \in V_k} I) = 0$ for all k.
- Each $I \in V_k$ has exactly one parent.
- If $I \in V_k$ and $I_1, \dots, I_n \in V_{k+1}$ are the children of I, then

$$0 < \mu(I) = \sum_{j=1}^{n} \mu(I_j). \tag{2-1}$$

For regular sets, the discretization has the following additional properties:

Lemma 2.1. *Let $L \geq 2$, $K > 0$ be integers and assume (X, μ_X) is δ-regular up to scale L^{-K} with regularity constant C_R, where $0 < \delta < 1$. Then the discretization of μ_X with base L has the following properties*:

(1) *Each $I \in V$ with $0 \leq H(I) \leq K$ satisfies, for $C_R' := (3C_R^2)^{\frac{1}{1-\delta}}$,*

$$L^{-H(I)} \leq |I| \leq C_R' L^{-H(I)}, \tag{2-2}$$

$$C_R^{-1} L^{-\delta H(I)} \leq \mu_X(I) \leq C_R (C_R')^{\delta} L^{-\delta H(I)}. \tag{2-3}$$

(2) *If I' is a child of $I \in V$ and $0 \leq H(I) < K$, then*

$$\frac{\mu_X(I')}{\mu_X(I)} \geq \frac{L^{-\delta}}{C_R'}. \tag{2-4}$$

(3) *Assume that*

$$L \geq (4C_R)^{\frac{6}{\delta(1-\delta)}}. \tag{2-5}$$

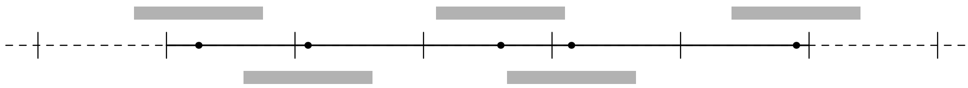

Figure 1. An illustration of the proof of the upper bound in (2-2). The ticks mark points in $L^{-k}\mathbb{Z}$, the solid interval is I, the dots mark the points x_q, and the shaded intervals are I_q. The intervals of length L^{-k} adjacent to I have zero measure μ_X.

Then for each $I \in V$ *with* $0 \le H(I) < K$, *there exist two children* I', I'' *of* I *such that*

$$\tfrac{1}{2} C_R^{-\frac{2}{\delta}} L^{-H(I)-\frac{2}{3}} \le |x' - x''| \le 2L^{-H(I)-\frac{2}{3}} \quad \textit{for all } x' \in I',\ x'' \in I''.$$

Remark. Parts (1) and (2) of the lemma state that the tree of intervals discretizing μ_X is approximately regular. Part (3), which is used at the end of Section 3B, states that once the base of discretization L is large enough, each interval I in the tree has two children which are $\sim L^{-H(I)-\frac{2}{3}}$ apart from each other. A similar statement would hold if $\frac{2}{3}$ were replaced by any number in $(0, 1)$.

Proof. (1) Put $k := H(I)$. The lower bound on $|I|$ follows from the construction of the discretization. To show the upper bound, assume that $I = [c, d]$ and $d - c = ML^{-k}$. For each $q \in L^{-k}\mathbb{Z}$ with $c \le q < d$, we have $\mu_X([q, q + L^{-k}]) > 0$; thus there exists $x_q \in [q, q + L^{-k}] \cap X$. Let I_q be the interval of size L^{-k} centered at x_q; see Figure 1. Then

$$\mu_X\Big(\bigcup_q I_q\Big) \le \mu_X([c - L^{-k}, d + L^{-k}]) = \mu_X(I) \le C_R(ML^{-k})^\delta.$$

On the other hand, each point is covered by at most three intervals I_q; therefore

$$MC_R^{-1}L^{-k\delta} \le \sum_q \mu_X(I_q) \le 3\mu_X\Big(\bigcup_q I_q\Big).$$

Together these two inequalities imply $M \le C'_R$, giving (2-2).

The upper bound on $\mu_X(I)$ follows from (2-2). To show the lower bound, take $x \in I \cap X$ and let I' be the interval of size L^{-k} centered at x. Then $\mu_X(I' \setminus I) = 0$; therefore $\mu_X(I) \ge \mu_X(I') \ge C_R^{-1}L^{-\delta k}$.

(2) This follows directly from (2-3) and the fact that $C_R^2(C'_R)^\delta \le C'_R$.

(3) Put $k := H(I)$. Take $x \in I \cap X$ and let J be the interval of size $L^{-k-\frac{2}{3}}$ centered at x. Let $I_1, \dots, I_n$ be all the intervals in V_{k+1} which intersect J; they all have to be children of I. Let $x_1, \dots, x_n$ be the centers of $I_1, \dots, I_n$. Define

$$T := L^{k+\frac{2}{3}} \max_{j,\ell} |x_j - x_\ell|.$$

By (2-2), we have $|I_j| \le C'_R L^{-k-1}$ and thus $T \le 1 + C'_R L^{-\frac{1}{3}}$. On the other hand, the union of $I_1, \dots, I_n$ is contained in an interval of size $TL^{-k-\frac{2}{3}} + C'_R L^{-k-1}$. Therefore

$$C_R^{-1} L^{-\delta(k+\frac{2}{3})} \le \mu_X(J) \le \sum_{j=1}^n \mu_X(I_j) \le C_R(TL^{-k-\frac{2}{3}} + C'_R L^{-k-1})^\delta.$$

This implies $T \ge C_R^{-\frac{2}{\delta}} - C'_R L^{-\frac{1}{3}}$.

Now, put $I' := I_j$, $I'' = I_\ell$, where j, ℓ are chosen so that $T = L^{k+\frac{2}{3}}|x_j - x_\ell|$. Then for each $x' \in I'$, $x'' \in I''$, we have by (2-5)

$$\tfrac{1}{2}C_R^{-\frac{2}{\delta}} \le C_R^{-\frac{2}{\delta}} - 2C_R' L^{-\frac{1}{3}} \le L^{k+\frac{2}{3}}|x' - x''| \le 1 + 2C_R' L^{-\frac{1}{3}} \le 2. \qquad \square$$

We finally have the following estimates on the Lebesgue measure of neighborhoods of a δ-regular set, which are used in Sections 4–5:

Lemma 2.2. *Assume that (Λ, μ_Λ) is δ-regular up to scale $h \in (0,1)$ with constant C_R. Let $X := \Lambda(h) = \Lambda + [-h, h]$ be the h-neighborhood of Λ and define the measure μ_X by*

$$\mu_X(A) := h^{\delta-1}\mu_L(X \cap A), \quad A \subset \mathbb{R}, \tag{2-6}$$

where μ_L denotes the Lebesgue measure. Then (X, μ_X) is δ-regular up to scale h with constant $C_R' := 30C_R^2$.

Proof. We follow [Dyatlov and Zahl 2016, Lemma 7.4]. Let $I \subset \mathbb{R}$ be an interval with $|I| \ge h$. Let $x_1, \dots, x_N \in \Lambda \cap I(h)$ be a maximal set of $2h$-separated points. Denote by I_n' the interval of size h centered at x_n. Since I_n' are disjoint and their union is contained in $I(2h)$, which is an interval of size $|I| + 4h \le 5|I|$, we have

$$N \cdot C_R^{-1} h^\delta \le \sum_{n=1}^{N} \mu_\Lambda(I_n') \le \mu_\Lambda(I(2h)) \le 5C_R|I|^\delta. \tag{2-7}$$

Next, let I_n be the interval of size $6h$ centered at x_n. Then $X \cap I$ is contained in the union of I_n and thus

$$\mu_L(X \cap I) \le \sum_{n=1}^{N} \mu_L(I_n) = 6hN. \tag{2-8}$$

Together (2-7) and (2-8) give the required upper bound

$$\mu_X(I) = h^{\delta-1}\mu_L(X \cap I) \le 30C_R^2|I|^\delta.$$

Now, assume additionally that $|I| \le 1$ and I is centered at a point in X. Let $y_1, \dots, y_M \in \Lambda \cap I$ be a maximal set of h-separated points. Denote by I_m the interval of size $2h$ centered at y_m. Then $\Lambda \cap I$ is contained in the union of I_m; therefore

$$C_R^{-1}|I|^\delta \le \mu_\Lambda(I) = \mu_\Lambda(\Lambda \cap I) \le \sum_{m=1}^{M} \mu_\Lambda(I_m) \le M \cdot 2C_R h^\delta. \tag{2-9}$$

Next, let I_m' be the interval of size h centered at y_m. Then $I_m' \subset X$ are nonoverlapping and each $I_m' \cap I$ has size at least $\frac{1}{2}h$; therefore

$$\mu_L(X \cap I) \ge \sum_{m=1}^{M} \mu_L(I_m' \cap I) \ge \tfrac{1}{2}Mh. \tag{2-10}$$

Combining (2-9) and (2-10) gives the required lower bound

$$\mu_X(I) = h^{\delta-1}\mu_L(X \cap I) \ge \frac{1}{4C_R^2}|I|^\delta. \qquad \square$$

2B. ***Functional spaces.*** For a constant $\theta > 0$ and an interval I, let $\mathcal{C}_\theta(I)$ be the space $C^1(I)$ with the norm

$$\|f\|_{\mathcal{C}_\theta(I)} := \max\bigl(\sup_I |f|, \theta|I|\cdot\sup_I |f'|\bigr).$$

The following lemma shows that multiplications by functions of the form $\exp(i\psi)$ have norm 1 when mapping $\mathcal{C}_\theta(I)$ into the corresponding space for a sufficiently small subinterval of I:

Lemma 2.3. *Consider intervals*

$$I' \subset I, \quad |I'| \le \tfrac{1}{4}|I|. \tag{2-11}$$

Assume that $\psi \in C^\infty(I;\mathbb{R})$ *and* $\theta > 0$ *are such that*

$$4\theta|I'|\cdot\sup_{I'}|\psi'| \le 1. \tag{2-12}$$

Then for each $f \in \mathcal{C}_\theta(I)$, *we have* $\|\exp(i\psi)f\|_{\mathcal{C}_\theta(I')} \le \|f\|_{\mathcal{C}_\theta(I)}$ *and*

$$\theta|I'|\cdot\sup_{I'}|(\exp(i\psi)f)'| \le \tfrac{1}{2}\|f\|_{\mathcal{C}_\theta(I)}. \tag{2-13}$$

Proof. The left-hand side of (2-13) is bounded from above by

$$\theta|I'|\cdot\bigl(\sup_{I'}|\psi' f| + \sup_{I'}|f'|\bigr).$$

From (2-12), (2-11) we get

$$\theta|I'|\cdot\sup_{I'}|\psi' f| \le \tfrac{1}{4}\|f\|_{\mathcal{C}_\theta(I)}, \quad \theta|I'|\cdot\sup_{I'}|f'| \le \tfrac{1}{4}\|f\|_{\mathcal{C}_\theta(I)},$$

which finishes the proof of (2-13). The bound (2-13) implies $\|\exp(i\psi)f\|_{\mathcal{C}_\theta(I')} \le \|f\|_{\mathcal{C}_\theta(I)}$. □

The following is a direct consequence of the mean value theorem:

Lemma 2.4. *Let* $f \in \mathcal{C}_\theta(I)$. *Then for all* $x, x' \in I$, *we have*

$$|f(x) - f(x')| \le \frac{|x - x'|}{\theta|I|}\cdot\|f\|_{\mathcal{C}_\theta(I)}. \tag{2-14}$$

2C. ***A few technical lemmas.*** The following is a two-dimensional analog of the mean value theorem:

Lemma 2.5. *Let* $I = [c_1, d_1]$ *and* $J = [c_2, d_2]$ *be two intervals and* $\Phi \in C^2(I \times J;\mathbb{R})$. *Then there exists* $(x_0, y_0) \in I \times J$ *such that*

$$\Phi(c_1, c_2) + \Phi(d_1, d_2) - \Phi(c_1, d_2) - \Phi(d_1, c_2) = |I|\cdot|J|\cdot\partial^2_{xy}\Phi(x_0, y_0).$$

Proof. Replacing $\Phi(x, y)$ by $\Phi(x, y) - \Phi(c_1, y) - \Phi(x, c_2) + \Phi(c_1, c_2)$, we may assume that $\Phi(c_1, y) = 0$ and $\Phi(x, c_2) = 0$ for all $x \in I$, $y \in J$. By the mean value theorem, we have $\Phi(d_1, d_2) = |I|\cdot\partial_x\Phi(x_0, d_2)$ for some $x_0 \in I$. Applying the mean value theorem again, we have $\partial_x\Phi(x_0, d_2) = |J|\cdot\partial^2_{xy}\Phi(x_0, y_0)$ for some $y_0 \in J$, finishing the proof. □

Lemma 2.6. *Assume that* $\tau \in \mathbb{R}$ *and* $|\tau| \le \pi$. *Then* $|e^{i\tau} - 1| \ge \frac{2}{\pi}|\tau|$.

Proof. We have

$$|e^{i\tau}-1| = 2\sin\big(\tfrac{1}{2}|\tau|\big).$$

It remains to use that $\sin x \ge \frac{2}{\pi}x$ when $0\le x\le \frac{\pi}{2}$, which follows from the concavity of $\sin x$ on that interval. □

The next lemma, used several times in Section 3B, is a quantitative version of the fact that balls in Hilbert spaces are strictly convex:

Lemma 2.7. *Assume that $\mathcal{H}$ is a Hilbert space, $f_1,\dots,f_n\in\mathcal{H}$, $p_1,\dots,p_n\ge 0$, and $p_1+\dots+p_n=1$. Then*

$$\Big\|\sum_{j=1}^n p_j f_j\Big\|_{\mathcal{H}}^2 = \sum_{j=1}^n p_j\|f_j\|_{\mathcal{H}}^2 - \sum_{1\le j<\ell\le n} p_j p_\ell \|f_j - f_\ell\|_{\mathcal{H}}^2. \tag{2-15}$$

If moreover for some $\varepsilon, R\ge 0$

$$\sum_{j=1}^n p_j\|f_j\|_{\mathcal{H}}^2 = R,\qquad \Big\|\sum_{j=1}^n p_j f_j\Big\|_{\mathcal{H}}^2 \ge (1-\varepsilon)R,\qquad p_{\min} := \min_j p_j \ge 2\sqrt{\varepsilon} \tag{2-16}$$

then for all j

$$\tfrac{1}{2}\sqrt{R}\le \|f_j\|_{\mathcal{H}} \le 2\sqrt{R}. \tag{2-17}$$

Proof. The identity (2-15) follows by a direct computation. To show (2-17), note that by (2-15) and (2-16) for each j,ℓ

$$\|f_j-f_\ell\|_{\mathcal{H}}^2 \le \frac{\varepsilon R}{p_{\min}^2}\le \tfrac{1}{4}R.$$

Put

$$f_{\max} := \max_j \|f_j\|_{\mathcal{H}},\qquad f_{\min} := \min_j \|f_j\|_{\mathcal{H}}.$$

Then

$$f_{\max}-f_{\min}\le \tfrac{1}{2}\sqrt{R},\qquad f_{\min}\le\sqrt{R}\le f_{\max},$$

which implies (2-17). □

Lemma 2.8. *Assume that $\alpha_j, p_j\ge 0$, $j=1,\dots,n$, $p_1+\dots+p_n=1$, and for some $\varepsilon, R\ge 0$*

$$\sum_{j=1}^n p_j\alpha_j \ge (1-\varepsilon)R,\qquad \max_j \alpha_j\le R,\qquad p_{\min} := \min_j p_j\ge 2\varepsilon.$$

Then for all j,

$$\alpha_j\ge \tfrac{1}{2}R.$$

Proof. We have

$$\sum_{j=1}^n p_j(R-\alpha_j)\le \varepsilon R.$$

All the terms in the sum are nonnegative; therefore for all j

$$R-\alpha_j \le \frac{\varepsilon R}{p_{\min}}\le \tfrac{1}{2}R.$$

□

3. Proof of Theorem 1

3A. ***The iterative argument.*** In this section, we prove the following statement which can be viewed as a special case of Theorem 1. Its proof relies on an inductive bound, Lemma 3.2, which is proved in Section 3B. In Section 3C, we deduce Theorem 1 from Proposition 3.1, in particular removing the condition (3-1).

Proposition 3.1. *Let $\delta, \delta' \in (0,1)$, $C_R > 1$, $I_0, J_0 \subset \mathbb{R}$ be some intervals, $G \in C^1(I_0 \times J_0; \mathbb{C})$, and the phase function $\Phi \in C^2(I_0 \times J_0; \mathbb{R})$ satisfy*

$$\tfrac{1}{2} < |\partial^2_{xy}\Phi(x,y)| < 2 \quad \textit{for all } (x,y) \in I_0 \times J_0. \tag{3-1}$$

Choose constants $C'_R > 0$ and $L \in \mathbb{N}$ such that

$$C'_R = (2C_R)^{\frac{2}{1-\max(\delta,\delta')}}, \qquad L \geq (2C'_R(6C_R)^{\frac{1}{\delta}+\frac{1}{\delta'}})^6. \tag{3-2}$$

Fix $K_0 \in \mathbb{N}_0$ and put $h := L^{-K}$ for some $K \in \mathbb{N}_0$, $K \geq 2K_0$. Assume that (X, μ_X) is δ-regular, and (Y, μ_Y) is δ'-regular, up to scale L^{K_0-K} with regularity constant C_R, and $X \subset I_0$, $Y \subset J_0$. Put

$$\varepsilon_1 := 10^{-5}(C_R^{\frac{1}{\delta}+\frac{1}{\delta'}} C'_R)^{-4} L^{-5}, \qquad \varepsilon_0 := -\frac{\log(1-\varepsilon_1)}{2\log L}. \tag{3-3}$$

Then for some C depending only on $K_0, G, \mu_X(X), \mu_Y(Y)$, and $\mathcal{B}_h$ defined in (1-3),

$$\|\mathcal{B}_h f\|_{L^2(X,\mu_X)} \leq C h^{\varepsilon_0} \|f\|_{L^2(Y,\mu_Y)} \quad \textit{for all } f \in L^2(Y,\mu_Y). \tag{3-4}$$

Remark. Proposition 3.1 has complicated hypotheses in order to make it useful for the proof of Theorem 1. However, the argument is essentially the same in the following special case which could simplify the reading of the proof below: $\delta = \delta'$, $G \equiv 1$, $\Phi(x,y) = xy$, $K_0 = 0$. Note that in this case $\mathcal{B}_h$ is related to the semiclassical Fourier transform (1-2).

To start the proof of Proposition 3.1, we extend Φ to a function in $C^2(\mathbb{R}^2; \mathbb{R})$ such that (3-1) still holds, and extend G to a function in $C^1(\mathbb{R}^2; \mathbb{C})$ such that $G, \partial_x G$ are uniformly bounded. Following Section 2A, consider the discretizations of μ_X, μ_Y with base L, denoting by V_X, V_Y the sets of vertices and by H the height functions.

Fix $f \in L^2(Y, \mu_Y)$. For each $J \in V_Y$, let y_J denote the center of J and define the function of $x \in \mathbb{R}$

$$F_J(x) = \frac{1}{\mu_Y(J)} \int_J \exp\left(\frac{i(\Phi(x,y) - \Phi(x,y_J))}{h}\right) G(x,y) f(y)\, d\mu_Y(y). \tag{3-5}$$

In terms of the operator $\mathcal{B}_h$ from (1-3), we may write

$$F_J(x) = \frac{1}{\mu_Y(J)} \exp\left(-\frac{i\Phi(x,y_J)}{h}\right) \mathcal{B}_h(\mathbb{1}_J f)(x). \tag{3-6}$$

Put

$$\theta := \frac{1}{8(C'_R)^2} \tag{3-7}$$

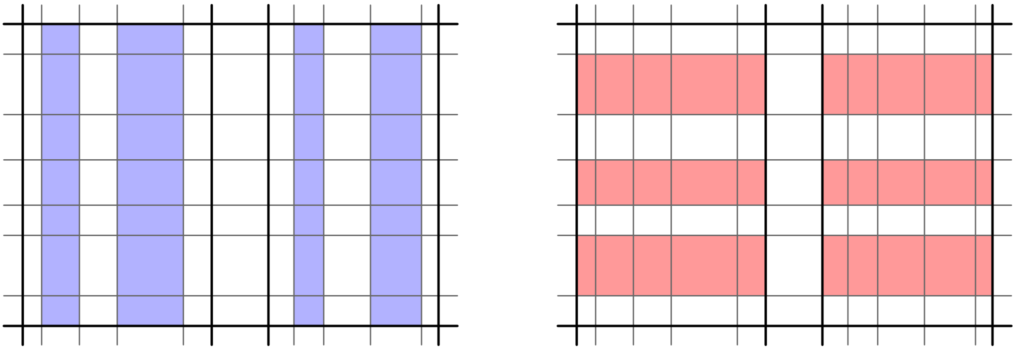

Figure 2. An illustration of (3-8) in the case $K = 1$. The vertical lines mark the endpoints of intervals in V_X and the horizontal lines mark the endpoints of intervals in V_Y. The thick lines correspond to intervals of height 0 and the thin lines to intervals of height 1. The shaded rectangles have the form $I \times J$, $I \in V_X$, $J \in V_Y$, where E_J is constant on I, and the shaded rectangles on the left/right correspond to the left-/right-hand sides of (3-9) for $H(J) = 0$.

and for $J \in V_Y$ define the piecewise constant function $E_J \in L^\infty(X, \mu_X)$ using the space $\mathcal{C}_\theta(I)$ defined in Section 2B:

$$E_J(x) = \|F_J\|_{\mathcal{C}_\theta(I)}, \quad \text{where } x \in I \in V_X,\ H(I) + H(J) = K. \tag{3-8}$$

See Figure 2. Note that $|F_J(x)| \le E_J(x)$ for μ_X-almost every x.

The L^2 norms of the functions E_J satisfy the following key bound, proved in Section 3B, which gives an improvement from one scale to the next. The use of the L^2 norm of E_J as the monotone quantity is convenient for several reasons. On one hand, the averaging provided by the L^2 norm means it is only necessary to show an improvement on F_J in sufficiently many places; more precisely we will show in (3-11) that such improvement happens on at least one child of each interval $I \in V(X)$ with $H(I) + H(J) = K - 1$. On the other hand, such improvement is obtained by a pointwise argument which also uses that the F_{J_b} are slowly varying on each interval I with $H(I) + H(J) = K - 1$ (see Lemma 3.7); this motivates the use of $\mathcal{C}_\theta(I)$ norms in the definition of E_{J_b}.

Lemma 3.2. *Let $J \in V_Y$ with $K_0 \le H(J) < K - K_0$ and $J_1, \dots, J_B \in V_Y$ be the children of J. Then, with ε_1 defined in (3-3),*

$$\|E_J\|^2_{L^2(X,\mu_X)} \le (1 - \varepsilon_1) \sum_{b=1}^{B} \frac{\mu_Y(J_b)}{\mu_Y(J)} \|E_{J_b}\|^2_{L^2(X,\mu_X)}. \tag{3-9}$$

Iterating Lemma 3.2, we obtain:

Proof of Proposition 3.1. First of all, we show that for all $J \in V_Y$ with $H(J) = K - K_0$, and some constant C_0 depending on G, $\mu_X(X)$ and defined below, we have

$$\|E_J\|^2_{L^2(X,\mu_X)} \le C_0 \frac{\|f\|^2_{L^2(J,\mu_Y)}}{\mu_Y(J)}. \tag{3-10}$$

Indeed, take $I\in V_X$ such that $H(I)=K_0$. By (2-2) and (3-1), for all $y\in J$

$$\frac{1}{h}\sup_{x\in I}|\partial_x\Phi(x,y)-\partial_x\Phi(x,y_J)|\le\frac{2}{h}|J|\le 2C_R'L^{K_0}$$

and thus by (2-2) and (3-7)

$$\frac{4\theta|I|}{h}\sup_{x\in I}|\partial_x\Phi(x,y)-\partial_x\Phi(x,y_J)|\le 1.$$

Arguing similarly to Lemma 2.3, we obtain for all $y\in J$

$$\left\|\exp\left(\frac{i(\Phi(x,y)-\Phi(x,y_J))}{h}\right)G(x,y)\right\|_{\mathcal{C}_\theta(I)}\le C_G:=\max(\sup|G|,\sup|\partial_x G|).$$

Using Hölder's inequality in (3-5), we obtain

$$E_J|_I=\|F_J\|_{\mathcal{C}_\theta(I)}\le\frac{C_G}{\mu_Y(J)}\int_J|f(y)|\,d\mu_Y(y)\le\frac{C_G\|f\|_{L^2(J,\mu_Y)}}{\sqrt{\mu_Y(J)}}$$

and (3-10) follows by integration in x, where we put $C_0:=C_G^2\mu_X(X)$.

Now, arguing by induction on $H(J)$ with (3-10) as the base case and (3-9) as the inductive step, we obtain for all $J\in V_Y$ with $K_0\le H(J)\le K-K_0$,

$$\|E_J\|^2_{L^2(X,\mu_X)}\le C_0(1-\varepsilon_1)^{K-K_0-H(J)}\frac{\|f\|^2_{L^2(J,\mu_Y)}}{\mu_Y(J)}.$$

In particular, for all $J\in V_Y$ with $H(J)=K_0$, we have by (3-6)

$$\left\|\frac{\mathcal{B}_h(\mathbb{1}_J f)}{\mu_Y(J)}\right\|^2_{L^2(X,\mu_X)}=\|F_J\|^2_{L^2(X,\mu_X)}\le\|E_J\|^2_{L^2(X,\mu_X)}\le C_1h^{2\varepsilon_0}\frac{\|f\|^2_{L^2(J,\mu_Y)}}{\mu_Y(J)},$$

where $C_1:=C_0(1-\varepsilon_1)^{-2K_0}$. Using the identity

$$\mathcal{B}_h f=\mu_Y(Y)\sum_{\substack{J\in V_Y\\ H(J)=K_0}}\frac{\mu_Y(J)}{\mu_Y(Y)}\cdot\frac{\mathcal{B}_h(\mathbb{1}_J f)}{\mu_Y(J)}$$

and (2-15), we estimate

$$\|\mathcal{B}_h f\|^2_{L^2(X,\mu_X)}\le C_1\mu_Y(Y)h^{2\varepsilon_0}\|f\|^2_{L^2(Y,\mu_Y)}$$

and (3-4) follows with $C:=C_G(1-\varepsilon_1)^{-K_0}\sqrt{\mu_X(X)\mu_Y(Y)}$. ☐

3B. *The inductive step.* In this section we prove Lemma 3.2. Let $J\in V_Y$ satisfy $K_0\le H(J)<K-K_0$ and $J_1,\dots,J_B$ be the children of J. It suffices to show that for all $I\in V_X$ with $H(I)+H(J)=K-1$ we have

$$\|E_J\|^2_{L^2(I,\mu_X)}\le(1-\varepsilon_1)\sum_{b=1}^{B}\frac{\mu_Y(J_b)}{\mu_Y(J)}\|E_{J_b}\|^2_{L^2(I,\mu_X)}. \tag{3-11}$$

Indeed, summing (3-11) over I, we obtain (3-9).

Fix $I \in V_X$ with $H(I) + H(J) = K - 1$ and let $I_1, \dots, I_A$ be the children of I. Define

$$p_a := \frac{\mu_X(I_a)}{\mu_X(I)}, \quad q_b := \frac{\mu_Y(J_b)}{\mu_Y(J)}.$$

Note that $p_a, q_b \geq 0$ and $p_1 + \cdots + p_A = q_1 + \cdots + q_B = 1$.

The functions F_J and F_{J_b} are related by the following formula:

$$F_J = \sum_{b=1}^{B} q_b \exp(i\Psi_b) F_{J_b}, \quad \Psi_b(x) := \frac{\Phi(x, y_{J_b}) - \Phi(x, y_J)}{h}. \tag{3-12}$$

That is, F_J is a convex combination of $F_{J_1}, \dots, F_{J_B}$ multiplied by the phase factors $\exp(i\Psi_b)$. At the end of this subsection we exploit cancellation between these phase factors to show (3-11). However there are several preparatory steps necessary. Before we proceed with the proof, we show the version of (3-11) with no improvement:

Lemma 3.3. *We have*

$$\|E_J\|^2_{L^2(I,\mu_X)} \leq \sum_{b=1}^{B} q_b \|E_{J_b}\|^2_{L^2(I,\mu_X)}. \tag{3-13}$$

Proof. By (2-2), (3-1), and (3-7), we have for all a, b

$$4\theta|I_a| \cdot \sup_I |\Psi_b'| \leq \frac{8\theta|I_a| \cdot |J|}{h} \leq 1. \tag{3-14}$$

Moreover, by (2-2) and (3-2) we have $|I_a| \leq \frac{1}{4}|I|$. Applying Lemma 2.3, we obtain

$$\|\exp(i\Psi_b)F_{J_b}\|_{C_\theta(I_a)} \leq \|F_{J_b}\|_{C_\theta(I)}.$$

By (3-12) and (2-15) we then have

$$\|F_J\|^2_{C_\theta(I_a)} \leq \left(\sum_{b=1}^{B} q_b \|F_{J_b}\|_{C_\theta(I)}\right)^2 \leq \sum_{b=1}^{B} q_b \|F_{J_b}\|^2_{C_\theta(I)}. \tag{3-15}$$

By (3-8), we have for all a, b

$$E_J|_{I_a} = \|F_J\|_{C_\theta(I_a)}, \quad E_{J_b}|_I = \|F_{J_b}\|_{C_\theta(I)}. \tag{3-16}$$

Now, summing both sides of (3-15) over a with weights $\mu_X(I_a)$, we obtain (3-13). ☐

The rest of this section is dedicated to the proof of (3-11), studying the situations in which the bound (3-13) is almost sharp and ultimately reaching a contradiction. The argument is similar in spirit to Lemma 1.2. In fact we can view Lemma 1.2 as the special degenerate case when $A = B = 2$, $p_a = q_b = \frac{1}{2}$, the intervals I_a are replaced by points x_a, $F_{J_b} \equiv f_b$ are constants, $u_a = F_J(x_a)$, and $\omega_{ab} = \Psi_b(x_a)$. The general case is more technically complicated. In particular we use Lemma 2.7 to deal with general convex combinations. We also use δ-regularity in many places, for instance to show that the coefficients p_a, q_b are bounded away from zero and to get the phase factor cancellations in (3-30) at the end of the proof. The reading of the argument below may be simplified by making the illegal choice $\varepsilon_1 := 0$.

We henceforth assume that (3-11) does not hold. Put

$$R := \sum_{b=1}^{B} q_b \|F_{J_b}\|^2_{\mathcal{C}_\theta(I)}. \tag{3-17}$$

By (3-16), the failure of (3-11) can be rewritten as

$$\sum_{a=1}^{A} p_a \|F_J\|^2_{\mathcal{C}_\theta(I_a)} > (1-\varepsilon_1)R. \tag{3-18}$$

We note for future use that p_a, q_b are bounded below by (2-4):

$$p_{\min} := \min_a p_a \ge \frac{L^{-\delta}}{C'_R}, \quad q_{\min} := \min_b q_b \ge \frac{L^{-\delta'}}{C'_R}. \tag{3-19}$$

We first deduce from (3-17) and the smallness of ε_1 an upper bound on each $\|F_{J_b}\|_{\mathcal{C}_\theta(I)}$ in terms of the averaged quantity R:

Lemma 3.4. *We have for all b,*

$$\|F_{J_b}\|_{\mathcal{C}_\theta(I)} \le 2\sqrt{R}. \tag{3-20}$$

Proof. The first inequality in (3-15) together with (3-18) implies

$$\left(\sum_{b=1}^{B} q_b \|F_{J_b}\|_{\mathcal{C}_\theta(I)}\right)^2 \ge \sum_{a=1}^{A} p_a \|F_J\|^2_{\mathcal{C}_\theta(I_a)} \ge (1-\varepsilon_1)R. \tag{3-21}$$

By (3-3) and (3-19) we have $q_{\min} \ge 2\sqrt{\varepsilon_1}$. Applying (2-17) to $f_b := \|F_{J_b}\|_{\mathcal{C}_\theta(I)}$ with (3-17) and (3-21), we obtain (3-20). ☐

We next obtain a version of (3-18) which gives a lower bound on the size of F_J, rather than on the norm $\|F_J\|_{\mathcal{C}_\theta(I_a)}$:

Lemma 3.5. *There exist $x_a \in I_a$, $a = 1, \dots, A$, such that*

$$\sum_{a=1}^{A} p_a |F_J(x_a)|^2 > (1-2\varepsilon_1)R. \tag{3-22}$$

Proof. By Lemma 2.3 and (3-14), we have

$$\theta |I_a| \cdot \sup_{I_a} |(\exp(i\Psi_b)F_{J_b})'| \le \tfrac{1}{2}\|F_{J_b}\|_{\mathcal{C}_\theta(I)}.$$

It follows by (3-12) and the triangle inequality that for all a,

$$\|F_J\|_{\mathcal{C}_\theta(I_a)} \le \max\left(\sup_{I_a} |F_J|, \frac{1}{2}\sum_{b=1}^{B} q_b \|F_{J_b}\|_{\mathcal{C}_\theta(I)}\right). \tag{3-23}$$

By (3-15) we have

$$\sup_{I_a} |F_J|^2 \le \|F_J\|^2_{\mathcal{C}_\theta(I_a)} \le R. \tag{3-24}$$

Therefore by (3-23) and the second inequality in (3-15)

$$\|F_J\|^2_{\mathcal{C}_\theta(I_a)} \le \tfrac{1}{2}\big(R + \sup_{I_a} |F_J|^2\big).$$

Summing this inequality over a with weights p_a, we see that (3-18) implies

$$\sum_{a=1}^{A} p_a \sup_{I_a} |F_J|^2 > (1-2\varepsilon_1)R,$$

which gives (3-22). ☐

Now, choose x_a as in Lemma 3.5 and put

$$F_{ab} := F_{J_b}(x_a) \in \mathbb{C}, \quad \omega_{ab} := \Psi_b(x_a) \in \mathbb{R}.$$

Note that by (3-12)

$$F_J(x_a) = \sum_{b=1}^{B} q_b \exp(i\omega_{ab}) F_{ab}.$$

Using (2-15) for $f_b = \exp(i\omega_{ab})F_{ab}$ and (3-22), we obtain

$$\sum_{a,b} p_a q_b |F_{ab}|^2 > (1-2\varepsilon_1)R + \sum_{\substack{a,b,b' \\ b<b'}} p_a q_b q_{b'} \big|\exp(i(\omega_{ab}-\omega_{ab'}))F_{ab} - F_{ab'}\big|^2. \tag{3-25}$$

From the definition (3-17) of R, we have for all a

$$\sum_{b=1}^{B} q_b |F_{ab}|^2 \le R. \tag{3-26}$$

Therefore, the left-hand side of (3-25) is bounded above by R. Using (3-19), we then get for all a, b, b' the following approximate equality featuring the phase terms ω_{ab}:

$$\begin{aligned}\big|\exp(i(\omega_{ab}-\omega_{ab'}))F_{ab} - F_{ab'}\big| &< \sqrt{\frac{2\varepsilon_1 R}{p_{\min} q_{\min}^2}} \\ &\le 2(C_R')^2 L^{\delta+\delta'} \sqrt{\varepsilon_1 R}.\end{aligned} \tag{3-27}$$

Using the smallness of ε_1, we obtain from here a lower bound on $|F_{ab}|$:

Lemma 3.6. *For all a, b we have*

$$|F_{ab}| \ge \tfrac{1}{2}\sqrt{R}. \tag{3-28}$$

Proof. By (3-3) and (3-19), we have $p_{\min} \ge 4\varepsilon_1$. Applying Lemma 2.8 to $\alpha_a = \sum_b q_b |F_{ab}|^2$ and using (3-25) and (3-26), we obtain for all a

$$\sum_{b=1}^{B} q_b |F_{ab}|^2 \ge \tfrac{1}{2} R. \tag{3-29}$$

We now argue similarly to the proof of (2-17). Fix a and let $F_{a,\min} = \min_b |F_{ab}|$, $F_{a,\max} = \max_b |F_{ab}|$. By (3-29) we have $F_{a,\max} \ge \sqrt{R/2}$. On the other hand the difference $F_{a,\max} - F_{a,\min}$ is bounded above by (3-27). By (3-3) we then have

$$F_{a,\min} \ge \sqrt{\tfrac{1}{2}R} - 2(C_R')^2 L^{\delta+\delta'}\sqrt{\varepsilon_1 R} \ge \tfrac{1}{2}\sqrt{R}. \qquad \square$$

We next estimate the discrepancy between the values F_{ab} for fixed b and different a, using the fact that we control the norm $\|F_{J_b}\|_{\mathcal{C}_\theta(I)}$ and thus the derivative of F_{J_b}:

Lemma 3.7. *For all a, a', b we have*

$$|F_{ab} - F_{a'b}| \le \frac{2\sqrt{R}\,|x_a - x_{a'}|}{\theta |I|} \le \frac{2\sqrt{R}}{\theta} L^{H(I)} \cdot |x_a - x_{a'}|.$$

Proof. This follows immediately by combining Lemma 2.4, Lemma 3.4, and (2-2). $\square$

Armed with the bounds obtained above, we are now ready to reach a contradiction and finish the proof of Lemma 3.2, using the discrepancy of the phase shifts ω_{ab} and the lower bound on $|\partial^2_{xy}\Phi|$ from (3-1).

Using part (3) of Lemma 2.1 and (3-2), choose a, a', b, b' such that

$$\tfrac{1}{2} C_R^{-\frac{2}{\delta}} L^{-\frac{2}{3}} \le L^{H(I)} \cdot |x_a - x_{a'}| \le 2L^{-\frac{2}{3}},$$
$$\tfrac{1}{2} C_R^{-\frac{2}{\delta'}} L^{-\frac{2}{3}} \le L^{H(J)} \cdot |y_b - y_{b'}| \le 2L^{-\frac{2}{3}}.$$

Recall that $x_a \in I_a$ is chosen in Lemma 3.5 and $y_b := y_{J_b}$ is the center of J_b. By Lemma 2.5, we have for some $(\tilde{x}, \tilde{y}) \in I \times J$,

$$\tau := \omega_{ab} + \omega_{a'b'} - \omega_{a'b} - \omega_{ab'} = \frac{(x_a - x_{a'})(y_b - y_{b'})}{h}\, \partial^2_{xy}\Phi(\tilde{x}, \tilde{y}).$$

By (3-1) and (3-2) and since $h = L^{-K}$, $H(I) + H(J) = K - 1$, we have

$$\tfrac{1}{8} C_R^{-\frac{2}{\delta}-\frac{2}{\delta'}} L^{-\frac{1}{3}} \le |\tau| \le 8L^{-\frac{1}{3}} \le \pi.$$

Therefore, by Lemma 2.6 the phase factor $e^{i\tau}$ is bounded away from 1, which combined with (3-28) gives a lower bound on the discrepancy:

$$|F_{ab}| \cdot |e^{i\tau} - 1| \ge \frac{|\tau|\sqrt{R}}{\pi} \ge \frac{C_R^{-\frac{2}{\delta}-\frac{2}{\delta'}}}{8\pi} L^{-\frac{1}{3}}\sqrt{R}. \tag{3-30}$$

On the other hand we can estimate the same discrepancy from above by (3-27), Lemma 3.7, and the triangle inequality:

$$\begin{aligned}|F_{ab}| \cdot |e^{i\tau} - 1| &= |e^{i(\omega_{ab}-\omega_{ab'})} F_{ab} - e^{i(\omega_{a'b}-\omega_{a'b'})} F_{ab}| \\ &\le |e^{i(\omega_{ab}-\omega_{ab'})} F_{ab} - F_{ab'}| + |F_{ab'} - F_{a'b'}| + |e^{i(\omega_{a'b}-\omega_{a'b'})} F_{a'b} - F_{a'b'}| + |F_{ab} - F_{a'b}| \\ &< 4(C_R')^2 L^{\delta+\delta'}\sqrt{\varepsilon_1 R} + 8\theta^{-1} L^{-\frac{2}{3}}\sqrt{R}.\end{aligned}$$

Comparing this with (3-30) and dividing by $\sqrt{R}$, we obtain

$$\frac{C_R^{-\frac{2}{\delta}-\frac{2}{\delta'}}}{8\pi}L^{-\frac{1}{3}} < 4(C_R')^2L^{\delta+\delta'}\sqrt{\varepsilon_1}+8\theta^{-1}L^{-\frac{2}{3}}.$$

This gives a contradiction with the following consequences of (3-2) and (3-3):

$$8\theta^{-1}L^{-\frac{2}{3}} \le \frac{C_R^{-\frac{2}{\delta}-\frac{2}{\delta'}}}{16\pi}L^{-\frac{1}{3}}, \quad 4(C_R')^2L^{\delta+\delta'}\sqrt{\varepsilon_1} \le \frac{C_R^{-\frac{2}{\delta}-\frac{2}{\delta'}}}{16\pi}L^{-\frac{1}{3}}.$$

3C. *Proof of Theorem 1.* We now show how to reduce Theorem 1 to Proposition 3.1. The idea is to split G into pieces using a partition of unity. On each piece, by appropriate rescaling we keep the regularity constant C_R and reduce to the case (3-1) and $h = L^{-K}$ for some fixed L satisfying (3-2) and some integer $K > 0$.

To be more precise, let (X, μ_X), (Y, μ_Y), δ, δ', I_0, J_0, Φ, G satisfy the hypotheses of Theorem 1. Using a partition of unity, we write G as a finite sum

$$G = \sum_\ell G_\ell, \quad G_\ell \in C^1(I_0 \times J_0; \mathbb{C}),\ \operatorname{supp} G_\ell \subset I_\ell \times J_\ell, \tag{3-31}$$

where $I_\ell \subset I_0$, $J_\ell \subset J_0$ are intervals such that for some $m = m(\ell) \in \mathbb{Z}$,

$$2^{m-1} < |\partial^2_{xy}\Phi| < 2^{m+1} \quad \text{on } I_\ell \times J_\ell.$$

It then suffices to show (1-4), where G is replaced by one of the functions G_ℓ. By changing Φ outside of the support of G (which does not change the operator $\mathcal{B}_h$), we then reduce to the case when

$$2^{m-1} < |\partial^2_{xy}\Phi| < 2^{m+1} \quad \text{on } I_0 \times J_0 \tag{3-32}$$

for some $m \in \mathbb{Z}$.

We next rescale $\mathcal{B}_h$ to an operator $\widetilde{\mathcal{B}}_{\tilde h}$ satisfying the hypotheses of Proposition 3.1. Fix the smallest $L \in \mathbb{Z}$ satisfying (3-2). Choose $K \in \mathbb{Z}$ and $\sigma \in [1, \sqrt{L})$ such that

$$\sigma^2 = 2^m\frac{\tilde h}{h}, \quad \tilde h := L^{-K}. \tag{3-33}$$

Put for all intervals I, J

$$\begin{aligned}\widetilde X &:= \sigma X \subset \widetilde I_0 := \sigma I_0, \quad \mu_{\widetilde X}(\sigma I) := \sigma^\delta \mu_X(I),\\ \widetilde Y &:= \sigma Y \subset \widetilde J_0 := \sigma J_0, \quad \mu_{\widetilde Y}(\sigma J) := \sigma^{\delta'} \mu_Y(J).\end{aligned}$$

Then $(\widetilde X, \mu_{\widetilde X})$ is δ-regular, and $(\widetilde Y, \mu_{\widetilde Y})$ is δ'-regular, up to scale σh with regularity constant C_R. Consider the unitary operators

$$\begin{aligned}U_X &: L^2(X, \mu_X) \to L^2(\widetilde X, \mu_{\widetilde X}), \quad U_X f(\tilde x) = \sigma^{-\frac{\delta}{2}} f(\sigma^{-1}\tilde x),\\ U_Y &: L^2(Y, \mu_Y) \to L^2(\widetilde Y, \mu_{\widetilde Y}), \quad U_Y f(\tilde y) = \sigma^{-\frac{\delta'}{2}} f(\sigma^{-1}\tilde y).\end{aligned}$$

Then the operator $\widetilde{\mathcal{B}}_{\tilde h} := U_X \mathcal{B}_h U_Y^{-1} : L^2(\widetilde{Y}, \mu_{\widetilde{Y}}) \to L^2(\widetilde{X}, \mu_{\widetilde{X}})$ has the form (1-3):

$$\widetilde{\mathcal{B}}_{\tilde h} f(\tilde x) = \int_{\widetilde{Y}} \exp\Big(\frac{i\widetilde{\Phi}(\tilde x, \tilde y)}{\tilde h}\Big)\widetilde{G}(\tilde x, \tilde y) f(\tilde y)\, d\mu_{\widetilde{Y}}(\tilde y),$$

where

$$\widetilde{\Phi}(\tilde x, \tilde y) = 2^{-m}\sigma^2 \Phi(\sigma^{-1}\tilde x, \sigma^{-1}\tilde y), \quad \widetilde{G}(\tilde x, \tilde y) = \sigma^{-\frac{\delta}{2}-\frac{\delta'}{2}} G(\sigma^{-1}\tilde x, \sigma^{-1}\tilde y).$$

By (3-32) the function $\widetilde{\Phi}$ satisfies (3-1). Fix smallest $K_0 \in \mathbb{N}_0$ such that $\sigma h \le L^{K_0 - K}$, that is,

$$L^{K_0} \ge \frac{2^m}{\sigma}.$$

Without loss of generality, we may assume that h is small enough depending on L, m so that $K \ge 2K_0$. Then Proposition 3.1 applies to $\widetilde{\mathcal{B}}_{\tilde h}$ and gives

$$\|\mathcal{B}_h\|_{L^2(Y,\mu_Y)\to L^2(X,\mu_X)} = \|\widetilde{\mathcal{B}}_{\tilde h}\|_{L^2(\widetilde{Y},\mu_{\widetilde{Y}})\to L^2(\widetilde{X},\mu_{\widetilde{X}})} \le C\tilde h^{\varepsilon_0} \le C(2^{-m}L)^{\varepsilon_0} h^{\varepsilon_0}$$

for ε_0 defined in (1-5) and some constant C depending only on $\delta, \delta', C_R, I_0, J_0, \Phi, G$. This finishes the proof of Theorem 1.

4. Application: spectral gap for hyperbolic surfaces

We now discuss applications of Theorem 1 to spectral gaps. We start with the case of hyperbolic surfaces, referring the reader to [Borthwick 2016; Dyatlov and Zahl 2016] for the terminology used here.

Let $M = \Gamma\backslash\mathbb{H}^2$ be a convex cocompact hyperbolic surface, $\Lambda_\Gamma \subset \mathbb{S}^1$ be its limit set, $\delta \in [0, 1)$ be the dimension of Λ_Γ, and μ be the Patterson–Sullivan measure, which is a probability measure supported on Λ_Γ; see for instance [Borthwick 2016, Section 14.1]. Since Λ_Γ is closed and is not equal to the entire $\mathbb{S}^1$, we may cut the circle $\mathbb{S}^1$ to turn it into an interval and treat Λ_Γ as a compact subset of $\mathbb{R}$. Then (Λ_Γ, μ) is δ-regular up to scale 0 with some constant C_R; see for instance [Borthwick 2016, Lemma 14.13]. The regularity constant C_R depends continuously on the surface, as explained in the case of three-funnel surfaces in [Dyatlov and Zahl 2016, Proposition 7.7].

The main result of this section is the following essential spectral gap for M. We formulate it here in terms of the scattering resolvent of the Laplacian. Another formulation is in terms of a zero-free region for the Selberg zeta function past the first pole; see for instance [Dyatlov and Zahl 2016]. See below for a discussion of previous work on spectral gaps.

Theorem 2. *Consider the meromorphic scattering resolvent*

$$R(\lambda) = \big(-\Delta_M - \tfrac14 - \lambda^2\big)^{-1} : \begin{cases} L^2(M) \to L^2(M), & \operatorname{Im}\lambda > 0, \\ L^2_{\text{comp}}(M) \to L^2_{\text{loc}}(M), & \operatorname{Im}\lambda \le 0. \end{cases}$$

Assume that $0 < \delta < 1$. *Then* M *has an essential spectral gap of size*

$$\beta = \tfrac12 - \delta + (13C_R)^{-\frac{320}{\delta(1-\delta)}}; \tag{4-1}$$

that is, $R(\lambda)$ *has only finitely many poles in* $\{\operatorname{Im}\lambda > -\beta\}$ *and it satisfies the cutoff estimates for each* $\psi \in C_0^\infty(M)$, $\varepsilon > 0$ *and some constant* C_0 *depending on* ε

$$\|\psi R(\lambda)\psi\|_{L^2\to L^2} \le C(\psi,\varepsilon)|\lambda|^{-1-2\min(0,\operatorname{Im}\lambda)+\varepsilon}, \quad \operatorname{Im}\lambda \in [-\beta, 1],\ |\operatorname{Re}\lambda| \ge C_0.$$

Proof. We use the strategy of [Dyatlov and Zahl 2016]. By Theorem 3 of that paper, it suffices to show the following fractal uncertainty principle: for each $\rho \in (0,1)$,

$$\beta_0 := \tfrac{1}{2} - \delta + (150C_R^2)^{-\frac{160}{\delta(1-\delta)}},$$

and each cutoff function $\chi \in C^\infty(\mathbb{S}^1 \times \mathbb{S}^1)$ supported away from the diagonal, there exists a constant C depending on M, χ, ρ such that for all $h \in (0,1)$

$$\|\mathbb{1}_{\Lambda_\Gamma(h^\rho)} B_{\chi,h} \mathbb{1}_{\Lambda_\Gamma(h^\rho)}\|_{L^2(\mathbb{S}^1)\to L^2(\mathbb{S}^1)} \le C h^{\beta_0 - 2(1-\rho)}, \tag{4-2}$$

where $\Lambda_\Gamma(h^\rho) \subset \mathbb{S}^1$ is the h^ρ neighborhood of Λ_Γ and the operator $B_{\chi,h}$ is defined by (here $|x-y|$ is the Euclidean distance between $x, y \in \mathbb{S}^1 \subset \mathbb{R}^2$)

$$B_{\chi,h} f(x) = (2\pi h)^{-\frac{1}{2}} \int_{\mathbb{S}^1} |x-y|^{\frac{2i}{h}} \chi(x,y) f(y)\,dy.$$

To show (4-2), we first note that by Lemma 2.2, (Y, μ_Y) is δ-regular up to scale h with constant $30C_R^2$, where $Y = \Lambda_\Gamma(h)$ and μ_Y is $h^{\delta-1}$ times the restriction of the Lebesgue measure to Y. We lift $\chi(x,y)$ to a compactly supported function on $\mathbb{R}^2$ (splitting it into pieces using a partition of unity) and write

$$B_{\chi,h} \mathbb{1}_{\Lambda_\Gamma(h)} f(x) = (2\pi)^{-\frac{1}{2}} h^{\frac{1}{2}-\delta} \mathcal{B}_h f(x),$$

where $\mathcal{B}_h$ has the form (1-3) with $G(x,y) = \chi(x,y)$ and (with $|x-y|$ still denoting the Euclidean distance between $x, y \in \mathbb{S}^1$)

$$\Phi(x,y) = 2\log|x-y|.$$

The function Φ is smooth and satisfies the condition $\partial^2_{xy}\Phi \neq 0$ on the open set $\mathbb{S}^1 \times \mathbb{S}^1 \setminus \{x=y\}$ which contains the support of G; see for instance [Bourgain and Dyatlov 2016, Section 4.3]. Applying Theorem 1 with $(X, \mu_X) := (Y, \mu_Y)$, we obtain

$$\|\mathbb{1}_{\Lambda_\Gamma(h)} B_{\chi,h} \mathbb{1}_{\Lambda_\Gamma(h)}\|_{L^2(\mathbb{S}^1)\to L^2(\mathbb{S}^1)} \le C h^{\beta_0}.$$

Similarly we have

$$\|\mathbb{1}_{\Lambda_\Gamma(h)+t} B_{\chi,h} \mathbb{1}_{\Lambda_\Gamma(h)+s}\|_{L^2(\mathbb{S}^1)\to L^2(\mathbb{S}^1)} \le C h^{\beta_0}, \quad t, s \in [-1, 1],$$

where $X + t$ is the result of rotating $X \subset \mathbb{S}^1$ by angle t. Covering $\Lambda_\Gamma(h^\rho)$ with at most $10h^{\rho-1}$ rotations of the set $\Lambda_\Gamma(h)$, see for instance the proof of [Bourgain and Dyatlov 2016, Proposition 4.2], and using triangle inequality, we obtain (4-2), finishing the proof. □

We now briefly discuss previous results on spectral gaps for hyperbolic surfaces:

- The works [Patterson 1976; Sullivan 1979] imply that $R(\lambda)$ has no poles with $\operatorname{Im}\lambda > \delta - \frac{1}{2}$. On the other hand, the fact that $R(\lambda)$ is the L^2 resolvent of the Laplacian in $\{\operatorname{Im}\lambda > 0\}$ shows that it has only has finitely many poles in this region. Together these two results give the essential spectral gap $\beta = \max(0, \frac{1}{2} - \delta)$. Thus Theorem 2 gives no new results when δ is much larger than $\frac{1}{2}$.

- Using the method developed by Dolgopyat [1998], Naud [2005] showed an essential spectral gap of size $\beta > \frac{1}{2} - \delta$ when $\delta > 0$. Oh and Winter [2016] showed that the size of the gap is uniformly controlled for towers of congruence covers in the arithmetic case.

- Dyatlov and Zahl [2016] introduced the fractal-uncertainty-principle approach to spectral gaps and used it together with tools from additive combinatorics to give an estimate of the size of the gap in terms of C_R in the case when δ is very close to $\frac{1}{2}$.

- Bourgain and Dyatlov [2016] showed that each convex cocompact hyperbolic surface has an essential spectral gap of some size $\beta = \beta(\delta, C_R) > 0$. Their result is new in the case $\delta > \frac{1}{2}$ and is thus complementary to the results mentioned above, as well as to Theorem 2.

More generally, spectral gaps have been studied for noncompact manifolds with hyperbolic trapped sets. (See for instance [Nonnenmacher 2011, Section 2.1] for a definition.) In this setting the Patterson–Sullivan gap $\frac{1}{2} - \delta$ generalizes to the *pressure gap* $-P\left(\frac{1}{2}\right)$ which has been established by Ikawa [1988], Gaspard and Rice [1989], and Nonnenmacher and Zworski [2009]. An improved gap $\beta > -P\left(\frac{1}{2}\right)$ has been proved in several cases; see in particular [Petkov and Stoyanov 2010; Stoyanov 2011; 2012]. We refer the reader to [Nonnenmacher 2011] for an overview of results on spectral gaps for general hyperbolic trapped sets.

5. Application: spectral gap for open quantum maps

In this section, we discuss applications of the fractal uncertainty principle to the spectral properties of open quantum maps. Following the notation in [Dyatlov and Jin 2017] we consider an open quantum baker's map B_N determined by a triple $(M, \mathcal{A}, \chi)$, where $M \in \mathbb{N}$ is called the base, $\mathcal{A} \subset \mathbb{Z}_M = \{0, 1, \dots, M-1\}$ is called the alphabet, and $\chi \in C_0^\infty((0,1); [0,1])$ is a cutoff function. The map B_N is a sequence of operators $B_N : \ell^2_N \to \ell^2_N$, $\ell^2_N = \ell^2(\mathbb{Z}_N)$, defined for every positive $N \in M\mathbb{Z}$ by

$$B_N = \mathcal{F}_N^* \begin{pmatrix} \chi_{N/M}\, \mathcal{F}_{N/M}\, \chi_{N/M} & & \\ & \ddots & \\ & & \chi_{N/M}\, \mathcal{F}_{N/M}\, \chi_{N/M} \end{pmatrix} I_{\mathcal{A},M}, \tag{5-1}$$

where $\mathcal{F}_N$ is the unitary Fourier transform given by the $N \times N$ matrix $(1/\sqrt{N})(e^{-\frac{2\pi i j \ell}{N}})_{j\ell}$, $\chi_{N/M}$ is the multiplication operator on $\ell^2_{N/M}$ discretizing χ, and $I_{\mathcal{A},M}$ is the diagonal matrix with ℓ-th diagonal entry equal to 1 if $\lfloor \ell/(N/M) \rfloor \in \mathcal{A}$ and 0 otherwise.

An important difference from [Dyatlov and Jin 2017] is that in the present paper we allow N to be any multiple of M, while they required that N be a power of M. To measure the size of N, we let k be the unique integer such that $M^k \le N < M^{k+1}$, i.e., $k = \lfloor \log N / \log M \rfloor$. Denote by δ the dimension of the Cantor set corresponding to M and $\mathcal{A}$, given by

$$\delta = \frac{\log |\mathcal{A}|}{\log M}.$$

The main result of this section is the following spectral gap, which was previously established in [Dyatlov and Jin 2017, Theorem 1] for the case when N is a power of M:

Theorem 3. *Assume that* $0 < \delta < 1$; *that is,* $1 < |\mathcal{A}| < M$. *Then there exists*

$$\beta = \beta(M, \mathcal{A}) > \max\big(0, \tfrac{1}{2} - \delta\big) \tag{5-2}$$

such that, with $\mathrm{Sp}(B_N) \subset \{\lambda \in \mathbb{C} : |\lambda| \le 1\}$ *denoting the spectrum of* B_N,

$$\limsup_{N\to\infty,\, N\in M\mathbb{Z}} \max\{|\lambda| : \lambda \in \mathrm{Sp}(B_N)\} \le M^{-\beta}. \tag{5-3}$$

The main component of the proof is a fractal uncertainty principle. For the case $N = M^k$, the following version of it was used in [Dyatlov and Jin 2017]:

$$\|\mathbb{1}_{\mathcal{C}_k} \mathcal{F}_N \mathbb{1}_{\mathcal{C}_k}\|_{\ell^2_N \to \ell^2_N} \le C N^{-\beta}, \tag{5-4}$$

where $\mathcal{C}_k$ is the discrete Cantor set given by

$$\mathcal{C}_k := \left\{ \sum_{j=0}^{k-1} a_j M^j : a_0, \dots, a_{k-1} \in \mathcal{A} \right\} \subset \mathbb{Z}_N. \tag{5-5}$$

For general $N \in M\mathbb{Z} \cap [M^k, M^{k+1})$, we define a similar discrete Cantor set in $\mathbb{Z}_N$ by

$$\mathcal{C}_k(N) := \{b_j(N) : j \in \mathcal{C}_k\} \subset \mathbb{Z}_N, \quad b_j(N) := \left\lceil \frac{jN}{M^k} \right\rceil. \tag{5-6}$$

In fact, in our argument we only need $b_j(N)$ to be some integer in $[jN/M^k, (j+1)N/M^k)$.

The uncertainty principle then takes the following form:

Theorem 4. *Assume that* $0 < \delta < 1$. *Then there exists*

$$\beta = \beta(M, \mathcal{A}) > \max\big(0, \tfrac{1}{2} - \delta\big) \tag{5-7}$$

such that for some constant C *and all* N,

$$\|\mathbb{1}_{\mathcal{C}_k(N)} \mathcal{F}_N \mathbb{1}_{\mathcal{C}_k(N)}\|_{\ell^2_N \to \ell^2_N} \le C N^{-\beta}. \tag{5-8}$$

In Section 5A below, we show that Theorem 4 implies Theorem 3. We prove Theorem 4 in Sections 5C and 5D using Ahlfors–David regularity of the Cantor set, which is verified in Section 5B.

5A. *Fractal uncertainty principle implies spectral gap.* We first show that Theorem 4 implies Theorem 3. The argument is essentially the same as in [Dyatlov and Jin 2017, Section 2.3], relying on the following generalization of Proposition 2.5 from that paper:

Proposition 5.1 (localization of eigenstates). *Fix* $\nu > 0$, $\rho \in (0, 1)$, *and assume that for some* $k \in \mathbb{N}$, $N \in M\mathbb{Z} \cap [M^k, M^{k+1})$, $\lambda \in \mathbb{C}$, $u \in \ell^2_N$, *we have*

$$B_N u = \lambda u, \quad |\lambda| \ge M^{-\nu}.$$

Define

$$X_\rho := \bigcup \{\mathcal{C}_k(N) + m : m \in \mathbb{Z},\ |m| \le (M+2)N^{1-\rho}\} \subset \mathbb{Z}_N.$$

Then

$$\|u\|_{\ell^2_N} \le M^\nu |\lambda|^{-\rho k} \, \|\mathbb{1}_{X_\rho} u\|_{\ell^2_N} + \mathcal{O}(N^{-\infty})\|u\|_{\ell^2_N}, \tag{5-9}$$

$$\|u - \mathcal{F}_N^* \mathbb{1}_{X_\rho} \mathcal{F}_N u\|_{\ell^2_N} = \mathcal{O}(N^{-\infty})\|u\|_{\ell^2_N}, \tag{5-10}$$

where the constants in $\mathcal{O}(N^{-\infty})$ depend only on ν, ρ, χ.

Proof. Following [Dyatlov and Jin 2017, (2.7)], let $\Phi = \Phi_{M,\mathcal{A}}$ be the expanding map defined by

$$\Phi : \bigsqcup_{a \in \mathcal{A}} \left(\frac{a}{M}, \frac{a+1}{M} \right) \to (0,1), \quad \Phi(x) = Mx - a,\ x \in \left(\frac{a}{M}, \frac{a+1}{M} \right). \tag{5-11}$$

Put

$$\tilde{k} := \lceil \rho k \rceil \in \{1, \dots, k\}. \tag{5-12}$$

With $d(\cdot,\cdot)$ denoting the distance function on the circle as in [Dyatlov and Jin 2017, Section 2.1], define

$$\mathcal{X}_\rho := \{ x \in [0,1] : d(x, \Phi^{-\tilde{k}}([0,1])) \le N^{-\rho} \}.$$

Then (5-9), (5-10) follow from the long time Egorov theorem [Dyatlov and Jin 2017, Proposition 2.4] (whose proof never used that N is a power of M) similarly to Proposition 2.5 of the same paper, as long as we show the following analog of [Dyatlov and Jin 2017, (2.30)]:

$$\ell \in \{0, \dots, N-1\}, \quad \frac{\ell}{N} \in \mathcal{X}_\rho \quad \Longrightarrow \quad \ell \in X_\rho. \tag{5-13}$$

To see (5-13), note that (with the intervals considered in $\mathbb{R}/\mathbb{Z}$)

$$\Phi^{-\tilde{k}}([0,1]) \subset \bigcup_{j \in \mathcal{C}_k} \left(\frac{j - M^{k-\tilde{k}}}{M^k}, \frac{j + M^{k-\tilde{k}}}{M^k} \right).$$

Assume that $\ell \in \{0, \dots, N-1\}$ and $\ell/N \in \mathcal{X}_\rho$. Then there exists $j \in \mathcal{C}_k$ such that

$$d\left(\frac{\ell}{N}, \frac{j}{M^k} \right) \le N^{-\rho} + M^{-\tilde{k}} \le (M+1)N^{-\rho}.$$

It follows that

$$d\left(\frac{\ell}{N}, \frac{b_j(N)}{N} \right) \le (M+2)N^{-\rho}$$

and thus $\ell \in X_\rho$ as required. □

Now, we assume that Theorem 4 holds and prove Theorem 3. Using the triangle inequality as in the proof of [Dyatlov and Jin 2017, Proposition 2.6], we obtain

$$\begin{aligned} \|\mathbb{1}_{X_\rho} \mathcal{F}_N^* \mathbb{1}_{X_\rho}\|_{\ell^2_N \to \ell^2_N} &\le (2M+5)^2 N^{2(1-\rho)} \|\mathbb{1}_{\mathcal{C}_k(N)} \mathcal{F}_N \mathbb{1}_{\mathcal{C}_k(N)}\|_{\ell^2_N \to \ell^2_N} \\ &\le C N^{2(1-\rho)-\beta}. \end{aligned} \tag{5-14}$$

Here C denotes a constant independent of N.

Assume that $\lambda \in \mathbb{C}$ is an eigenvalue of B_N such that $|\lambda| \geq M^{-\beta}$ and $u \in \ell^2_N$ is a normalized eigenfunction of B_N with eigenvalue λ. By (5-9), (5-10), and (5-14)

$$\begin{aligned} 1 = \|u\|_{\ell^2_N} &\leq M^\beta |\lambda|^{-\rho k} \|\mathbb{1}_{X_\rho} u\|_{\ell^2_N} + \mathcal{O}(N^{-\infty}) \\ &\leq M^\beta |\lambda|^{-\rho k} \|\mathbb{1}_{X_\rho} \mathcal{F}_N^* \mathbb{1}_{X_\rho} \mathcal{F}_N u\|_{\ell^2_N} + \mathcal{O}(N^{-\infty}) \\ &\leq C|\lambda|^{-\rho k} N^{2(1-\rho)-\beta} + \mathcal{O}(N^{-\infty}). \end{aligned} \tag{5-15}$$

It follows that $|\lambda|^{\rho k} \leq C N^{-\beta+2(1-\rho)}$ or equivalently

$$|\lambda| \leq C^{\frac{1}{\rho} k} M^{\frac{2(1-\rho)-\beta}{\rho}}.$$

This implies

$$\limsup_{N\to\infty} \max\{|\lambda| : \lambda \in \operatorname{Sp}(B_N)\} \leq \max\{M^{-\beta}, M^{\frac{2(1-\rho)-\beta}{\rho}}\}.$$

Letting $\rho \to 1$, we conclude the proof of Theorem 3.

5B. *Regularity of discrete Cantor sets.* Theorem 4 will be deduced from Theorem 1 and the results of [Bourgain and Dyatlov 2016]. To apply these, we establish Ahlfors–David regularity of the Cantor set $\mathcal{C}_k(N) \subset \mathbb{Z}_N = \{0, \dots, N-1\}$ in the following discrete sense.

Definition 5.2. We say that $X \subset \mathbb{Z}_N$ is δ-regular with constant C_R if

- for each interval J of size $|J| \geq 1$, we have $\#(J \cap X) \leq C_R |J|^\delta$, and
- for each interval J with $1 \leq |J| \leq N$ which is centered at a point in X, we have $\#(J \cap X) \geq C_R^{-1} |J|^\delta$.

Definition 5.2 is related to Definition 1.1 as follows:

Lemma 5.3. *Let $X \subset \mathbb{Z}_N$. Define $\widetilde{X} := N^{-1} X \subset [0,1]$ which supports the measure*

$$\mu_{\widetilde{X}}(A) := N^{-\delta} \cdot \#(\widetilde{X} \cap A), \quad A \subset \mathbb{R}. \tag{5-16}$$

Then X is δ-regular with constant C_R in the sense of Definition 5.2 if and only if $(\widetilde{X}, \mu_{\widetilde{X}})$ is δ-regular up to scale N^{-1} with constant C_R in the sense of Definition 1.1.

Proof. This follows directly from the two definitions. □

We first establish the regularity of the discrete Cantor set $\mathcal{C}_k$ defined in (5-5):

Lemma 5.4. *The set $\mathcal{C}_k \subset \mathbb{Z}_{M^k}$ is δ-regular with constant $C_R = 2M^{2\delta}$.*

Proof. We notice that for all integers $k' \in [0,k]$ and $j' \in \mathbb{Z}$

$$\#\big(\mathcal{C}_k \cap [j' M^{k'}, (j'+1) M^{k'})\big) = \begin{cases} |\mathcal{A}|^{k'} = M^{\delta k'}, & j' \in \mathcal{C}_{k-k'}, \\ 0, & j' \notin \mathcal{C}_{k-k'}. \end{cases} \tag{5-17}$$

Let J be an interval in $\mathbb{R}$, with $1 \leq |J| \leq N = M^k$. Choose an integer $k' \in [0, k-1]$ such that $M^{k'} \leq |J| \leq M^{k'+1}$. Then there exists some $j' \in \mathbb{Z}$ such that

$$J \subset [j' M^{k'+1}, (j'+2) M^{k'+1}).$$

Therefore by (5-17)

$$\#(\mathcal{C}_k \cap J) \le 2M^{\delta(k'+1)} \le 2M^{\delta}|J|^{\delta} \le C_R|J|^{\delta}.$$

On the other hand, if $|J| > N$ then

$$\#(\mathcal{C}_k \cap J) \le \#(\mathcal{C}_k) = N^{\delta} \le |J|^{\delta}.$$

This gives the required upper bound on $\#(\mathcal{C}_k \cap J)$.

Now, assume that $1 \le |J| \le N$ and J is centered at some $j \in \mathcal{C}_k$. Choose k' as before. If $k' = 0$ then

$$\#(\mathcal{C}_k \cap J) \ge 1 \ge M^{-\delta}|J|^{\delta} \ge C_R^{-1}|J|^{\delta}.$$

We henceforth assume that $1 \le k' \le k-1$. Let $j' \in \mathcal{C}_{k-k'+1}$ be the unique element such that $j'M^{k'-1} \le j < (j'+1)M^{k'-1}$. Since $M \ge 2$, we have $|J| \ge M^{k'} \ge 2M^{k'-1}$ and thus

$$[j'M^{k'-1}, (j'+1)M^{k'-1}] \subset [j - M^{k'-1}, j + M^{k'-1}] \subset J.$$

Therefore by (5-17)

$$\#(\mathcal{C}_k \cap J) \ge M^{\delta(k'-1)} \ge M^{-2\delta}|J|^{\delta} \ge C_R^{-1}|J|^{\delta}.$$

This gives the required lower bound on $\#(\mathcal{C}_k \cap J)$, finishing the proof. □

We now establish regularity of the dilated Cantor set $\mathcal{C}_k(N)$:

Proposition 5.5. *Assume that $M^k \le N < M^{k+1}$ and let $\mathcal{C}_k(N) \subset \mathbb{Z}_N$ be given by (5-6). Then $\mathcal{C}_k(N)$ is δ-regular with constant $C_R = 8M^{3\delta}$.*

Proof. For any interval J, we have

$$\#(\mathcal{C}_k(N) \cap J) = \#\{j \in \mathcal{C}_k : b_j(N) \in J\} = \#\Big\{j \in \mathcal{C}_k : \frac{M^k}{N} b_j(N) \in \frac{M^k}{N} J\Big\}.$$

By our choice of $b_j(N)$, we have $(M^k/N)b_j(N) \in [j, j+1)$. Therefore

$$\#\Big(\mathcal{C}_k \cap \frac{M^k}{N} J\Big) - 1 \le \#(\mathcal{C}_k(N) \cap J) \le \#\Big(\mathcal{C}_k \cap \frac{M^k}{N} J\Big) + 1.$$

We apply Lemma 5.4 to see that for any interval J with $|J| \ge 1$

$$\#(\mathcal{C}_k(N) \cap J) \le 2M^{2\delta}|J|^{\delta} + 1 \le 3M^{2\delta}|J|^{\delta} \le C_R|J|^{\delta}.$$

Now, assume that J is an interval with $8^{\frac{1}{\delta}}M^3 \le |J| \le N$ centered at $b_j(N)$ for some $j \in \mathcal{C}_k$. Then $(M^k/N)J$ contains the interval of size $\frac{1}{2M}|J|$ centered at j. Therefore, by Lemma 5.4

$$\#(\mathcal{C}_k(N) \cap J) \ge \frac{1}{2M^{2\delta}}\left(\frac{|J|}{2M}\right)^{\delta} - 1 \ge \frac{|J|^{\delta}}{8M^{3\delta}} \ge C_R^{-1}|J|^{\delta}.$$

Finally, if J is an interval with $1 \le |J| \le 8^{\frac{1}{\delta}}M^3$ centered at a point in $\mathcal{C}_k(N)$, then

$$\#(\mathcal{C}_k(N) \cap J) \ge 1 \ge C_R^{-1}|J|^{\delta}.$$

□

5C. *Fractal uncertainty principle for $\delta \le \frac{1}{2}$.* The proof of Theorem 4 in the case $\delta \le \frac{1}{2}$ relies on the following corollary of Theorem 1:

Proposition 5.6. *Let $X, Y \subset \mathbb{Z}_N$ be δ-regular with constant C_R and $0 < \delta < 1$. Then*

$$\|\mathbb{1}_X \mathcal{F}_N \mathbb{1}_Y\|_{\ell^2_N \to \ell^2_N} \le C N^{-(\frac{1}{2}-\delta+\varepsilon_0)}, \tag{5-18}$$

where C only depends on δ, C_R and

$$\varepsilon_0 = (5C_R)^{-\frac{160}{\delta(1-\delta)}}. \tag{5-19}$$

Proof. Put $h := N^{-1}$, $\widetilde{X} := hX$, $\widetilde{Y} := hY$, and define the measures $\mu_{\widetilde{X}}, \mu_{\widetilde{Y}}$ by (5-16). By Lemma 5.3, $(\widetilde{X}, \mu_{\widetilde{X}})$ and $(\widetilde{Y}, \mu_{\widetilde{Y}})$ are δ-regular up to scale h with constant C_R. Consider the operator

$$\mathcal{B}_h : L^1(\widetilde{Y}, \mu_{\widetilde{Y}}) \to L^\infty(\widetilde{X}, \mu_{\widetilde{X}})$$

defined by

$$\mathcal{B}_h f(x) = \int_{\widetilde{Y}} \exp\left(-\frac{2\pi i x y}{h}\right) f(y)\, d\mu_{\widetilde{Y}}(y)$$

and note that it has the form (1-3) with $\Phi(x, y) = -2\pi x y$, $G \equiv 1$. By Theorem 1

$$\|\mathcal{B}_h\|_{L^2(\widetilde{Y}, \mu_{\widetilde{Y}}) \to L^2(\widetilde{X}, \mu_{\widetilde{X}})} \le C h^{\varepsilon_0}.$$

Comparing the formula

$$\mathcal{B}_h f\left(\frac{j}{N}\right) = N^{-\delta} \sum_{\ell \in Y} \exp\left(-\frac{2\pi i j \ell}{N}\right) f\left(\frac{\ell}{N}\right), \quad j \in X,$$

with the definition of the discrete Fourier transform $\mathcal{F}_N$, we see that

$$\|\mathbb{1}_X \mathcal{F}_N \mathbb{1}_Y\|_{\ell^2_N \to \ell^2_N} = N^{\delta - \frac{1}{2}} \|\mathcal{B}_h\|_{L^2(\widetilde{Y}, \mu_{\widetilde{Y}}) \to L^2(\widetilde{X}, \mu_{\widetilde{X}})},$$

which finishes the proof. □

Combining Propositions 5.5 and 5.6, we get (5-8) for

$$\beta = \tfrac{1}{2} - \delta + (40 M^{3\delta})^{-\frac{160}{\delta(1-\delta)}} \tag{5-20}$$

which finishes the proof of Theorem 4 for $\delta \le \frac{1}{2}$.

5D. *Fractal uncertainty principle for $\delta > \frac{1}{2}$.* For $\delta > \frac{1}{2}$, Theorem 1 does not in general give an improvement over the trivial gap $\beta = 0$. Instead, we shall use the following reformulation of [Bourgain and Dyatlov 2016, Theorem 4]:

Proposition 5.7. *Let $0 \le \delta < 1$, $C_R \ge 1$, $N \ge 1$ and assume that $\widetilde{X}, \widetilde{Y} \subset [-1, 1]$ and $(\widetilde{X}, \mu_{\widetilde{X}})$ and $(\widetilde{Y}, \mu_{\widetilde{Y}})$ are δ-regular up to scale N^{-1} with constant C_R in the sense of Definition 1.1, for some finite measures $\mu_{\widetilde{X}}, \mu_{\widetilde{Y}}$ supported on $\widetilde{X}, \widetilde{Y}$.*

Then there exist $\beta_0 > 0$, C_0 depending only on δ, C_R such that for all $f \in L^2(\mathbb{R})$,

$$\operatorname{supp} \hat{f} \subset N \cdot \widetilde{Y} \quad \Longrightarrow \quad \|f\|_{L^2(\widetilde{X})} \le C_0 N^{-\beta_0} \|f\|_{L^2(\mathbb{R})}. \tag{5-21}$$

Here $\hat f$ denotes the Fourier transform of f:

$$htf(\xi) = \mathcal{F}f(\xi) = \int_{\mathbb{R}} e^{-2\pi i x\xi} f(x)\,dx. \tag{5-22}$$

Proposition 5.7 implies the following discrete fractal uncertainty principle:

Proposition 5.8. *Let $X, Y \subset \mathbb{Z}_N$ be δ-regular with constant C_R and $0 \le \delta < 1$. Then*

$$\|\mathbb{1}_X \mathcal{F}_N \mathbb{1}_Y\|_{\ell^2_N \to \ell^2_N} \le CN^{-\beta}, \tag{5-23}$$

where $C, \beta > 0$ only depend on δ, C_R.

Proof. Put $h := N^{-1}$,

$$\widetilde{X} := hX + [-h, h], \quad \widetilde{Y} := hY + [-h, h],$$

and define the measures $\mu_{\widetilde{X}}, \mu_{\widetilde{Y}}$ on $\widetilde{X}, \widetilde{Y}$ by (2-6). By Lemmas 5.3 and 2.2, $(\widetilde{X}, \mu_{\widetilde{X}})$ and $(\widetilde{Y}, \mu_{\widetilde{Y}})$ are δ-regular up to scale h with constant $30C_R^2$. Applying Proposition 5.7, we obtain for some constants $\beta_0 > 0$, C_0 depending only on δ, C_R and all $f \in L^2(\mathbb{R})$

$$\operatorname{supp} htf \subset N \cdot \widetilde{Y} \quad \Longrightarrow \quad \|f\|_{L^2(\widetilde{X})} \le C_0 N^{-\beta_0} \|f\|_{L^2(\mathbb{R})}. \tag{5-24}$$

To pass from (5-24) to (5-23), fix a cutoff function χ such that for some constant $c > 0$

$$\chi \in C_0^\infty\big((-\tfrac{1}{2}, \tfrac{1}{2})\big), \quad \|\chi\|_{L^2} = 1, \quad \inf_{[0,1]} |\mathcal{F}^{-1}\chi| \ge c.$$

This is possible since for any $\chi \in C_0^\infty(\mathbb{R})$ which is not identically 0, $\mathcal{F}^{-1}\chi$ extends to an entire function and thus has no zeros on $\{\operatorname{Im} z = s\}$ for all but countably many choices of $s \in \mathbb{R}$. Choosing such s we see that $\mathcal{F}^{-1}(e^{-s\xi}\chi(\xi))$ has no real zeros.

Now, take arbitrary $u \in \ell^2_N$. Consider the function $f \in L^2(\mathbb{R})$ defined by

$$\hat f(\xi) = \sum_{\ell \in Y} u(\ell)\chi(\xi - \ell).$$

Then $\operatorname{supp} \hat f \subset N \cdot \widetilde{Y}$ and $\|f\|_{L^2(\mathbb{R})} \le \|u\|_{\ell^2_N}$, so by (5-24)

$$\|f\|_{L^2(\widetilde{X})} \le C_0 N^{-\beta_0} \|u\|_{\ell^2_N}. \tag{5-25}$$

On the other hand, for all $j \in \mathbb{Z}_N$, we have for all $j \in X$

$$\frac{1}{\sqrt{N}} f\Big(\frac{j}{N}\Big) = \mathcal{F}_N^* \mathbb{1}_Y u(j) \cdot (\mathcal{F}^{-1}\chi)\Big(\frac{j}{N}\Big). \tag{5-26}$$

Consider the nonoverlapping collection of intervals

$$I_j := \Big[\frac{j}{N} - \frac{1}{2N}, \frac{j}{N} + \frac{1}{2N}\Big] \subset \widetilde{X}, \quad j \in X.$$

Using that $(|f|^2)' = 2\,\mathrm{Re}(\bar f f')$, we have

$$|\mathcal{F}_N^* \mathbb{1}_Y u(j)|^2 \le \frac{1}{c^2 N}\left| f\left(\frac{j}{N}\right)\right|^2 \le C \int_{I_j} |f(x)|^2\,dx + \frac{C}{N}\int_{I_j} |f(x)|\cdot|f'(x)|\,dx,$$

where C denotes some constant depending only on δ, C_R, χ. Summing over $j \in X$ and using the Cauchy–Schwarz inequality, we obtain

$$\|\mathbb{1}_X \mathcal{F}_N^* \mathbb{1}_Y u\|^2_{\ell^2_N} \le C\|f\|^2_{L^2(\tilde X)} + \frac{C}{N}\|f\|_{L^2(\tilde X)}\cdot\|f'\|_{L^2(\mathbb{R})}.$$

Since $\operatorname{supp}\hat f \subset [-N, N]$, we have $\|f'\|_{L^2(\mathbb{R})} \le 10N\|f\|_{L^2(\mathbb{R})} \le 10N\|u\|_{\ell^2_N}$ and thus by (5-25)

$$\|\mathbb{1}_X \mathcal{F}_N^* \mathbb{1}_Y u\|^2_{\ell^2_N} \le C N^{-\beta_0}\|u\|^2_{\ell^2_N},$$

which gives (5-23) with $\beta = \frac{1}{2}\beta_0$. □

Combining Propositions 5.5 and 5.8, we obtain (5-8) for $\frac{1}{2} \le \delta < 1$, finishing the proof of Theorem 4.

Acknowledgements

We would like to thank Jean Bourgain for inspiring discussions on the fractal uncertainty principle, and Maciej Zworski for many helpful comments. We would also like to thank an anonymous referee for many suggestions to improve the manuscript. This research was conducted during the period Dyatlov served as a Clay Research Fellow.

References

[Borthwick 2016] D. Borthwick, *Spectral theory of infinite-area hyperbolic surfaces*, 2nd ed., Progress in Mathematics **318**, Springer, 2016. MR Zbl

[Bourgain and Dyatlov 2016] J. Bourgain and S. Dyatlov, "Spectral gaps without the pressure condition", preprint, 2016. To appear in *Ann. Math.* (2). arXiv

[Dolgopyat 1998] D. Dolgopyat, "On decay of correlations in Anosov flows", *Ann. of Math.* (2) **147**:2 (1998), 357–390. MR Zbl

[Dyatlov and Jin 2017] S. Dyatlov and L. Jin, "Resonances for open quantum maps and a fractal uncertainty principle", *Comm. Math. Phys.* **354**:1 (2017), 269–316. MR Zbl

[Dyatlov and Zahl 2016] S. Dyatlov and J. Zahl, "Spectral gaps, additive energy, and a fractal uncertainty principle", *Geom. Funct. Anal.* **26**:4 (2016), 1011–1094. MR Zbl

[Gaspard and Rice 1989] P. Gaspard and S. A. Rice, "Scattering from a classically chaotic repellor", *J. Chem. Phys.* **90**:4 (1989), 2225–2241. MR

[Ikawa 1988] M. Ikawa, "Decay of solutions of the wave equation in the exterior of several convex bodies", *Ann. Inst. Fourier (Grenoble)* **38**:2 (1988), 113–146. MR Zbl

[Magee et al. 2017] M. Magee, H. Oh, and D. Winter, "Uniform congruence counting for Schottky semigroups in $\mathrm{SL}_2(\mathbb{Z})$", *J. Reine Angew. Math.* (online publication January 2017).

[Naud 2005] F. Naud, "Expanding maps on Cantor sets and analytic continuation of zeta functions", *Ann. Sci. École Norm. Sup.* (4) **38**:1 (2005), 116–153. MR Zbl

[Nonnenmacher 2011] S. Nonnenmacher, "Spectral problems in open quantum chaos", *Nonlinearity* **24**:12 (2011), R123–R167. Zbl

[Nonnenmacher and Zworski 2009] S. Nonnenmacher and M. Zworski, "Quantum decay rates in chaotic scattering", *Acta Math.* **203**:2 (2009), 149–233. MR Zbl

[Oh and Winter 2016] H. Oh and D. Winter, "Uniform exponential mixing and resonance free regions for convex cocompact congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$", *J. Amer. Math. Soc.* **29**:4 (2016), 1069–1115. MR Zbl

[Patterson 1976] S. J. Patterson, "The limit set of a Fuchsian group", *Acta Math.* **136**:3-4 (1976), 241–273. MR Zbl

[Petkov and Stoyanov 2010] V. Petkov and L. Stoyanov, "Analytic continuation of the resolvent of the Laplacian and the dynamical zeta function", *Anal. PDE* **3**:4 (2010), 427–489. MR Zbl

[Stoyanov 2011] L. Stoyanov, "Spectra of Ruelle transfer operators for axiom A flows", *Nonlinearity* **24**:4 (2011), 1089–1120. MR Zbl

[Stoyanov 2012] L. Stoyanov, "Non-integrability of open billiard flows and Dolgopyat-type estimates", *Ergodic Theory Dynam. Systems* **32**:1 (2012), 295–313. MR Zbl

[Sullivan 1979] D. Sullivan, "The density at infinity of a discrete group of hyperbolic motions", *Inst. Hautes Études Sci. Publ. Math.* **50** (1979), 171–202. MR Zbl

Received 23 Feb 2017. Revised 26 Oct 2017. Accepted 12 Jan 2018.

SEMYON DYATLOV: dyatlov@math.mit.edu
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, United States

LONG JIN: long249@purdue.edu
Department of Mathematics, Purdue University, West Lafayette, IN, United States

ANALYSIS AND PDE
Vol. 11, No. 6, 2018
dx.doi.org/10.2140/apde.2018.11.1487

DINI AND SCHAUDER ESTIMATES FOR NONLOCAL FULLY NONLINEAR PARABOLIC EQUATIONS WITH DRIFTS

HONGJIE DONG, TIANLING JIN AND HONG ZHANG

We obtain Dini- and Schauder-type estimates for concave fully nonlinear nonlocal parabolic equations of order $\sigma \in (0, 2)$ with rough and nonsymmetric kernels and drift terms. We also study such linear equations with only measurable coefficients in the time variable, and obtain Dini-type estimates in the spacial variable. This is a continuation of work by the authors Dong and Zhang.

1. Introduction and main results

The paper is a continuation of [Dong and Zhang 2016a; 2016b] by the first and last authors, where they obtained Schauder-type estimates for concave fully nonlinear nonlocal parabolic equations and Dini-type estimates for concave fully nonlinear nonlocal elliptic equations. Here, we consider concave fully nonlinear nonlocal parabolic equations with Dini continuous coefficients, drifts and nonhomogeneous terms, and establish a C^σ estimate under these assumptions.

The study of second-order equations with Dini continuous coefficients and data dates back to at least 1970s, when Burch [1978] first considered divergence-type linear elliptic equations with Dini continuous coefficients and data, and estimated the modulus of continuity of the derivatives of solutions. Later work for second-order linear or concave fully nonlinear elliptic and parabolic equations with Dini data includes, for example, [Sperner 1981; Lieberman 1987; Safonov 1988; Kovats 1997; Bao 2002; Duzaar and Gastel 2002; Wang 2006; Maz'ya and McOwen 2011; Li 2017], and many others.

The regularity theory for nonlocal elliptic and parabolic equations has been developed extensively in recent years. For example, C^α estimates, $C^{1,\alpha}$ estimates, an Evans–Krylov-type theorem, and Schauder estimates were established in the past decade. See, for instance, [Caffarelli and Silvestre 2009; 2011; Dong and Kim 2012; 2013; Kim and Lee 2013; Lara and Dávila 2014; Mikulevičius and Pragarauskas 2014; Chang-Lara and Kriventsov 2017; Jin and Xiong 2015; 2016; Serra 2015; Mou 2016; Imbert et al. 2016]. In particular, Mou [2016] investigated a class of concave fully nonlinear nonlocal elliptic equations with smooth symmetric kernels, and obtained the C^σ estimate under a slightly stronger assumption than the usual Dini continuity on the coefficients and data. He implemented a recursive Evans–Krylov theorem, which was first studied by Jin and Xiong [2016], as well as a perturbation-type argument. By using a novel perturbation-type argument, the first and last authors proved the C^σ estimate for

H. Dong and H. Zhang were partially supported by the NSF under agreements DMS-1056737 and DMS-1600593. T. Jin was partially supported by Hong Kong RGC grant ECS 26300716 and HKUST initiation grant IGN16SC04.
MSC2010: primary 35B45, 35K55, 35R09; secondary 35B65, 60J75.
Keywords: nonlocal fully nonlinear parabolic equations, Dini- and Schauder-type estimates, critical drift terms.

concave fully nonlinear elliptic equations in [Dong and Zhang 2016a], which relaxed the regularity assumption to simply Dini continuity and also removed the symmetry and smoothness assumptions on the kernels.

In this paper, we extend the results in [Dong and Zhang 2016a] from elliptic equations to parabolic equations with drifts; that is, we study fully nonlinear nonlocal parabolic equations in the form

$$\partial_t u = \inf_{\beta\in\mathcal{A}} (L_\beta u + b_\beta Du + f_\beta), \tag{1-1}$$

where $\mathcal{A}$ is an index set and for each $\beta \in \mathcal{A}$,

$$L_\beta u = \int_{\mathbb{R}^d} \delta u(t,x,y) K_\beta(t,x,y)\,dy,$$

$$\delta u(t,x,y) = \begin{cases} u(t,x+y) - u(t,x) & \text{for } \sigma \in (0,1), \\ u(t,x+y) - u(t,x) - y\cdot Du(t,x)\chi_{B_1} & \text{for } \sigma = 1, \\ u(t,x+y) - u(t,x) - y\cdot Du(t,x) & \text{for } \sigma \in (1,2), \end{cases}$$

and

$$K_\beta(t,x,y) = a_\beta(t,x,y)|y|^{-d-\sigma}.$$

This type of nonlocal operator was first investigated by Komatsu [1984], Mikulevičius and Pragarauskas [1992; 2014], and later by Dong and Kim [2012; 2013], and Schwab and Silvestre [2016].

We assume

$$(2-\sigma)\lambda \le a_\beta(\cdot,\cdot,\cdot) \le (2-\sigma)\Lambda \quad \text{for all } \beta \in \mathcal{A},$$

for some ellipticity constants $0 < \lambda \le \Lambda$, and is merely measurable with respect to the y-variable. When $\sigma = 1$, we additionally assume

$$\int_{S_r} yK_\beta(t,x,y)\,ds_y = 0 \tag{1-2}$$

for any $r > 0$, where S_r is the sphere of radius r centered at the origin.

We also assume $b_\beta \equiv 0$ when $\sigma < 1$ and $b_\beta = b(t,x)$ is independent of β when $\sigma = 1$.

We say that a function f is Dini continuous if its modulus of continuity ω_f is a Dini function, i.e.,

$$\int_0^1 \frac{\omega_f(r)}{r}\,dr < \infty.$$

We need the Dini continuity assumptions on the coefficients of (1-1):

$$\begin{gathered} \sup_{\beta\in\mathcal{A}} \int_{B_{2r}\setminus B_r} |a_\beta(t,x,y) - a_\beta(t',x',y)|\,dy \le \Lambda r^d \omega_a(\max\{|x-x'|, |t-t'|^{\frac{1}{\sigma}}\}) \quad \text{for all } r>0, \\ \sup_{\beta\in\mathcal{A}} \|f_\beta\|_{L_\infty(Q_1)} < \infty, \quad \sup_{\beta\in\mathcal{A}} |f_\beta(t,x) - f_\beta(t',x')| \le \omega_f(\max\{|x-x'|, |t-t'|^{\frac{1}{\sigma}}\}), \\ \sup_{\beta\in\mathcal{A}} \|b_\beta\|_{L_\infty(Q_1)} \le N_0, \quad \sup_{\beta\in\mathcal{A}} |b_\beta(t,x) - b_\beta(t',x')| \le \omega_b(\max\{|x-x'|, |t-t'|^{\frac{1}{\sigma}}\}), \\ \text{where } N_0 > 0, \text{ and } \omega_a, \omega_b, \omega_f \text{ are all Dini functions.} \end{gathered} \tag{1-3}$$

In Theorem 1.1 below, ω_u denotes the modulus of continuity of u in $(-1,0)\times\mathbb{R}^d$; that is,

$$|u(t,x)-u(t',x')|\le\omega_u(\max\{|x-x'|,|t-t'|^{\frac{1}{\sigma}}\})\quad\text{for all }(t,x),(t,x')\in(-1,0)\times\mathbb{R}^d.$$

We also use the notation $C^{1,\sigma^+}(Q_1)$ to denote $C^{1,\sigma+\varepsilon}_{t,x}(Q_1)$ for some arbitrarily small $\varepsilon>0$. This condition is only needed for $L_\beta u$ to be well defined, and may be replaced by other weaker conditions.

Theorem 1.1. *Let $\sigma\in(0,2)$, $0<\lambda\le\Lambda<\infty$, and $\mathcal{A}$ be an index set. Assume for each $\beta\in\mathcal{A}$, K_β satisfies (1-2) when $\sigma=1$, and the Dini continuity assumption (1-3) holds for all $(t,x),(t',x')\in Q_1$. Suppose $u\in C^{1,\sigma^+}(Q_1)$ is a solution of (1-1) in Q_1 and is Dini continuous in $(-1,0)\times\mathbb{R}^d$. Then we have $\partial_t u$ is uniformly continuous and the a priori estimate*

$$\|\partial_t u\|_{L_\infty(Q_{1/2})}+[u]^x_{\sigma;Q_{1/2}}\le C\sum_{j=0}^{\infty}(2^{-j\sigma}\omega_u(2^j)+\omega_u(2^{-j})+\omega_f(2^{-j})),\tag{1-4}$$

where $C>0$ is a constant depending only on d, σ, λ, Λ, N_0, ω_b, and ω_a. Moreover, when $\sigma\ne1$, we have

$$\sup_{(t_0,x_0)\in Q_{1/2}}[u]^x_{\sigma;Q_r(t_0,x_0)}\to0\quad\textit{as } r\to0$$

with a decay rate depending only on d, σ, λ, Λ, ω_a, ω_f, ω_u, N_0, and ω_b. When $\sigma=1$, Du is uniformly continuous in $Q_{1/2}$ with a modulus of continuity controlled by the quantities before.

This theorem improves Theorem 1.1 in [Dong and Zhang 2016a] in the following two ways. First, (1-1) is parabolic and has drift terms. Second, the right-hand side of the estimate (1-4) depends only on the seminorms of u and f, in particular, not on $\sup_{\beta\in\mathcal{A}}\|f_\beta\|_{L_\infty(Q_1)}$.

Remark 1.2. When $\sigma\in(1,2)$ in Theorem 1.1, by interpolation inequalities we have

$$[Du]^t_{\frac{\sigma-1}{\sigma};Q_{1/2}}\le C(\|\partial_t u\|_{L_\infty(Q_{1/2})}+[u]^x_{\sigma;Q_{1/2}})\le C\sum_{j=0}^{\infty}(2^{-j\sigma}\omega_u(2^j)+\omega_u(2^{-j})+\omega_f(2^{-j})).$$

The same proof of Theorem 1.1 can be used to prove Schauder estimates for concave fully nonlinear nonlocal parabolic equations with drifts. To this end, we need the Hölder continuity assumptions on the coefficients of (1-1):

$$\begin{gathered}\sup_{\beta\in\mathcal{A}}\int_{B_{2r}\setminus B_r}|a_\beta(t,x,y)-a_\beta(t',x',y)|\,dy\le\Lambda r^d\max\{|x-x'|^\gamma,|t-t'|^{\frac{\gamma}{\sigma}}\}\quad\text{for all } r>0,\\ \sup_{\beta\in\mathcal{A}}\|f_\beta\|_{L_\infty(Q_1)}<\infty,\quad \sup_{\beta\in\mathcal{A}}|f_\beta(t,x)-f_\beta(t',x')|\le C_f\max\{|x-x'|^\gamma,|t-t'|^{\frac{\gamma}{\sigma}}\},\\ \sup_{\beta\in\mathcal{A}}\|b_\beta\|_{L_\infty(Q_1)}\le N_0,\quad \sup_{\beta\in\mathcal{A}}|b_\beta(t,x)-b_\beta(t',x')|\le C_b\max\{|x-x'|^\gamma,|t-t'|^{\frac{\gamma}{\sigma}}\},\\ \text{where } N_0,C_f,C_b>0,\text{ and }\gamma\in(0,1).\end{gathered}\tag{1-5}$$

Recall that we assume $b_\beta\equiv0$ when $\sigma<1$, and $b_\beta=b(t,x)$ is independent of β when $\sigma=1$.

Theorem 1.3. *Let $\sigma \in (0,2)$, $0 < \lambda \le \Lambda < \infty$, and $\mathcal{A}$ be an index set. There exists $\hat\alpha$ depending only on d, λ, Λ and σ (uniform as $\sigma \to 2^-$) such that the following holds. Let $\gamma \in (0, \hat\alpha)$ such that $\sigma + \gamma < 2$ is not an integer. Assume for each $\beta \in \mathcal{A}$, K_β satisfies (1-2) when $\sigma = 1$, and the Hölder continuity assumptions (1-5) hold for all $(t,x), (t',x') \in Q_1$. Suppose $u \in C^{1+\frac{\gamma}{\sigma}, \sigma+\gamma}(Q_1) \cap C^{\frac{\gamma}{\sigma},\gamma}((-1,0)\times\mathbb{R}^d)$ is a solution of (1-1) in Q_1; then we have the a priori estimate*

$$[u]_{1+\frac{\gamma}{\sigma},\sigma+\gamma;Q_{1/2}} \le C\|u\|_{\frac{\gamma}{\sigma},\gamma;(-1,0)\times\mathbb{R}^d} + CC_f, \tag{1-6}$$

where $C > 0$ is a constant depending only on $d, \sigma, \gamma, \lambda, \Lambda, N_0$, and C_b.

The essential new part of Theorem 1.3 is for the case $\sigma = 1$. For $\sigma < 1$, Theorem 1.3 is just Theorem 1.1 in [Dong and Zhang 2016b]. Even though the Hölder continuity assumption appeared slightly differently, the proof in [Dong and Zhang 2016b] can be carried out with minimal modifications. For $\sigma > 1$, the drift is a lower-order perturbation and the conclusion can be proved without assuming $\sigma + \gamma < 2$ by using Theorem 1.1 in [Dong and Zhang 2016b] and interpolation inequalities.

In the case of the linear equation

$$\partial_t u = Lu + bDu + f, \tag{1-7}$$

the estimate (1-6) holds for all $\gamma \in (0,\sigma)$. Again, we assume $b \equiv 0$ when $\sigma < 1$.

Theorem 1.4. *Let $\sigma \in (0,2)$, $0 < \lambda \le \Lambda < \infty$, and $\gamma \in (0,\sigma)$ such that $\sigma + \gamma$ is not an integer. Assume K satisfies (1-2) when $\sigma = 1$, and the Hölder continuity assumptions (1-5) hold for all $(t,x), (t',x') \in Q_1$. Suppose $u \in C^{1+\frac{\gamma}{\sigma},\sigma+\gamma}(Q_1) \cap C^{\frac{\gamma}{\sigma},\gamma}((-1,0)\times\mathbb{R}^d)$ is a solution of (1-7) in Q_1; then we have the a priori estimate*

$$[u]_{1+\frac{\gamma}{\sigma},\sigma+\gamma;Q_{1/2}} \le C\|u\|_{\frac{\gamma}{\sigma},\gamma;(-1,0)\times\mathbb{R}^d)} + CC_f, \tag{1-8}$$

where $C > 0$ is a constant depending only on $d, \sigma, \gamma, \lambda, \Lambda, N_0$, and C_b.

It is natural to assume $\gamma < \sigma$ in Theorem 1.4, since (1-5) will imply that f is independent of t if $0 < \sigma < \gamma$. In many applications, a will be independent of t as well under the assumptions of (1-5) and $\sigma < \gamma$. Then, we can always differentiate (1-7) in t, and obtain higher-order regularity in t by applying the result of Theorem 1.4 above.

We are also interested in the linear equation (1-7) when K, b, and f are Dini continuous in x but only measurable in the time variable t, that is, they satisfy

$$\begin{aligned}
&\int_{B_{2r}\setminus B_r} |a(t,x,y) - a(t,x',y)|\,dy \le \Lambda r^d \omega_a(|x-x'|) \quad \text{for all } r > 0,\\
&\|f\|_{L_\infty(Q_1)} < \infty, \quad |f(t,x) - f(t,x')| \le \omega_f(|x-x'|),\\
&\|b\|_{L_\infty(Q_1)} \le N_0, \quad |b(t,x) - b(t,x')| \le \omega_b(|x-x'|),\\
&\text{where } N_0 > 0, \text{ and } \omega_a, \omega_b, \omega_f \text{ are all Dini functions.}
\end{aligned} \tag{1-9}$$

In Theorem 1.5 below, ω_u denotes the modulus of continuity of u in x uniform for all t; that is,

$$|u(t,x) - u(t,x')| \le \omega_u(|x-x'|) \quad \text{for all } (t,x), (t,x') \in (-1,0)\times\mathbb{R}^d.$$

Theorem 1.5. *Let $\sigma \in (0,2)$, $0<\lambda\le\Lambda<\infty$. Assume K satisfies (1-2) when $\sigma=1$, and the Dini continuity assumption (1-9) holds for all $(t,x),(t,x')\in Q_1$. Suppose $u\in C^{1,\sigma^+}(Q_1)$ is a solution of (1-7) in Q_1 and is Dini continuous in x in $(-1,0)\times\mathbb{R}^d$. Then we have the a priori estimate: for $\sigma\in(0,2)$,*

$$\|\partial_t u\|_{L_\infty(Q_{1/2})}+[u]^x_{\sigma;Q_{1/2}}\le C\sum_{j=0}^{\infty}(2^{-j\sigma}\omega_u(2^j)+\omega_u(2^{-j})+\omega_f(2^{-j})), \tag{1-10}$$

where $C>0$ is a constant depending only on d, σ, λ, Λ, N_0, ω_b, and ω_a. Moreover, when $\sigma\ne1$, we have

$$\sup_{(t_0,x_0)\in Q_{1/2}}[u]^x_{\sigma;Q_r(t_0,x_0)}\to 0\quad \text{as } r\to0$$

with a decay rate depending only on d, σ, λ, Λ, ω_a, ω_f, ω_u, N_0, and ω_b. When $\sigma=1$, Du is uniformly continuous in x in $Q_{1/2}$ with a modulus of continuity controlled by the quantities before. Also, $\partial_t u$ is uniformly continuous in x in $Q_{1/2}$ with a modulus of continuity controlled by d, σ, λ, Λ, ω_a, ω_f, ω_u, N_0, ω_b, and $\|u\|_{L_\infty}$.

If K, b, and f in (1-7) are Hölder continuous in x locally but only measurable in the time variable t, that is, they satisfy

$$\begin{gathered}\int_{B_{2r}\setminus B_r}|a(t,x,y)-a(t,x',y)|\,dy\le\Lambda r^d|x-x'|^\gamma\quad\text{for all } r>0,\\ \|f\|_{L_\infty(Q_1)}<\infty,\quad |f(t,x)-f(t,x')|\le C_f|x-x'|^\gamma,\\ \|b\|_{L_\infty(Q_1)}\le N_0,\quad |b(t,x)-b(t,x')|\le C_b|x-x'|^\gamma,\\ \text{where } N_0,C_a,C_b>0,\text{ and }\gamma\in(0,1),\end{gathered}\tag{1-11}$$

then we have:

Theorem 1.6. *Let $\sigma\in(0,2)$, $0<\lambda\le\Lambda<\infty$, and $\gamma\in(0,1)$ such that $\sigma+\gamma$ is not an integer. Assume K satisfies (1-2) when $\sigma=1$, and the Hölder continuity assumptions (1-11) hold for all $(t,x),(t,x')\in Q_1$. Suppose $u\in C^{1,\sigma+\gamma}(Q_1)\cap C^\gamma_x((-1,0)\times\mathbb{R}^d)$ is a solution of (1-7) in Q_1; then we have the a priori estimate*

$$[\partial_t u]^x_{\gamma;Q_{1/2}}+[u]^x_{\sigma+\gamma,;Q_{1/2}}\le C\|u\|^x_{\gamma;(-1,0)\times\mathbb{R}^d}+CC_f, \tag{1-12}$$

where $C>0$ is a constant depending only on d, σ,γ, λ, Λ, N_0, and C_b.

Note that here we assume $\gamma\in(0,1)$ for all $\sigma\in(0,2)$, since all the estimates only involve x. This theorem improves Theorem 1.1 in [Jin and Xiong 2015], which does not include drifts and requires the Hölder continuity of a and f in the time variable t as well. In the second-order case, similar results were obtained a long time ago by Knerr [1980/81] and Lieberman [1992].

A few remarks are in order.

Remark 1.7. It is evident that Theorems 1.1, 1.3, 1.4, 1.5, and 1.6 hold for corresponding elliptic equations as well.

Remark 1.8. Our proof does not tell whether the a priori estimates in Theorems 1.1 and 1.5 can be made uniformly bounded as $\sigma\to2^-$, even if we replace Λ by $(2-\sigma)\Lambda$ in both (1-5) and (1-9).

The ideas of our proofs are in the spirit of the approach first developed in [Campanato 1966], which has been used in [Dong and Zhang 2016a] for nonlocal fully nonlinear elliptic equations. A similar idea was also used in the literature to derive Cordes–Nirenberg-type estimates; see, e.g., [Nirenberg 1954]. Here, we adapt the methods in [Dong and Zhang 2016a] from elliptic settings to parabolic settings, with extra efforts to deal with the drift term especially when $\sigma = 1$ and some simplification of the proofs.

The key idea is that instead of estimating the C^σ seminorm of the solution, we construct and bound certain seminorms of the solution; see Lemma 2.1. When $\sigma < 1$, we define such a seminorm as a series of lower-order Hölder seminorms of u. In order for the nonlocal operator to be well defined, the solution needs to be smoother than C^σ. This motivates us to divide the integral domain into annuli, and use a lower-order seminorm to estimate the integral in each annulus. The proof of the case when $\sigma \geq 1$ is more involved mainly due to the fact that the series of lower-order Hölder seminorms of the solution itself is no longer sufficient to estimate the C^σ norm. Therefore, we need to subtract a polynomial from the solution in the construction of the seminorm. In some sense, the polynomial should be chosen to minimize the series. It turns out that when $\sigma \geq 1$, we can make use of the first-order Taylor's expansion of the mollification of the solution.

The organization of this paper is as follows. In the next section, we introduce some notation and preliminary results that are necessary in the proofs of our main theorems. In Section 3, we show the Dini estimates for nonlocal nonlinear parabolic equations in Theorem 1.1. In Section 4, we prove the Schauder estimates for equations with a drift in Theorems 1.3 and 1.4. The last section is devoted to linear parabolic equations with measurable coefficients in the time variable t, where Theorems 1.5 and 1.6 are proved.

2. Preliminaries

We will use the following notation:

- For $r > 0$, we set $Q_r(t_0, x_0) = (t_0 - r^\sigma, t_0] \times B_r(x_0)$ and $\hat{Q}_r(t_0, x_0) = (t_0 - r^\sigma, t_0 + r^\sigma) \times B_r(x_0)$, where $B_r(x_0) \subset \mathbb{R}^d$ is the ball of radius r centered at x_0. We write $Q_r = Q_r(0,0)$ for brevity.
- $\mathcal{P}_t$ (or $\mathcal{P}_x$) is the set of first-order polynomials in t (or x), respectively.
- $\mathcal{P}_1$ is the set of first-order polynomials in both t and x.
- For $\alpha, \beta > 0$,

$$[u]_{\alpha,\beta;Q_r(t_0,x_0)} = [u]_{C^{\alpha,\beta}_{t,x}(Q_r(t_0,x_0))},$$

$$[u]^x_{\beta;Q_r(t_0,x_0)} = \sup_{t\in(t_0-r^\sigma,t_0)} [u(t,\cdot)]_{C^\beta(B_r(x_0))},$$

$$[u]^t_{\alpha;Q_r(t_0,x_0)} = \sup_{x\in B_r(x_0)} [u(\cdot,x)]_{C^\alpha((t_0-r^\sigma,t_0))}.$$

If β (or α) is an integer, the above seminorms mean the Lipschitz norm of $D^{|\beta|-1}$ (or $\partial_t^{|\alpha|-1}$). If there is no subscript about the region where the norm is taken, then it means the whole domain where the function is defined (e.g., $\mathbb{R}^d$ or $(-t_0, 0] \times \mathbb{R}^d$ for some $t_0 > 0$).

- We say $u \in C^{1,\sigma^+}(Q_1)$ if $u \in C^{1,\sigma+\varepsilon}(Q_1)$ for some small $\varepsilon > 0$.

• We will also use the following Lipschitz–Zygmund seminorms. Let $\Omega \subset \mathbb{R}^d$ be a domain, $r > 0$, and $Q = (t_0 - r, t_0] \times \Omega$. For $\alpha, \beta \in (0, 2)$, we define

$$[u]^x_{\mathbf{\Lambda}^\beta(Q)} = \sup_{t\in(t_0-r^\sigma,t_0]} [u(t,\cdot)]_{\mathbf{\Lambda}^\beta(\Omega)} = \sup_{t\in(t_0-r^\sigma,t_0]} \sup_{\substack{x_1,x_2,x_3\in\Omega\\ x_1\neq x_3,\ x_1+x_3=2x_2}} \frac{|u(t,x_1)+u(t,x_3)-2u(t,x_2)|}{|x_1-x_2|^\alpha},$$

$$[u]^t_{\mathbf{\Lambda}^\alpha(Q)} = \sup_{x\in\Omega}[u(\cdot,x)]_{\mathbf{\Lambda}^\alpha((t_0-r^\sigma,t_0))} = \sup_{x\in\Omega} \sup_{\substack{t_1,t_2,t_3\in(t_0-r^\sigma,t_0]\\ t_1\neq t_3,\ t_1+t_3=2t_2}} \frac{|u(t_1,x)+u(t_3,x)-2u(t_2,x)|}{|t_1-t_2|^\alpha},$$

$$[u]_{\mathbf{\Lambda}^{\alpha,\beta}(Q)} = \sup_{\substack{(t_1,x_1),(t_2,x_2),(t_3,x_3)\in Q\\ (t_1,x_1)\neq(t_3,x_3),\ (t_1,x_1)+(t_3,x_3)=2(t_2,x_2)}} \frac{|u(t_1,x_1)+u(t_3,x_3)-2u(t_2,x_2)|}{|t_1-t_2|^\alpha+|x_1-x_2|^\beta}.$$

We will frequently use the identities

$$2^j(u(t,x+2^{-j}l)-u(t,x))-(u(t,x+l)-u(t,x)) = \sum_{k=1}^{j} 2^{k-1}\big(2u(t,x+2^{-k}l)-u(t,x+2^{-k+1}l)-u(t,x)\big), \tag{2-1}$$

$$2^j(u(t-2^{-j},x)-u(t,x))-(u(t-1,x)-u(t,x)) = \sum_{k=1}^{j} 2^{k-1}\big(2u(t-2^{-k},x)-u(t-2^{-k+1},x)-u(t,x)\big), \tag{2-2}$$

which hold for any unit vector $l \in \mathbb{R}^d$ and $j \in \mathbb{N}$.

Lemma 2.1. *Let $\alpha \in (0,\sigma)$ be a constant. Let Q be a convex cylinder such that $Q_{\frac{1}{2}} \subset Q \subset Q_1$.*

(i) *When $\sigma \in (0,1)$, we have*

$$[u]^x_{\sigma;Q} + \|\partial_t u\|_{L_\infty(Q)} \le C\sum_{k=0}^{\infty} 2^{k(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}(t_0,x_0)} + C\|u\|_{L_\infty(Q_{2^{1/\sigma}})}, \tag{2-3}$$

where C is a constant depending only on d, σ, and α. Moreover, the modulus of continuity of $\partial_t u$ is bounded by the tail of the summation on the right-hand side of (2-3).

(ii) *When $\sigma \in (1,2)$, we have*

$$[u]^x_{\sigma;Q} + \|\partial_t u\|_{L_\infty(Q)} + [Du]^t_{\frac{\sigma-1}{\sigma};Q} \le C\sum_{k=0}^{\infty} 2^{k(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q} \inf_{p\in\mathcal{P}_1} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}(t_0,x_0)} + C\|u\|_{L_\infty(Q_2)}, \tag{2-4}$$

where C is a constant depending on d, α, and σ. The modulus of continuity of $\partial_t u$ is bounded by the tail of the summation above.

(iii) *When $\sigma = 1$, we have*

$$\|Du\|_{L_\infty(Q)} + \|\partial_t u\|_{L_\infty(Q)} \le C \sum_{k=0}^{\infty} 2^{k(1-\alpha)} \sup_{(t_0,x_0)\in Q} \inf_{p\in\mathcal{P}_1} [u-p]_{\alpha,\alpha;Q_{2^{-k}}(t_0,x_0)} + C \sup_{\substack{(t,x),(t',x')\in Q_2\\ \max\{|t-t'|,|x-x'|\}=1}} |u(t,x)-u(t',x')|, \quad (2\text{-}5)$$

where C is a constant depending on d, α, and σ. The modulus of continuity of $\partial_t u$ and Du are bounded by the tail of the summation above.

Proof. We first prove the estimate of $\partial_t u$ for $\sigma \in (0,2)$ by showing that

$$\|\partial_t u\|_{L_\infty(Q)} \le C \sum_{k=0}^{\infty} 2^{k(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q} \inf_{p\in\mathcal{P}_t} [u-p]^t_{\frac{\alpha}{\sigma};Q_{2^{-k}}(t_0,x_0)} + 2 \sup_{(t_0,x_0)\in Q} |u(t_0-1,x_0)-u(t_0,x_0)|. \quad (2\text{-}6)$$

Indeed, from (2-2),

$$\begin{aligned} 2^j |u(t-2^{-j},x)-u(t,x)| &\le |u(t-1,x)-u(t,x)| + \sum_{k=1}^{\infty} 2^{k-1} |2u(t-2^{-k},x)-u(t-2^{-k+1},x)-u(t,x)| \\ &\le |u(t-1,x)-u(t,x)| + C \sum_{k=1}^{\infty} 2^{k(1-\frac{\alpha}{\sigma})} [u]^t_{\Lambda^{\alpha/\sigma}(Q_{2^{-k^*}}(t,x))}, \end{aligned} \quad (2\text{-}7)$$

where C only depends on σ and $k^* = [(k-1)/\sigma]$, i.e., the largest integer which is smaller than $(k-1)/\sigma$. The right-hand side of the above inequality is less than

$$\begin{aligned} &|u(t-1,x)-u(t,x)| + C \sum_{k=1}^{\infty} 2^{(k^*\sigma+\sigma)(1-\frac{\alpha}{\sigma})} [u]^t_{\Lambda^{\alpha/\sigma}(Q_{2^{-k^*}}(t,x))} \\ &\qquad \le |u(t-1,x)-u(t,x)| + C \sum_{k=1}^{\infty} 2^{k^*(\sigma-\alpha)} \inf_{p\in\mathcal{P}_t} [u-p]^t_{\Lambda^{\alpha/\sigma}(Q_{2^{-k^*}}(t,x))}. \end{aligned}$$

By using the definition of k^*, it is easy to see the second term on the right-hand side of the above inequality is bounded by

$$C \sum_{k=0}^{\infty} 2^{k(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q} \inf_{p\in\mathcal{P}_t} [u-p]^t_{\frac{\alpha}{\sigma};Q_{2^{-k}}(t_0,x_0)}.$$

Therefore, by sending $j \to \infty$ in (2-7), we prove that $\|\partial_t u\|_{L_\infty(Q)}$ is bounded by the right-hand side of (2-6). Since

$$\inf_{p\in\mathcal{P}_t} [u-p]^t_{\frac{\alpha}{\sigma};Q_{2^{-k}}(t_0,x_0)} \le \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}(t_0,x_0)},$$

$$\inf_{p\in\mathcal{P}_t} [u-p]^t_{\frac{\alpha}{\sigma};Q_{2^{-k}}(t_0,x_0)} \le \inf_{p\in\mathcal{P}_1} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}(t_0,x_0)},$$

the right-hand side of (2-6) is bounded by that of (2-3)–(2-5). We obtain the bound of $\|\partial_t u\|_{L_\infty(Q)}$.

Next, we bound the modulus of continuity of $\partial_t u$ in Q. Assume

$$|t-t'|+|x-x'|^\sigma \in [2^{-(i+1)}, 2^{-i}) \quad \text{for some } i \geq 1.$$

From (2-2), for any $j \geq i+1$,

$$2^j(u(t-2^{-j},x)-u(t,x)) - 2^i(u(t-2^{-i},x)-u(t,x)) = \sum_{k=i+1}^{j} 2^{k-1}(2u(t-2^{-k},x)-u(t-2^{-k+1},x)-u(t,x)),$$

and the same identity holds with (t',x') in place of (t,x). Then we have

$$\begin{aligned}|\partial_t u(t,x)-\partial_t u(t',x')| &= \lim_{j\to\infty} |2^j(u(t-2^{-j},x)-u(t,x)) - 2^j(u(t'-2^{-j},x')-u(t',x'))| \\ &\leq |2^i(u(t-2^{-i},x)-u(t,x)) - 2^i(u(t'-2^{-i},x')-u(t',x'))| \\ &\quad + C\sum_{k=i+1}^{\infty} \sup_{(t_0,x_0)\in Q} 2^{k(1-\frac{\alpha}{\sigma})}[u]^t_{\mathbf{\Lambda}^{\alpha/\sigma}(Q_{2^{-k^*}}(t_0,x_0))},\end{aligned}$$

where k^* is defined above. By the triangle inequality, the first term on the right-hand side is bounded by

$$2^i|u(t-2^{-i},x)+u(t',x')-2u(\bar t,\bar x)| + 2^i|u(t'-2^{-i},x')-2u(\bar t,\bar x)+u(t,x)|,$$

where $\bar t = (t+t'-2^{-i})/2$ and $\bar x = (x+x')/2$. This is further bounded by

$$2^{i(1-\frac{\alpha}{\sigma})} \sup_{(t_0,x_0)\in Q} [u]_{\mathbf{\Lambda}^{\alpha/\sigma,\alpha}(Q_{2^{-i^*}}(t_0,x_0))},$$

where $i^* = [(i-1)/\sigma]$. Therefore,

$$\begin{aligned}|\partial_t u(t,x)-\partial_t u(t',x')| &\leq C\sum_{k=i}^{\infty} \sup_{(t_0,x_0)\in Q} 2^{k(1-\frac{\alpha}{\sigma})}[u]_{\mathbf{\Lambda}^{\alpha/\sigma,\alpha}(Q_{2^{-i^*}}(t_0,x_0))} \\ &\leq C\sum_{k=i}^{\infty} \sup_{(t_0,x_0)\in Q} 2^{k(1-\frac{\alpha}{\sigma})} \inf_{p\in\mathcal{P}_1} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-i^*}}(t_0,x_0)},\end{aligned}$$

which, from the definition of i^*, converges to 0 as $i\to\infty$.

In the rest of the proof, we consider the three cases separately.

Case 1: $\sigma\in(0,1)$. The estimates of $[u]^x_\sigma$ are the same as [Dong and Zhang 2016a, Lemma 2.1] and we only provide a sketch here. Let $(t,x),(t,x')\in Q$ be two different points. Suppose $h := |x-x'| \in (0,1)$. Since

$$h^{-\sigma}|u(t,x')-u(t,x)| \leq \sup_{x\in Q} h^{\alpha-\sigma}[u(t,\cdot)]_{\alpha;B_h(x)},$$

by taking the supremum with respect to t, x, and x' for $h<1$ on both sides, we get

$$[u]^x_{\sigma;Q} \leq \sup_{(t_0,x_0)\in Q}\ \sup_{0<h<1} h^{\alpha-\sigma}[u]^x_{\alpha;Q_h(t_0,x_0)} \leq C\sum_{k=0}^{\infty} 2^{k(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q} [u]^x_{\alpha;Q_{2^{-k}}(t_0,x_0)}.$$

Notice that

$$[u]^x_{\alpha;Q_{2^{-k}}(t_0,x_0)} = \inf_{p\in\mathcal{P}_t}[u-p]^x_{\alpha;Q_{2^{-k}}(t_0,x_0)} \le \inf_{p\in\mathcal{P}_t}[u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}(t_0,x_0)}.$$

The proof of Case 1 is completed.

Case 2: $\sigma\in(1,2)$. Similar to the previous case, we only provide the sketch of the proof following that of [Dong and Zhang 2016a, Lemma 2.1]. Let $\ell\in\mathbb{R}^d$ be a unit vector and $\varepsilon\in\big(0,\frac{1}{16}\big)$ be a small constant to be specified later. For any two distinct points $(t,x),(t,x')\in Q$ such that $h=|x-x'|<\frac{1}{2}$, there exist $\bar{x},\bar{x}'\in Q$ such that $|x-\bar{x}|<\varepsilon h$, $\bar{x}+\varepsilon hl\in Q$, and $|x'-\bar{x}'|<\varepsilon h$, $\bar{x}'+\varepsilon hl\in Q$. By the triangle inequality,

$$h^{1-\sigma}|D_\ell u(t,x)-D_\ell u(t,x')|\le I_1+I_2+I_3, \tag{2-8}$$

where

$$\begin{aligned} I_1 &:= h^{1-\sigma}|D_\ell u(t,x)-(\varepsilon h)^{-1}(u(t,\bar{x}+\varepsilon h\ell)-u(t,\bar{x}))|,\\ I_2 &:= h^{1-\sigma}|D_\ell u(t,x')-(\varepsilon h)^{-1}(u(t,\bar{x}'+\varepsilon h\ell)-u(t,\bar{x}'))|,\\ I_3 &:= h^{1-\sigma}(\varepsilon h)^{-1}|(u(t,\bar{x}+\varepsilon h\ell)-u(t,\bar{x}))-(u(t,\bar{x}'+\varepsilon h\ell)-u(t,\bar{x}'))|. \end{aligned}$$

By the mean value theorem,

$$I_1+I_2\le 2^\sigma\varepsilon^{\sigma-1}[u]^x_{\sigma;Q}. \tag{2-9}$$

Now we choose and fix an ε sufficiently small depending only on σ such that $2^\sigma\varepsilon^{\sigma-1}\le\frac{1}{2}$. Using the triangle inequality, we have

$$I_3\le Ch^{-\sigma}\big(|u(t,\bar{x}+\varepsilon h\ell)+u(t,\bar{x}')-2u(t,\tilde{x})|+|u(t,\bar{x}'+\varepsilon h\ell)+u(t,\bar{x})-2u(t,\tilde{x})|\big),$$

where $\tilde{x}=(\bar{x}+\varepsilon h\ell+\bar{x}')/2$. Thus,

$$I_3\le Ch^{\alpha-\sigma}[u(t,\cdot)]^x_{\Lambda^\alpha(Q_h(t,\tilde{x}))}. \tag{2-10}$$

Combining (2-8), (2-9), and (2-10), we get

$$[u]^x_{\sigma;Q}\le C\sum_{k=0}^{\infty}2^{k(\sigma-\alpha)}\sup_{(t_0,x_0)\in Q}\inf_{p\in\mathcal{P}_x}[u-p]^x_{\alpha;Q_{2^{-k}}(t_0,x_0)}.$$

Because

$$\inf_{p\in\mathcal{P}_x}[u-p]^x_{\alpha;Q_{2^{-k}}(t_0,x_0)}\le\inf_{p\in\mathcal{P}_1}[u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}(t_0,x_0)},$$

we bound $[u]^x_{\sigma;Q}$ by the right-hand side of (2-4).

It follows from [Krylov 1996, Section 3.3] that $[Du]^t_{\frac{\sigma-1}{\sigma};Q}$ is bounded by $\|\partial_t u\|_{L_\infty(Q)}+[u]^x_{\sigma;Q}$. Therefore, (2-4) is proved.

Case 3: $\sigma=1$. We give the estimate of $\|Du\|_{L_\infty}$. It follows from (2-1) that

$$2^j|u(t,x+2^{-j}\ell)-u(t,x)|\le|u(t,x+\ell)-u(t,x)|+\sum_{k=1}^{j}2^{k(1-\alpha)}[u(t,\cdot)]^x_{\Lambda^\alpha(B_{2^{-k}}(x+2^{-k}\ell))}.$$

Taking $j \to \infty$, we obtain that

$$\|Du\|_{L_\infty} \le C \sum_{k=1}^{\infty} 2^{k(1-\alpha)} \sup_{(t_0,x_0)\in Q} \inf_{p\in\mathcal{P}_x} [u-p]^x_{\alpha;Q_{2^{-k}}(t_0,x_0)} + \sup_{\substack{(t,x),(t,x')\in Q_2\\ |x-x'|=1}} |u(t,x)-u(t,x')|.$$

The estimate of the continuity of Du is the same as $\partial_t u$, and thus omitted. □

Let η be a smooth nonnegative function in $\mathbb{R}$ with unit integral and vanishing outside $(0, 1)$. For $R > 0$ and $\sigma \in (0, 1)$, we define the mollification of u with respect to t as

$$u^{(R)}(t,x) = \int_{\mathbb{R}} u(t - R^\sigma s, x)\eta(s)\, ds.$$

For the case $\sigma \in [1, 2)$, we define $u^{(R)}$ differently by mollifying the x-variable as well. Let $\zeta \in C_0^\infty(B_1)$ be a radial nonnegative function with unit integral. For $R > 0$, we define

$$u^{(R)}(t,x) = \int_{\mathbb{R}^{d+1}} u(t - R^\sigma s, x - Ry)\eta(s)\zeta(y)\, dy\, ds.$$

The following lemma is for the case $\sigma \in (0, 1)$.

Lemma 2.2. *Let $\sigma \in (0,1)$, $\alpha \in (0,\sigma)$, and $R > 0$ be constants. Let $p_0 = p_0(t)$ be the first-order Taylor expansion of $u^{(R)}$ at the origin in t and $\tilde{u} = u - p_0$. Then for any integer $j \ge 0$, we have*

$$[\tilde{u}]_{\frac{\alpha}{\sigma},\alpha;(-R^\sigma,0)\times B_{2^j R}} \le C \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;(-R^\sigma,0)\times B_{2^j R}}, \tag{2-11}$$

where C is a constant only depending on d and α.

Proof. It is easily seen that $\tilde{u}$ is invariant up to a constant if we replace u by $u - p$ for any $p \in \mathcal{P}_t$. Thus to prove the lemma, we only need to bound the left-hand side of (2-11) by

$$C[u]_{\frac{\alpha}{\sigma},\alpha;(-R^\sigma,0)\times B_{2^j R}}.$$

Since $\tilde{u} = u - p(t)$, it suffices to observe that

$$[p]^t_{\frac{\alpha}{\sigma};(-R^\sigma,0)} = R^{\sigma-\alpha}|\partial_t u^{(R)}(0,0)| \le C[u]^t_{\frac{\alpha}{\sigma};Q_R}.$$ □

The following lemma is useful in dealing with the case $\sigma \in (1, 2)$.

Lemma 2.3. *Let $\alpha \in (0,1)$ and $\sigma \in (1,2)$ be constant. Then for any $u \in C^1$ and any cylinder Q, we have*

$$\sum_{k=0}^{\infty} 2^{k(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q} [u-p_0]^x_{\alpha;Q_{2^{-k}}(t_0,x_0)} \le C \sum_{k=0}^{\infty} 2^{k(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q} \inf_{p\in\mathcal{P}_x} [u-p]^x_{\alpha;Q_{2^{-k}}(t_0,x_0)}, \tag{2-12}$$

where p_0 is the first-order Taylor's expansion of u in the x-variable at (t_0, x_0), and $C > 0$ is a constant depending only on d, α, and σ.

Proof. Define

$$b_k := 2^{k(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q} \inf_{p\in\mathcal{P}_x} [u-p]^x_{\alpha;Q_{2^{-k}}(t_0,x_0)}.$$

Then for any $(t_0,x_0)\in Q$ and each $k=0,1,\dots$, there exists $\mathfrak{p}_k\in\mathcal{P}_x$ such that

$$[u-\mathfrak{p}_k]^x_{\alpha;Q_{2^{-k}}(t_0,x_0)}\le 2b_k2^{-k(\sigma-\alpha)}.$$

By the triangle inequality, for $k\ge 1$ we have

$$[\mathfrak{p}_{k-1}-\mathfrak{p}_k]^x_{\alpha;Q_{2^{-k}}(t_0,x_0)}\le 2b_k2^{-k(\sigma-\alpha)}+2b_{k-1}2^{-(k-1)(\sigma-\alpha)}. \tag{2-13}$$

It is easily seen that

$$[\mathfrak{p}_{k-1}-\mathfrak{p}_k]^x_{\alpha;Q_{2^{-k}}(t_0,x_0)}=|\nabla\mathfrak{p}_{k-1}-\nabla\mathfrak{p}_k|2^{-(k-1)(1-\alpha)},$$

which together with (2-13) implies

$$|\nabla\mathfrak{p}_{k-1}-\nabla\mathfrak{p}_k|\le C2^{-k(\sigma-1)}(b_{k-1}+b_k). \tag{2-14}$$

Since $\sum_k b_k<\infty$, from (2-14) we see that $\{\nabla\mathfrak{p}_k\}$ is a Cauchy sequence in $\mathbb{R}^d$. Let $q=q(t_0,x_0)\in\mathbb{R}^d$ be the limit, which clearly satisfies for each $k\ge 0$,

$$|q-\nabla\mathfrak{p}_k|\le C\sum_{j=k}^{\infty}2^{-j(\sigma-1)}b_j.$$

By the triangle inequality, we get

$$\begin{aligned}[u-q\cdot x]^x_{\alpha;Q_{2^{-k}}(t_0,x_0)}&\le[u-\mathfrak{p}_k]^x_{\alpha;Q_{2^{-k}}(t_0,x_0)}+[\mathfrak{p}_k-q\cdot x]^x_{\alpha;Q_{2^{-k}}(t_0,x_0)}\\&\le C2^{-k(1-\alpha)}\sum_{j=k}^{\infty}2^{-j(\sigma-1)}b_j\le C2^{-k(\sigma-\alpha)},\end{aligned} \tag{2-15}$$

which implies

$$\|u(t_0,\cdot)-u(t_0,x_0)-q\cdot(x-x_0)\|_{L_\infty(B_{2^{-k}}(x_0))}\le C2^{-k\sigma},$$

and thus $q=\nabla u(t_0,x_0)$. It then follows from (2-15) that

$$\begin{aligned}\sum_{k=0}^{\infty}2^{k(\sigma-\alpha)}\sup_{(t_0,x_0)\in Q}[u-p_0]^x_{\alpha;Q_{2^{-k}}(t_0,x_0)}&\le C\sum_{k=0}^{\infty}2^{k(\sigma-1)}\sum_{j=k}^{\infty}2^{-j(\sigma-1)}b_j\\&=C\sum_{j=0}^{\infty}2^{-j(\sigma-1)}b_j\sum_{k=0}^{j}2^{k(\sigma-1)}\le C\sum_{j=0}^{\infty}b_j.\end{aligned}$$

This completes the proof of (2-12). □

The last lemma in this section is for the case when $\sigma\in[1,2)$.

Lemma 2.4. *Let $\alpha\in(0,1)$, $\sigma\in[1,2)$, and $R>0$ be constants. Let $p_0=p_0(t,x)$ be the first-order Taylor's expansion of $u^{(R)}$ at the origin and $\tilde u=u-p_0$. Then for any integer $j\ge 0$, we have*

$$\sup_{\substack{(t,x),(t',x')\in(-R^\sigma,0)\times B_{2^jR}\\(t,x)\ne(t',x'),\,0\le|x-x'|<2R}}\frac{|\tilde u(t,x)-\tilde u(t',x')|}{|x-x'|^\alpha+|t-t'|^{\frac{\alpha}{\sigma}}}\le C\inf_{p\in\mathcal{P}_1}[u-p]_{\frac{\alpha}{\sigma},\alpha;(-R^\sigma,0)\times B_{2^jR}}, \tag{2-16}$$

where $C>0$ is a constant depending only on d, α, and σ.

Proof. It is easily seen that $\tilde{u}$ is invariant up to a constant if we replace u by $u-p$ for any $p \in \mathcal{P}_1$. Thus to show (2-16), we only bound the left-hand side of (2-16) by

$$C[u]_{\frac{\alpha}{\sigma},\alpha;(-R^\sigma,0)\times B_{2^jR}}.$$

Since $\tilde{u} = u - p_0$, it suffices to observe that for any two distinct $(t,x),(t',x') \in (-R^\sigma,0)\times B_{2^jR}$ such that $0 \le |x-x'| < 2R$,

$$\begin{aligned}|p_0(t,x)-p_0(t',x')| &\le |x-x'|\,|Du^{(R)}(0,0)| + |t-t'|\,|\partial_t u^{(R)}(0,0)|\\ &\le C|x-x'|R^{\alpha-1}[u]^x_{\alpha;Q_R} + C|t-t'|R^{\sigma(\frac{\alpha}{\sigma}-1)}[u]^t_{\frac{\alpha}{\sigma};Q_R}\\ &\le C(|x-x'|^\alpha + |t-t'|^{\frac{\alpha}{\sigma}})[u]_{\frac{\alpha}{\sigma},\alpha;Q_R}.\end{aligned}$$ □

3. Dini estimates for nonlocal nonlinear parabolic equations

The following proposition is a further refinement of [Dong and Zhang 2016b, Corollary 4.6].

Proposition 3.1. *Let $\sigma \in (0,2)$ and $0 < \lambda \le \Lambda$. Assume for any $\beta \in \mathcal{A}$ that K_β only depends on y. There is a constant $\hat{\alpha}$ depending on d, σ, λ, and Λ (uniformly as $\sigma \to 2^-$) so that the following holds. Let $\alpha \in (0,\hat{\alpha})$ such that $\sigma+\alpha$ is not an integer. Suppose $u \in C^{1+\frac{\alpha}{\sigma},\sigma+\alpha}(Q_1)\cap C^{\frac{\alpha}{\sigma},\alpha}((-1,0)\times\mathbb{R}^d)$ is a solution of*

$$\partial_t u = \inf_{\beta\in\mathcal{A}}(L_\beta u + f_\beta) \quad \text{in } Q_1.$$

Then,

$$[u]_{1+\frac{\alpha}{\sigma},\alpha+\sigma;Q_{1/2}} \le C\sum_{j=1}^{\infty} 2^{-j\sigma}M_j + C\sup_\beta [f_\beta]_{\frac{\alpha}{\sigma},\alpha;Q_1},$$

where

$$M_j = \sup_{\substack{(t,x),(t',x')\in(-1,0)\times B_{2^j}\\ (t,x)\neq(t',x'),\,0\le|x-x'|<2}} \frac{|u(t,x)-u(t',x')|}{|x-x'|^\alpha + |t-t'|^{\frac{\alpha}{\sigma}}},$$

and $C>0$ depends only on d,λ,Λ,α and σ, and is uniformly bounded as $\sigma\to 2^-$.

Proof. This follows from the proof of [Dong and Zhang 2016b, Corollary 4.6] by observing that in the estimate of $[h_\beta]_{\frac{\alpha}{\sigma},\alpha;Q_1}$, the term $[u]_{\frac{\alpha}{\sigma},\alpha;(-1,0)\times B_{2^j}}$ can be replaced by M_j. Moreover, by replacing u by $u-u(0,0)$, we see that

$$\|u\|_{\frac{\alpha}{\sigma},\alpha;(-1,0)\times B_2} \le C[u]_{\frac{\alpha}{\sigma},\alpha;(-1,0)\times B_2}.$$ □

In the rest of this section, we consider three cases separately.

The case $\sigma \in (0,1)$.

Proposition 3.2. *Suppose (1-1) is satisfied in $Q_{2^{1/\sigma}}$. Then under the conditions of Theorem 1.1, we have*

$$[u]^x_{\sigma;Q_{1/2}} + \|\partial_t u\|_{L_\infty;Q_{1/2}} \le C\|u\|_{\frac{\alpha}{\sigma},\alpha} + C\sum_{k=1}^{\infty}\omega_f(2^{-k}), \tag{3-1}$$

where $C>0$ is a constant depending only on d, λ, Λ, ω_a, and σ.

Proof. For $k \in \mathbb{N}$, let v be the solution of

$$\begin{cases} \partial_t v = \inf_{\beta\in\mathcal{A}}(L_\beta(0,0)v + f_\beta(0,0) - \partial_t p_0) & \text{in } Q_{2^{-k}}, \\ v = u - p_0(t) & \text{in } ((-2^{-k\sigma},0)\times B^c_{2^{-k}}) \cup (\{t=-2^{-k\sigma}\}\times B_{2^{-k}}), \end{cases}$$

where $L_\beta(0,0)$ is the operator with kernel $K_\beta(0,0,y)$, and $p_0(t)$ is the Taylor's expansion of $u^{(2^{-k})}$ in t at the origin. Then by Proposition 3.1 with scaling, we have

$$[v]_{1+\frac{\alpha}{\sigma},\alpha+\sigma;Q_{2^{-k-1}}} \le C\sum_{j=1}^{\infty} 2^{(k-j)\sigma} M_j + C2^{k\sigma}[v]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}}, \tag{3-2}$$

where $\alpha \in (0,\hat{\alpha})$ satisfying $\sigma+\alpha<1$,

$$M_j = \sup_{\substack{(t,x),(t',x')\in(-2^{-k\sigma},0)\times B_{2^{j-k}} \\ (t,x)\neq(t',x'),\, 0\le|x-x'|<2^{-k+1}}} \frac{|\tilde{u}(t,x)-\tilde{u}(t',x')|}{|x-x'|^\alpha + |t-t'|^{\frac{\alpha}{\sigma}}},$$

and $\tilde{u} = u - p_0$.

Let $k_0 \ge 1$ be an integer to be specified and $p_1 = p_1(t)$ be the Taylor's expansion of v in t at the origin. By the mean value formula,

$$\|v-p_1\|_{L_\infty(Q_{2^{-k-k_0}})} \le 2^{-(k+k_0)(\sigma+\alpha)}[v]_{1+\frac{\alpha}{\sigma},\sigma+\alpha;Q_{2^{-k-k_0}}},$$

and the interpolation inequality

$$[v-p_1]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k-k_0}}} \le C\big(2^{(k+k_0)\alpha}\|v-p_1\|_{L_\infty(Q_{2^{-k-k_0}})} + 2^{-(k+k_0)\sigma}[v-p_1]_{1+\frac{\alpha}{\sigma},\alpha+\sigma;Q_{2^{-k-k_0}}}\big),$$

we obtain

$$[v-p_1]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k-k_0}}} \le C2^{-(k+k_0)\sigma}[v]_{1+\frac{\alpha}{\sigma},\alpha+\sigma;Q_{2^{-k-k_0}}}.$$

From Lemma 2.2, we have

$$M_j \le C \inf_{p\in\mathcal{P}_t}[u-p]_{\frac{\alpha}{\sigma},\alpha;(-2^{-k\sigma},0)\times B_{2^{j-k}}} \le C[u]_{\frac{\alpha}{\sigma},\alpha;(-2^{-k\sigma},0)\times\mathbb{R}^d}. \tag{3-3}$$

These and (3-2) give

$$\begin{aligned}
&[v-p_1]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k-k_0}}} \\
&\quad\le C2^{-(k+k_0)\sigma}\sum_{j=1}^{k}2^{(k-j)\sigma}M_j + C2^{-(k+k_0)\sigma}[u]_{\frac{\alpha}{\sigma},\alpha;(-2^{-k\sigma},0)\times\mathbb{R}^d} + C2^{-k_0\sigma}[v]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}} \\
&\quad\le C2^{-(k+k_0)\sigma}\sum_{j=1}^{k}2^{(k-j)\sigma}M_j + C2^{-(k+k_0)\sigma}[u]_{\frac{\alpha}{\sigma},\alpha} + C2^{-k_0\sigma}[v]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}}.
\end{aligned} \tag{3-4}$$

Next, $w := u - p_0 - v$ satisfies

$$\begin{cases} w_t - \mathcal{M}^+ w \le C_k & \text{in } Q_{2^{-k}}, \\ w_t - \mathcal{M}^- w \ge -C_k & \text{in } Q_{2^{-k}}, \\ w = 0 & \text{in } ((-2^{-k\sigma},0)\times B^c_{2^{-k}}) \cup (\{t=-2^{-k\sigma}\}\times B_{2^{-k}}), \end{cases} \tag{3-5}$$

where $\mathcal{M}^+$ and $\mathcal{M}^-$ are the Pucci extremal operators, see, e.g., [Dong and Zhang 2016b], and

$$C_k = \sup_{\beta\in\mathcal{A}} \| f_\beta - f_\beta(0,0) + (L_\beta - L_\beta(0,0))u\|_{L_\infty(Q_{2^{-k}})}.$$

It is easily seen that

$$C_k \le \omega_f(2^{-k}) + C\omega_a(2^{-k})\Big(\sup_{(t_0,x_0)\in Q_{2^{-k}}} \sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)}[u]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty}\Big).$$

Then by the Hölder estimate [Dong and Zhang 2016b, Lemma 2.5], we have

$$\begin{aligned}[][w]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}} &\le C2^{-k(\sigma-\alpha)}C_k \\ &\le C2^{-k(\sigma-\alpha)}\Big[\omega_f(2^{-k})+\omega_a(2^{-k})\Big(\sup_{(t_0,x_0)\in Q_{2^{-k}}} \sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)}[u]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)}+\|u\|_{L_\infty}\Big)\Big]\end{aligned} \tag{3-6}$$

for some $\alpha > 0$. This α can be the same as the one in (3-2) since α is always small. By the triangle inequality and Lemma 2.2 with $j = 0$

$$\begin{aligned}[][v]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}} &\le [w]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}} + [u-p_0]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}} \\ &\le [w]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}} + C \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}}.\end{aligned} \tag{3-7}$$

Combining (3-4), (3-6), (3-3), and (3-7) yields

$$\begin{aligned}&2^{(k+k_0)(\sigma-\alpha)}[u-p_0-p_1]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k-k_0}}} \\ &\quad= 2^{(k+k_0)(\sigma-\alpha)}[w+v-p_1]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k-k_0}}} \\ &\quad\le C2^{-(k+k_0)\alpha}\sum_{j=1}^{k} 2^{(k-j)\sigma} \inf_{p\in\mathcal{P}_t}[u-p]_{\frac{\alpha}{\sigma},\alpha;(-2^{-k\sigma},0)\times B_{2^{j-k}}} \\ &\qquad + C2^{-(k+k_0)\alpha}[u]_{\frac{\alpha}{\sigma},\alpha} + C2^{-k_0\alpha+k(\sigma-\alpha)} \inf_{p\in\mathcal{P}_t}[u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}} + C2^{k_0(\sigma-\alpha)}\omega_f(2^{-k}) \\ &\qquad + C2^{k_0(\sigma-\alpha)}\omega_a(2^{-k})\Big(\sup_{(t_0,x_0)\in Q_{2^{-k}}} \sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)}[u]^x_{\alpha;Q_{2^{-j}(t_0,x_0)}} + \|u\|_{L_\infty}\Big).\end{aligned} \tag{3-8}$$

Let $\ell_0 \ge 1$ be an integer such that

$$\frac{1}{2^\sigma} + \sum_{l=\ell_0+1}^{\infty} \frac{1}{2^{l\sigma}} \le 1.$$

Set $Q^{\ell_0} = Q_{\frac{1}{2}}$ and for $l = \ell_0+1, \ell_0+2, \dots,$ we define

$$Q^l := \Big(-\frac{1}{2^\sigma} - \sum_{j=\ell_0+1}^{l} \frac{1}{2^{j\sigma}}, 0\Big] \times \Big\{x : |x| < \frac{1}{2} + \sum_{j=\ell_0+1}^{l} \frac{1}{2^j}\Big\}.$$

The choice of ℓ_0 will ensure that $Q^l \subset Q_1$ for all $l \geq \ell_0$, and the definition of Q^l will ensure that for $l \geq \ell_0,\ k \geq l+1$, there holds

$$Q^l + Q_{2^{-k}}(t_0, x_0) \subset Q^{l+1} \quad \text{for all } (t_0, x_0) \in Q^l.$$

By translation of the coordinates, from (3-8) we have for any $l \geq \ell_0$ and $k \geq l+1$,

$$\begin{aligned}
&2^{(k+k_0)(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} [u-p_0-p_1]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k-k_0}}(t_0,x_0)} \\
&\quad \leq C 2^{-(k+k_0)\alpha} \sup_{(t_0,x_0)\in Q^l} \sum_{j=0}^{k} 2^{(k-j)\sigma} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;(t_0-2^{-k\sigma},t_0)\times B_{2^{j-k}}(x_0)} + C 2^{-(k+k_0)\alpha}[u]_{\frac{\alpha}{\sigma},\alpha} \\
&\qquad + C 2^{k_0(\sigma-\alpha)} \Big[\omega_f(2^{-k}) \\
&\qquad\qquad + \omega_a(2^{-k}) \Big(\sup_{(t_0,x_0)\in Q^{l+1}} \sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty} \Big) \Big]. \qquad (3\text{-}9)
\end{aligned}$$

Then we take the sum (3-9) in $k = l+1, l+2, \ldots$ to obtain

$$\begin{aligned}
&\sum_{k=l+1}^{\infty} 2^{(k+k_0)(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k-k_0}}(t_0,x_0)} \\
&\quad \leq C \sum_{k=l+1}^{\infty} 2^{-(k+k_0)\alpha} \sup_{(t_0,x_0)\in Q^l} \sum_{j=0}^{k} 2^{(k-j)\sigma} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;(t_0-2^{-k\sigma},t_0)\times B_{2^{j-k}}(x_0)} \\
&\qquad + C 2^{-(l+k_0)\alpha}[u]_{\frac{\alpha}{\sigma},\alpha} + C 2^{k_0(\sigma-\alpha)} \sum_{k=l+1}^{\infty} \omega_f(2^{-k}) \\
&\qquad + C 2^{k_0(\sigma-\alpha)} \sum_{k=l+1}^{\infty} \omega_a(2^{-k}) \Big(\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty} \Big).
\end{aligned}$$

By switching the order of summations and then replacing k by $k+j$, the first term on the right-hand side is bounded by

$$\begin{aligned}
&C 2^{-k_0\alpha} \sum_{j=0}^{\infty} 2^{-j\sigma} \sum_{k=j}^{\infty} 2^{k(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;(t_0-2^{-k\sigma},t_0)\times B_{2^{j-k}}(x_0)} \\
&\qquad \leq C 2^{-k_0\alpha} \sum_{j=0}^{\infty} 2^{-j\alpha} \sum_{k=0}^{\infty} 2^{k(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;(t_0-2^{-k\sigma},t_0)\times B_{2^{-k}}(x_0)} \\
&\qquad \leq C 2^{-k_0\alpha} \sum_{k=0}^{\infty} 2^{k(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}(t_0,x_0)}.
\end{aligned}$$

With the above inequality, we have

$$\begin{aligned}
&\sum_{k=l+1}^{\infty} 2^{(k+k_0)(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k-k_0}}(t_0,x_0)} \\
&\quad\le C2^{-k_0\alpha}\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j}}(t_0,x_0)} \\
&\qquad + C2^{-(l+k_0)\alpha}[u]_{\frac{\alpha}{\sigma},\alpha} + C2^{k_0(\sigma-\alpha)}\sum_{k=l+1}^{\infty}\omega_f(2^{-k}) \\
&\qquad + C2^{k_0(\sigma-\alpha)}\sum_{k=l+1}^{\infty}\omega_a(2^{-k})\Bigl(\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty}\Bigr).
\end{aligned}$$

The bound above, together with the obvious inequality

$$\sum_{j=0}^{l+k_0} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j}}(t_0,x_0)} \le C2^{(l+k_0)(\sigma-\alpha)}[u]_{\frac{\alpha}{\sigma},\alpha},$$

implies

$$\begin{aligned}
&\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j}}(t_0,x_0)} \\
&\quad\le C2^{-k_0\alpha}\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j}}(t_0,x_0)} \\
&\qquad + C2^{(l+k_0)(\sigma-\alpha)}[u]_{\frac{\alpha}{\sigma},\alpha} + C2^{k_0(\sigma-\alpha)}\sum_{k=l}^{\infty}\omega_f(2^{-k}) \\
&\qquad + C2^{k_0(\sigma-\alpha)}\sum_{k=l}^{\infty}\omega_a(2^{-k})\Bigl(\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty}\Bigr).
\end{aligned}$$

By first choosing k_0 sufficiently large, and then ℓ_0 sufficiently large (recalling that $l \ge \ell_0$), we get

$$\begin{aligned}
&\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}(t_0,x_0)} \\
&\quad\le \frac{1}{4}\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}(t_0,x_0)} + C2^{(l+k_0)(\sigma-\alpha)}\|u\|_{\frac{\alpha}{\sigma},\alpha} + C\sum_{k=1}^{\infty}\omega_f(2^{-k}).
\end{aligned}$$

Multiplying both sides by 4^{-l}, taking the sum in l, we have

$$4^{-l}\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}(t_0,x_0)} \le C\|u\|_{\frac{\alpha}{\sigma},\alpha} + C\sum_{k=1}^{\infty}\omega_f(2^{-k}). \tag{3-10}$$

This, together with Lemma 2.1(i) and the fact that $Q^{\ell_0} = Q_{\frac{1}{2}}$, gives (3-1) and the continuity of $\partial_t u$. □

The case when $\sigma \in (1, 2)$.

Proposition 3.3. *Suppose (1-1) is satisfied in Q_2. Then under the conditions of Theorem 1.1, we have for $\sigma \in (1,2)$*

$$[u]^x_{\sigma;Q_{1/2}} + [Du]^t_{\frac{\sigma-1}{\sigma};Q_{1/2}} + \|\partial_t u\|_{L_\infty(Q_{1/2})} \le C\|u\|_{\frac{\alpha}{\sigma},\alpha} + C\sum_{k=1}^{\infty}\omega_f(2^{-k}), \tag{3-11}$$

where $C > 0$ is a constant depending only on d, λ, Λ, ω_a, ω_b, N_0, and σ.

Proof. For $k \in \mathbb{N}$, let v_M be the solution of

$$\begin{cases} \partial_t v_M = \inf_{\beta\in\mathcal{A}}\big(L_\beta(0,0)v_M + f_\beta(0,0) + b_\beta(0,0)Du(0,0) - \partial_t p_0\big) & \text{in } Q_{2^{-k}}, \\ v_M = g_M & \text{in } ((-2^{-k\sigma},0)\times B^c_{2^{-k}})\cup(\{t=-2^{-k\sigma}\}\times B_{2^{-k}}), \end{cases}$$

where $M \ge 2\|u-p_0\|_{L_\infty(Q_{2^{-k}})}$ is a constant to be specified later,

$$g_M = \max(\min(u-p_0, M), -M),$$

and $p_0 = p_0(t,x)$ is the first-order Taylor's expansion of $u^{(2^{-k})}$ at the origin. By Proposition 3.1, we have

$$[v_M]_{1+\frac{\alpha}{\sigma},\alpha+\sigma;Q_{2^{-k-1}}} \le C\sum_{j=1}^{\infty} 2^{(k-j)\sigma}M_j + C2^{k\sigma}[v_M]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}},$$

where $\alpha \in (0, \min\{\hat{\alpha}, (\sigma-1)/2, 2-\sigma\})$ and

$$M_j = \sup_{\substack{(t,x),(t',x')\in(-2^{-k\sigma},0)\times B_{2^{j-k}} \\ (t,x)\neq(t',x'),\, 0\le|x-x'|<2^{-k+1}}} \frac{|u(t,x)-p_0(t,x)-u(t',x')+p_0(t',x')|}{|t-t'|^{\frac{\alpha}{\sigma}}+|x-x'|^\alpha}.$$

From Lemma 2.4 with $\sigma \in (1,2)$, it follows

$$M_j \le C\inf_{p\in\mathcal{P}_1}[u-p]_{\frac{\alpha}{\sigma},\alpha;(-2^{-k\sigma},0)\times B_{2^{j-k}}}. \tag{3-12}$$

In particular, for $j > k$, we have

$$M_j \le C[u]_{\frac{\alpha}{\sigma},\alpha;(-2^{-k\sigma},0)\times\mathbb{R}^d},$$

and thus,

$$\begin{aligned}[v_M]_{1+\frac{\alpha}{\sigma},\alpha+\sigma;Q_{2^{-k-1}}} &\le C\sum_{j=1}^{\infty}2^{(k-j)\sigma}M_j + C2^{k\sigma}[v_M]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}} \\ &\le C\sum_{j=1}^{k}2^{(k-j)\sigma}M_j + C[u]_{\frac{\alpha}{\sigma},\alpha;(-2^{-k\sigma},0)\times\mathbb{R}^d} + C2^{k\sigma}[v_M]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}}. \end{aligned} \tag{3-13}$$

From (3-13), and the mean value formula (recalling that $\alpha < 2-\sigma$),

$$\|v_M - p_1\|_{L_\infty(Q_{2^{-k-k_0}})} \le C2^{-(k+k_0)(\sigma+\alpha)} \sum_{j=1}^{k} 2^{(k-j)\sigma} M_j + C2^{-(k+k_0)(\sigma+\alpha)}[u]_{\frac{\alpha}{\sigma},\alpha;(-2^{-k\sigma},0)\times\mathbb{R}^d} + C2^{-k\alpha-k_0(\sigma+\alpha)}[v_M]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}},$$

where p_1 is the first-order Taylor's expansion of v_M at the origin. The above inequality, (3-13), and the interpolation inequality imply

$$[v_M - p_1]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k-k_0}}} \le C2^{-(k+k_0)\sigma} \sum_{j=1}^{k} 2^{(k-j)\sigma} M_j + C2^{-(k+k_0)\sigma}[u]_{\frac{\alpha}{\sigma},\alpha;(-2^{-k\sigma},0)\times\mathbb{R}^d} + C2^{-k_0\sigma}[v_M]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}}. \quad (3\text{-}14)$$

Next $w_M := g_M - v_M$ satisfies

$$\begin{cases} \partial_t w_M \le \mathcal{M}^+ w_M + h_M + C_k & \text{in } Q_{2^{-k}}, \\ \partial_t w_M \ge \mathcal{M}^- w_M + \hat{h}_M - C_k & \text{in } Q_{2^{-k}}, \\ \quad w_M = 0 & \text{in } ((-2^{-k\sigma},0)\times B^c_{2^{-k}}) \cup (\{t=-2^{-k\sigma}\}\times B_{2^{-k}}), \end{cases}$$

where

$$h_M := \mathcal{M}^+(u - p_0 - g_M), \quad \hat{h}_M := \mathcal{M}^-(u - p_0 - g_M).$$

Here

$$C_k = \sup_{\beta\in\mathcal{A}} \big\| f_\beta - f_\beta(0,0) + b_\beta Du - b_\beta(0,0)Du(0,0) + (L_\beta - L_\beta(0,0))u \big\|_{L_\infty(Q_{2^{-k}})}.$$

It follows easily that

$$C_k \le \omega_f(2^{-k}) + \omega_b(2^{-k})\|Du\|_{L_\infty(Q_{2^{-k}})} + \sup_\beta \|b_\beta\|_{L_\infty} 2^{-k\alpha}[Du]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}} + C\omega_a(2^{-k})\Big(\sup_{(t_0,x_0)\in Q_{2^{-k}}} \sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)}[u - p_{t_0,x_0}]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)} + \|Du\|_{L_\infty(Q_{2^{-k}})} + \|u\|_{L_\infty}\Big),$$

where $p_{t_0,x_0} = p_{t_0,x_0}(x)$ is the first-order Taylor's expansion of u with respect to x at (t_0,x_0). From Lemma 2.3, we obtain

$$C_k \le \omega_f(2^{-k}) + \omega_b(2^{-k})\|Du\|_{L_\infty(Q_{2^{-k}})} + \sup_\beta \|b_\beta\|_{L_\infty} 2^{-k\alpha}[Du]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}} + C\omega_a(2^{-k})\Big(\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q_{2^{-k}}} \inf_{p\in\mathcal{P}_x} [u-p]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)} + \|Du\|_{L_\infty(Q_{2^{-k}})} + \|u\|_{L_\infty}\Big).$$

By the dominated convergence theorem, it is easy to see that

$$\|h_M\|_{L_\infty(Q_{2^{-k}})}, \quad \|\hat{h}_M\|_{L_\infty(Q_{2^{-k}})} \to 0 \quad \text{as } M\to\infty.$$

Thus similar to (3-6), choosing M sufficiently large so that

$$\|h_M\|_{L_\infty(Q_{2^{-k}})}, \quad \|\hat{h}_M\|_{L_\infty(Q_{2^{-k}})} \le \tfrac{1}{2}C_k,$$

we have

$$[w_M]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}} \le C2^{-k(\sigma-\alpha)}\Big[\omega_f(2^{-k})+(\omega_b(2^{-k})+\omega_a(2^{-k}))\|Du\|_{L_\infty(Q_{2^{-k}})}+2^{-k\alpha}[Du]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}} + \omega_a(2^{-k})\Big(\sum_{j=0}^{\infty}2^{j(\sigma-\alpha)}\sup_{(t_0,x_0)\in Q_{2^{-k}}}\inf_{p\in\mathcal{P}_x}[u-p]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)}+\|u\|_{L_\infty}\Big)\Big]. \tag{3-15}$$

Clearly,

$$\inf_{p\in\mathcal{P}_x}[u-p]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)} \le \inf_{p\in\mathcal{P}_1}[u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j}}(t_0,x_0)}. \tag{3-16}$$

From the triangle inequality and Lemma 2.4 with $j=0$,

$$[v_M]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}} \le [w_M]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}}+[u-p_0]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}} \le [w_M]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}}+C\inf_{p\in\mathcal{P}_1}[u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}}.$$

For all $l=1,2,\dots$, we define $Q^l=Q_{1-2^{-l}}$. Combining (3-14), (3-15) with (3-16), and (3-12), similar to (3-9), we get that for all $l\ge 1$ and $k\ge l+1$,

$$\begin{aligned}
&2^{(k+k_0)(\sigma-\alpha)}\sup_{(t_0,x_0)\in Q^l}\inf_{p\in\mathcal{P}_1}[u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-(k_0+k)}}(t_0,x_0)}\\
&\le C2^{-(k+k_0)\alpha}\sup_{(t_0,x_0)\in Q^l}\sum_{j=0}^{k}2^{(k-j)\sigma}\inf_{p\in\mathcal{P}_1}[u-p]_{\frac{\alpha}{\sigma},\alpha;(t_0-2^{-k\sigma},t_0)\times B_{2^{j-k}}(x_0)}\\
&\quad+C2^{-(k+k_0)\alpha}[u]_{\frac{\alpha}{\sigma},\alpha}+C2^{-k\alpha+k_0(\sigma-\alpha)}[Du]_{\frac{\alpha}{\sigma},\alpha;Q^{l+1}}\\
&\quad+C2^{k_0(\sigma-\alpha)}\Big[\omega_f(2^{-k})+(\omega_b(2^{-k})+\omega_a(2^{-k}))\|Du\|_{L_\infty(Q^{l+1})}\\
&\qquad+\omega_a(2^{-k})\Big(\sum_{j=0}^{\infty}2^{j(\sigma-\alpha)}\sup_{(t_0,x_0)\in Q^{l+1}}\inf_{p\in\mathcal{P}_1}[u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j}}(t_0,x_0)}+\|u\|_{L_\infty}\Big)\Big].
\end{aligned}\tag{3-17}$$

Summing the above inequality in $k=l+1,l+2,\dots$ as before, we obtain

$$\begin{aligned}
&\sum_{k=l+1}^{\infty}2^{(k+k_0)(\sigma-\alpha)}\sup_{(t_0,x_0)\in Q^l}\inf_{p\in\mathcal{P}_1}[u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k-k_0}}(t_0,x_0)}\\
&\le C2^{-k_0\alpha}\sum_{j=0}^{\infty}2^{j(\sigma-\alpha)}\sup_{(t_0,x_0)\in Q^{l+1}}\inf_{p\in\mathcal{P}_1}[u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j}}(t_0,x_0)}\\
&\quad+C2^{-(k_0+l)\alpha}[u]_{\frac{\alpha}{\sigma},\alpha}+C2^{k_0(\sigma-\alpha)}\sum_{k=l+1}^{\infty}2^{-k\alpha}[Du]_{\frac{\alpha}{\sigma},\alpha;Q^{l+1}}\\
&\quad+C2^{k_0(\sigma-\alpha)}\sum_{k=l+1}^{\infty}\big(\omega_f(2^{-k})+(\omega_b(2^{-k})+\omega_a(2^{-k}))\|Du\|_{L_\infty(Q^{l+1})}\big)\\
&\quad+C2^{k_0(\sigma-\alpha)}\sum_{k=l+1}^{\infty}\omega_a(2^{-k})\Big(\sum_{j=0}^{\infty}2^{j(\sigma-\alpha)}\sup_{(t_0,x_0)\in Q^{l+1}}\inf_{p\in\mathcal{P}_1}[u-p]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)}+\|u\|_{L_\infty}\Big),
\end{aligned}\tag{3-18}$$

and

$$\begin{aligned}
&\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_1} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j}}(t_0,x_0)} \\
&\quad \le C2^{(k_0+l)(\sigma-\alpha)}[u]_{\frac{\alpha}{\sigma},\alpha} + C2^{-k_0\alpha}\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \inf_{p\in\mathcal{P}_1} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j}}(t_0,x_0)} \\
&\qquad + C2^{k_0(\sigma-\alpha)-l\alpha}[Du]_{\frac{\alpha}{\sigma},\alpha;Q^{l+1}} \\
&\qquad + C2^{k_0(\sigma-\alpha)} \sum_{k=l+1}^{\infty} \big(\omega_f(2^{-k}) + (\omega_b(2^{-k})+\omega_a(2^{-k}))\|Du\|_{L_\infty(Q^{l+1})}\big) \\
&\qquad + C2^{k_0(\sigma-\alpha)} \sum_{k=l+1}^{\infty} \omega_a(2^{-k})\bigg(\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \inf_{p\in\mathcal{P}_1} [u-p]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty}\bigg).
\end{aligned}$$

By choosing k_0 and l sufficiently large, and using (2-4) and interpolation inequalities (recalling that $\alpha < (\sigma-1)/2$), we obtain

$$\begin{aligned}
&\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_1} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j}}(t_0,x_0)} \\
&\quad \le \frac{1}{4}\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \inf_{p\in\mathcal{P}_1} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j}}(t_0,x_0)} + C2^{(k_0+l)(\sigma-\alpha)}\|u\|_{\frac{\alpha}{\sigma},\alpha} + C\sum_{k=1}^{\infty}\omega_f(2^{-k}).
\end{aligned}$$

Therefore,

$$\frac{1}{4^l}\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_1} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j}}(t_0,x_0)} \le C\|u\|_{\frac{\alpha}{\sigma},\alpha} + C\sum_{k=1}^{\infty}\omega_f(2^{-k}), \tag{3-19}$$

which together with Lemma 2.1(ii) gives (3-11) and the continuity of $\partial_t u$. □

The case when $\sigma = 1$.

Proposition 3.4. *Suppose (1-1) is satisfied in Q_2. Then under the conditions of Theorem 1.1,*

$$\|Du\|_{L_\infty(Q_{1/2})} + \|\partial_t u\|_{L_\infty(Q_{1/2})} \le C\|u\|_{\alpha,\alpha} + C\sum_{k=1}^{\infty}\omega_f(2^{-k}), \tag{3-20}$$

where $C > 0$ is a constant depending only on d, λ, Λ, N_0, ω_a, and ω_b.

Proof. Set $b_0 = b(0,0)$ and we define

$$\hat{u}(t,x) = u(t,x-b_0t), \quad \hat{f}_\beta(t,x) = f_\beta(t,x-b_0t), \quad \text{and} \quad \hat{b}(t,x) = b(t,x-b_0t).$$

It is easy to see that in Q_δ for some $\delta > 0$,

$$\partial_t\hat{u}(t,x) = \partial_t u(t,x-b_0t) - b_0\nabla u(t,x-b_0t),$$

and for $(t,x)\in Q_{2^{-k}}$,

$$|\hat f_\beta(t,x)-\hat f_\beta(0,0)|\le \omega_f((1+N_0)2^{-k}),$$
$$|\hat b-b_0|\le \omega_b((1+N_0)2^{-k}).$$

It follows immediately that

$$\hat u_t=\inf_\beta(\hat L_\beta \hat u+\hat f_\beta+(\hat b-b_0)\nabla \hat u), \tag{3-21}$$

where $\hat L$ is the operator with kernel $a(t,x-b_0t,y)|y|^{-d-\sigma}$. Furthermore,

$$\|Du\|_{L_\infty}+\|\partial_t u\|_{L_\infty}\le (1+N_0)(\|D\hat u\|_{L_\infty}+\|\partial_t \hat u\|_{L_\infty}).$$

Therefore, it is sufficient to bound $\hat u$. In the rest of the proof, we estimate the solution to (3-21) and abuse the notation to use u instead of $\hat u$ for simplicity. By scaling, translation and covering arguments, we also assume u satisfies the equation in Q_2.

The proof is similar to the case $\sigma\in(1,2)$ and we indeed proceed as in the previous case. Take p_0 to be the first-order Taylor's expansion of $u^{(2^{-k})}$ at the origin. We also assume that the solution v to the equations

$$\begin{cases}\partial_t v=\inf_{\beta\in\mathcal A}(L_\beta(0,0)v+f_\beta(0,0)-\partial_t p_0) & \text{in } Q_{2^{-k}},\\ \quad v=u-p_0 & \text{in } ((-2^{-k\sigma},0)\times B^c_{2^{-k}})\cup(\{t=-2^{-k\sigma}\}\times B_{2^{-k}})\end{cases}$$

exists without carrying out another approximation argument. By Proposition 3.1 and Lemma 2.4 with $\sigma=1$,

$$\begin{aligned}[v]_{1+\alpha,1+\alpha;Q_{2^{-k-1}}}&\le C\sum_{j=1}^{\infty}2^{k-j}M_j+C2^k[v]_{\alpha,\alpha;Q_{2^{-k}}}\\ &\le C\sum_{j=1}^{\infty}2^{k-j}\inf_{p\in\mathcal P_1}[u-p]_{\alpha,\alpha;(-2^{-k},0)\times B_{2^{j-k}}}+C2^k[v]_{\alpha,\alpha;Q_{2^{-k}}}\\ &\le C\sum_{j=1}^{k}2^{k-j}\inf_{p\in\mathcal P_1}[u-p]_{\alpha,\alpha;(-2^{-k},0)\times B_{2^{j-k}}}+C[u]_{\alpha,\alpha}+C2^k[v]_{\alpha,\alpha;Q_{2^{-k}}}.\end{aligned} \tag{3-22}$$

From (3-22) and the interpolation inequality, we obtain

$$\begin{aligned}&[v-p_1]_{\alpha,\alpha;Q_{2^{-k-k_0}}}\\ &\le C2^{-(k+k_0)}\sum_{j=1}^{k}2^{k-j}\inf_{p\in\mathcal P_1}[u-p]_{\alpha,\alpha;(-2^{-k},0)\times B_{2^{j-k}}}+C2^{-k_0}[v]_{\alpha,\alpha;Q_{2^{-k}}}+C2^{-(k+k_0)}[u]_{\alpha,\alpha},\end{aligned} \tag{3-23}$$

where p_1 is the first-order Taylor's expansion of v at the origin. Next $w:=u-p_0-v$ satisfies (3-5), where by the cancellation property,

$$\begin{aligned}C_k&\le \omega_f((1+N_0)2^{-k})+\omega_b((1+N_0)2^{-k})\|Du\|_{L_\infty(Q_{2^{-k}})}\\ &\quad+C\omega_a((1+N_0)2^{-k})\Big(\sup_{(t_0,x_0)\in Q_{2^{-k}}}\sum_{j=0}^{\infty}2^{j(1-\alpha)}\inf_{p\in\mathcal P_x}[u-p]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)}+\|u\|_{L_\infty}\Big).\end{aligned}$$

Clearly, for any $r \geq 0$,

$$\omega_\bullet((1+N_0)r) \leq (2+N_0)\omega_\bullet(r).$$

Therefore, similar to (3-6), we have

$$\begin{aligned}[][w]_{\alpha,\alpha;Q_{2^{-k}}} \\ \leq C2^{-k(1-\alpha)}\Bigg[&\omega_f(2^{-k}) + \omega_b(2^{-k})\|Du\|_{L_\infty(Q_{2^{-k}})} \\ &+ \omega_a(2^{-k})\Bigg(\sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{(t_0,x_0)\in Q_{2^{-k}}} \inf_{p\in\mathcal{P}_x} [u-p]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty}\Bigg)\Bigg]. \end{aligned} \tag{3-24}$$

From (2-16) and the triangle inequality,

$$[v]_{\alpha,\alpha;Q_{2^{-k}}} \leq [w]_{\alpha,\alpha;Q_{2^{-k}}} + [u-p_0]_{\alpha,\alpha;Q_{2^{-k}}} \leq [w]_{\alpha,\alpha;Q_{2^{-k}}} + C\inf_{p\in\mathcal{P}_1}[u-p]_{\alpha,\alpha;Q_{2^{-k}}}.$$

For all $l = 1, 2, \ldots$, we define $Q^l = Q_{1-2^{-l}}$. Similar to (3-9), by combining (3-23) and (3-24), shifting the coordinates, and using the above inequality, we obtain for all $l \geq 1$ and $k \geq l+1$,

$$\begin{aligned} 2^{(k+k_0)(1-\alpha)} &\sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_1} [u-p]_{\alpha,\alpha;Q_{2^{-k-k_0}}(t_0,x_0)} \\ &\leq C2^{-(k+k_0)\alpha} \sup_{(t_0,x_0)\in Q^l} \sum_{j=0}^{k} 2^{k-j} \inf_{p\in\mathcal{P}_1} [u-p]_{\alpha,\alpha;(t_0-2^{-k},t_0)\times B_{2^{j-k}}(x_0)} \\ &\quad + C2^{k_0(1-\alpha)}\Bigg[\omega_f(2^{-k}) + \omega_b(2^{-k})\|Du\|_{L_\infty(Q^{l+1})} \\ &\qquad + \omega_a(2^{-k})\Bigg(\sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \inf_{p\in\mathcal{P}_x} [u-p]^x_{\alpha,Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty}\Bigg)\Bigg] \\ &\quad + C2^{-(k+k_0)\alpha}[u]_{\alpha,\alpha}, \end{aligned} \tag{3-25}$$

which by summing in $k = l+1, l+2, \ldots$ implies

$$\begin{aligned} \sum_{k=l+1}^{\infty} 2^{(k+k_0)(1-\alpha)} &\sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_1} [u-p]_{\alpha,\alpha;Q_{2^{-k-k_0}}(t_0,x_0)} \\ &\leq C2^{-k_0\alpha} \sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \inf_{p\in\mathcal{P}_1} [u-p]_{\alpha,\alpha;Q_{2^{-j}}(t_0,x_0)} \\ &\quad + C2^{-(k_0+l)\alpha}[u]_{\alpha,\alpha} + C2^{k_0(1-\alpha)} \sum_{k=l+1}^{\infty} \omega_f(2^{-k}) \\ &\quad + C2^{k_0(1-\alpha)} \sum_{k=l+1}^{\infty} \Bigg[\omega_b(2^{-k})\|Du\|_{L_\infty(Q^{l+1})} + \omega_a(2^{-k}) \\ &\qquad \cdot \Bigg(\sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \inf_{p\in\mathcal{P}_x} [u-p]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty}\Bigg)\Bigg], \end{aligned}$$

where for the first term on the right-hand side, we replaced j by $k-j$, switched the order of the summation, and bounded it by

$$\sum_{k=0}^{\infty} 2^{-(k+k_0)\alpha} \sum_{j=0}^{k} 2^j \sup_{(t_0,x_0)\in Q^{l+1}} \inf_{p\in\mathcal{P}_1} [u-p]_{\alpha,\alpha;Q_{2^{-j}}(t_0,x_0)}$$
$$= 2^{-k_0\alpha} \sum_{j=0}^{\infty} 2^j \sup_{(t_0,x_0)\in Q^{l+1}} \inf_{p\in\mathcal{P}_1} [u-p]_{\alpha,\alpha;Q_{2^{-j}}(t_0,x_0)} \sum_{k=j}^{\infty} 2^{-k\alpha}$$
$$\le C 2^{-k_0\alpha} \sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \inf_{p\in\mathcal{P}_1} [u-p]_{\alpha,\alpha;Q_{2^{-j}}(t_0,x_0)}.$$

Therefore,

$$\sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{(t_0,x_0)\in Q^{l}} \inf_{p\in\mathcal{P}_1} [u-p]_{\alpha,\alpha;Q_{2^{-j}}(t_0,x_0)}$$
$$\le C 2^{-k_0\alpha} \sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \inf_{p\in\mathcal{P}_1} [u-p]_{\alpha,\alpha;Q_{2^{-j}}(t_0,x_0)}$$
$$+ C 2^{(l+k_0)(1-\alpha)} [u]_{\alpha,\alpha} + C 2^{k_0(1-\alpha)} \sum_{k=l+1}^{\infty} \omega_f(2^{-k})$$
$$+ C 2^{k_0(1-\alpha)} \sum_{k=l+1}^{\infty} \omega_b(2^{-k}) \|Du\|_{L_\infty(Q^{l+1})}$$
$$+ C 2^{k_0(1-\alpha)} \sum_{k=l+1}^{\infty} \omega_a(2^{-k}) \Bigl(\sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \inf_{p\in\mathcal{P}_1} [u-p]_{\alpha,\alpha;Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty} \Bigr). \tag{3-26}$$

Then we choose k_0 and l sufficiently large, and apply Lemma 2.1(iii) to obtain

$$\sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{(t_0,x_0)\in Q^{l}} \inf_{p\in\mathcal{P}_1} [u-p]_{\alpha,\alpha;Q_{2^{-j}}(t_0,x_0)}$$
$$\le \frac{1}{4} \sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \inf_{p\in\mathcal{P}_1} [u-p]_{\alpha,\alpha;Q_{2^{-j}}(t_0,x_0)} + C 2^{(l+k_0)(1-\alpha)} \|u\|_{\frac{\alpha}{\sigma},\alpha} + C \sum_{k=1}^{\infty} \omega_f(2^{-k}),$$

and thus,

$$\frac{1}{4^l} \sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{(t_0,x_0)\in Q^{l}} \inf_{p\in\mathcal{P}_1} [u-p]_{\alpha,\alpha;Q_{2^{-j}}(t_0,x_0)} \le C \|u\|_{\frac{\alpha}{\sigma},\alpha} + C \sum_{k=1}^{\infty} \omega_f(2^{-k}), \tag{3-27}$$

from which (3-20) follows. The proposition is proved. □

Proof of Theorem 1.1. We use the localization argument to prove Theorem 1.1.

Without loss of generality, we assume the equation holds in Q_3. We divide the proof into three steps.

Step 1: For $k = 1, 2, \ldots$, define $Q^k := Q_{1-2^{-k}}$. Let $\eta_k \in C_0^\infty(\hat{Q}^{k+3})$ be a sequence of nonnegative smooth cutoff functions satisfying $\eta_k \equiv 1$ in Q^{k+2}, $|\eta_k| \le 1$ in Q^{k+3}, $\|\partial_t^j D^i \eta_k\|_{L_\infty} \le C 2^{k(i+j)}$ for each $i, j \ge 0$. Set $v_k := u\eta_k \in C^{1,\sigma+}$ and notice that in Q^{k+1},

$$\begin{aligned}\partial_t v_k = \eta_k \partial_t u + \partial_t \eta_k u &= \inf_{\beta\in\mathcal{A}} (\eta_k L_\beta u + \eta_k b_\beta Du + \eta_k f_\beta + \partial_t \eta_k u)\\ &= \inf_{\beta\in\mathcal{A}} (L_\beta v_k + b_\beta D v_k - b_\beta u D\eta_k + h_{k\beta} + \eta_k f_\beta + \partial_t \eta_k u),\end{aligned}$$

where

$$h_{k\beta} = \eta_k L_\beta u - L_\beta v_k = \int_{\mathbb{R}^d} \frac{\xi_k(t,x,y) a_\beta(t,x,y)}{|y|^{d+\sigma}}\, dy,$$

and

$$\begin{aligned}\xi_k(t,x,y) &= u(t,x+y)(\eta_k(t,x+y) - \eta_k(t,x)) - y \cdot D\eta_k(t,x) u(t,x)(\chi_{\sigma=1}\chi_{B_1} + \chi_{\sigma>1})\\ &= u(t,x+y)(\eta_k(t,x+y) - \eta_k(t,x)) \quad \text{since } D\eta_k \equiv 0 \text{ in } Q^{k+1}.\end{aligned}$$

We will apply Proposition 3.3 to the equation of v_k in Q^{k+1} and obtain corresponding estimates for v_k in Q^k.

Obviously, in Q^{k+1} we have $\eta_k f_\beta \equiv f_\beta$, $b_\beta u D\eta_k \equiv 0$, and $\partial_t \eta_k u \equiv 0$. Thus, we only need to estimate the modulus of continuity of $h_{k\beta}$ in Q^{k+1}.

Step 2: For $(t,x) \in Q^{k+1}$ and $|y| \le 2^{-k-3}$, we have

$$\xi_k(t,x,y) = 0.$$

Also,

$$\begin{aligned}|\xi_k(t,x,y)| &= |u(t,x+y)(\eta_k(t,x+y) - \eta_k(t,x))|\\ &\le \begin{cases} 2\omega_u(|y|) + 2|u(t,x)| & \text{when } |y| \ge 1,\\ C2^k |u(t,x+y)|\,|y| & \text{when } 2^{-k-3} < |y| < 1.\end{cases}\end{aligned}$$

For $(t,x), (t',x') \in Q^{k+1}$, by the triangle inequality,

$$\begin{aligned}&|h_{k\beta}(t,x) - h_{k\beta}(t',x')|\\ &\quad\le \int_{\mathbb{R}^d} \frac{|(\xi_k(t,x,y) - \xi_k(t',x',y)) a_\beta(t,x,y)|}{|y|^{d+\sigma}} + \frac{|\xi_k(t',x',y)(a_\beta(t,x,y) - a_\beta(t',x',y))|}{|y|^{d+\sigma}}\, dy\\ &\quad := \mathrm{I} + \mathrm{II}.\end{aligned} \tag{3-28}$$

By the estimates of $|\xi_k(t,x,y)|$ above, we have

$$\mathrm{II} \le C\Big(2^{k(\sigma+1)} \|u\|_{L_\infty(Q_2)} + \sum_{j=0}^\infty 2^{-j\sigma} \omega_u(2^j)\Big) \omega_a(\max\{|x-x'|, |t-t'|^{\frac{1}{\sigma}}\}), \tag{3-29}$$

where C depends on d, σ, and Λ. For I, by the fundamental theorem of calculus,

$$\xi_k(t,x,y) - \xi_k(t',x',y) = y \cdot \int_0^1 \big(u(t,x+y) D\eta_k(t,x+sy) - u(t',x'+y) D\eta_k(t',x'+sy)\big)\, ds.$$

When $2^{-k-3} \le |y| < 2$, similar to the estimate of $\xi_k(t,x,y)$, it follows that

$$|\xi_k(t,x,y) - \xi_k(t',x',y)| \le C|y|\big(2^k \omega_u(\max\{|x-x'|, |t-t'|^{\frac{1}{\sigma}}\}) + 2^{2k}\|u\|_{L_\infty(Q_3)}(|x-x'|+|t-t'|)\big). \tag{3-30}$$

When $|y| \ge 2$, we have

$$|\xi_k(t,x,y) - \xi_k(t',x',y)| = |u(t,x+y) - u(t',x'+y)| \le \omega_u(\max\{|x-x'|, |t-t'|^{\frac{1}{\sigma}}\}),$$

which implies

$$\mathrm{I} \le C2^{k(\sigma+1)}\omega_u(\max\{|x-x'|, |t-t'|^{\frac{1}{\sigma}}\}) + C2^{k(\sigma+2)}\|u\|_{L_\infty(Q_3)}(|x-x'|+|t-t'|).$$

Therefore,

$$|h_{k\beta}(t,x) - h_{k\beta}(t',x')| \le \omega_h(\max\{|x-x'|, |t-t'|^{\frac{1}{\sigma}}\}),$$

where

$$\omega_h(r) := C\Bigg(2^{k(\sigma+1)}\|u\|_{L_\infty(Q_3)} + \sum_{j=0}^{\infty} 2^{-j\sigma}\omega_u(2^j)\Bigg)\omega_a(r) + C2^{k(\sigma+1)}\omega_u(r) + C2^{k(\sigma+2)}\|u\|_{L_\infty(Q_3)}(r+r^\sigma) \tag{3-31}$$

is a Dini function.

Step 3: In this last step, we only present the detailed proof for $\sigma \in (1,2)$. We omit the details for the proof of the case $\sigma \in (0,1]$, since it is almost the same as and actually even simpler than the case $\sigma \in (1,2)$. We apply Proposition 3.3, together with a scaling and covering argument, to v_k to obtain

$$\begin{aligned}
&\|\partial_t v_k\|_{L_\infty(Q^k)} + [v_k]^x_{\sigma;Q^k} + [Dv_k]^t_{\frac{\sigma-1}{\sigma};Q^k} \\
&\quad\le C2^{k\sigma}\|v_k\|_{L_\infty} + C2^{k(\sigma-\alpha)}[v_k]_{\frac{\alpha}{\sigma},\alpha} + C\sum_{j=1}^{\infty}(\omega_h(2^{-j}) + \omega_f(2^{-j})) \\
&\quad\le C2^{k(\sigma+2)}\|u\|_{L_\infty(Q_3)} + C_0 2^{k(\sigma-\alpha)}[u]_{\frac{\alpha}{\sigma},\alpha;Q^{k+3}} \\
&\qquad\qquad + C\sum_{j=0}^{\infty} 2^{-j\sigma}\omega_u(2^j) + C\sum_{j=0}^{\infty}(2^{k(\sigma+1)}\omega_u(2^{-j}) + \omega_f(2^{-j})),
\end{aligned}$$

where C and C_0 depend on d, λ, Λ, σ, N_0, ω_b, and ω_a, but are independent of k. Since $\eta_k \equiv 1$ in Q^k, it follows that

$$\begin{aligned}
&\|\partial_t u\|_{L_\infty(Q^k)} + [u]^x_{\sigma;Q^k} + [Du]^t_{\frac{\sigma-1}{\sigma};Q^k} \\
&\quad\le C2^{k(\sigma+2)}\|u\|_{L_\infty(Q_3)} + C_0 2^{k(\sigma-\alpha)}[u]_{\frac{\alpha}{\sigma},\alpha;Q^{k+3}} \\
&\qquad\qquad + C\sum_{j=0}^{\infty} 2^{-j\sigma}\omega_u(2^j) + C\sum_{j=0}^{\infty}(2^{k(\sigma+1)}\omega_u(2^{-j}) + \omega_f(2^{-j})).
\end{aligned} \tag{3-32}$$

By the interpolation inequality, for any $\varepsilon \in (0,1)$,

$$[u]_{\frac{\alpha}{\sigma},\alpha;Q^{k+3}} \le \varepsilon(\|\partial_t u\|_{L_\infty(Q^{k+3})} + [u]^x_{\sigma;Q^{k+3}}) + C\varepsilon^{-\frac{\alpha}{\sigma-\alpha}}\|u\|_{L_\infty(Q_3)}. \tag{3-33}$$

Recall that $\alpha \le (\sigma-1)/2$ and thus,

$$\frac{\alpha}{\sigma-\alpha} \le \frac{\sigma-1}{\sigma+1} < \frac{1}{2}.$$

Combining (3-32) and (3-33) with $\varepsilon = C_0^{-1}2^{-3k-16}$, we obtain

$$\begin{aligned}\|\partial_t u\|_{L_\infty(Q^k)} &+ [u]^x_{\sigma;Q^k} + [Du]^t_{\frac{\sigma-1}{\sigma};Q^k}\\ &\le 2^{-16}([u]^x_{\sigma;Q^{k+3}} + \|\partial_t u\|_{L_\infty(Q^{k+3})} + [Du]^t_{\frac{\sigma-1}{\sigma};Q^{k+3}}) + C2^{4k}\|u\|_{L_\infty(Q_3)}\\ &\qquad + C\sum_{j=0}^{\infty} 2^{-j\sigma}\omega_u(2^j) + C\sum_{j=0}^{\infty}(2^{k(\sigma+1)}\omega_u(2^{-j}) + \omega_f(2^{-j})).\end{aligned}$$

Then we multiply 2^{-5k} to both sides of the above inequality and get

$$\begin{aligned}2^{-5k}&(\|\partial_t u\|_{L_\infty(Q^k)} + [u]^x_{\sigma;Q^k} + [Du]^t_{\frac{\sigma-1}{\sigma};Q^k})\\ &\le 2^{-5(k+3)-1}(\|\partial_t u\|_{L_\infty(Q^{k+3})} + [u]^x_{\sigma;Q^{k+3}} + [Du]^t_{\frac{\sigma-1}{\sigma};Q^{k+3}})\\ &\qquad + C2^{-k}\|u\|_{L_\infty(Q_3)} + C2^{-2k}\sum_{j=0}^{\infty}(2^{-j\sigma}\omega_u(2^j) + \omega_u(2^{-j}) + \omega_f(2^{-j})).\end{aligned}$$

We sum up the both sides of the above inequality and obtain

$$\begin{aligned}\sum_{k=1}^{\infty} &2^{-5k}(\|\partial_t u\|_{L_\infty(Q^k)} + [u]^x_{\sigma;Q^k} + [Du]^t_{\frac{\sigma-1}{\sigma};Q^k})\\ &\le \frac{1}{2}\sum_{k=4}^{\infty} 2^{-5k}(\|\partial_t u\|_{L_\infty(Q^k)} + [u]^x_{\sigma;Q^k} + [Du]^t_{\frac{\sigma-1}{\sigma};Q^k})\\ &\qquad + C\|u\|_{L_\infty(Q_3)} + C\sum_{j=0}^{\infty}(2^{-j\sigma}\omega_u(2^j) + \omega_u(2^{-j}) + \omega_f(2^{-j})),\end{aligned}$$

which further implies

$$\begin{aligned}\sum_{k=1}^{\infty} &2^{-5k}(\|\partial_t u\|_{L_\infty(Q^k)} + [u]^x_{\sigma;Q^k} + [Du]^t_{\frac{\sigma-1}{\sigma};Q^k})\\ &\le C\|u\|_{L_\infty(Q_3)} + C\sum_{j=0}^{\infty}(2^{-j\sigma}\omega_u(2^j) + \omega_u(2^{-j}) + \omega_f(2^{-j})),\end{aligned}$$

where C depends on d, λ, Λ, σ, ω_b, N_0, and ω_a. By applying this estimate to $u - u(0,0)$, we obtain

$$\|\partial_t u\|_{L_\infty(Q^4)} + [u]^x_{\sigma;Q^4} + [Du]^t_{\frac{\sigma-1}{\sigma};Q^4} \le C\sum_{j=0}^{\infty}(2^{-j\sigma}\omega_u(2^j) + \omega_u(2^{-j}) + \omega_f(2^{-j})). \tag{3-34}$$

This proves (1-4).

Finally, since $\|v_1\|_{\frac{\alpha}{\sigma},\alpha}$ is bounded by the right-hand side of (3-34), from (3-19), we see that

$$\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_1} [v_1 - p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j}}(t_0,x_0)} \le C$$

for some large l. This and (3-18) with u replaced by v_1 and f_β replaced by $h_{1\beta}+\eta_1 f_\beta+\partial_t\eta_1 u-b_\beta u D\eta_1$ give

$$\sum_{j=k_1+1}^{\infty} 2^{(j+k_0)(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^{k_1}} \inf_{p\in\mathcal{P}_1} [v_1 - p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j-k_0}}(t_0,x_0)}$$
$$\le C2^{-k_0\alpha} + C2^{k_0(\sigma-\alpha)} \sum_{j=k_1}^{\infty} (\omega_f(2^{-j}) + \omega_a(2^{-j}) + \omega_u(2^{-j}) + \omega_b(2^{-j}) + 2^{-j\alpha}).$$

Here we also used (3-31) with $k = 1$. Therefore, for any small $\varepsilon > 0$, we can find k_0 sufficiently large, then k_1 sufficiently large, depending only on C, σ, N_0, α, ω_f, ω_a, ω_f, ω_b, and ω_u, such that

$$\sum_{j=k_1+1}^{\infty} 2^{(j+k_0)(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^{k_1}} \inf_{p\in\mathcal{P}_1} [v_1 - p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j-k_0}}(t_0,x_0)} < \varepsilon,$$

which, together with the fact that $v_1 = u$ in $Q_{\frac{1}{2}}$ and the proof of Lemma 2.1(ii), indicates that

$$\sup_{(t_0,x_0)\in Q_{1/2}} ([u]^x_{\sigma;Q_r(t_0,x_0)} + [Du]^t_{\frac{\sigma-1}{\sigma};Q_r(t_0,x_0)}) \to 0 \quad \text{as } r \to 0$$

with a decay rate depending only on d, λ, N_0, Λ, ω_a, ω_f, ω_b, ω_u, and σ. Hence, the proof of the case when $\sigma \in (1, 2)$ is completed. □

4. Schauder estimates for equations with drifts

We now are going to prove Theorems 1.3 and 1.4. Here, the main difference from the theorems in [Dong and Zhang 2016b] is that our equation may have a drift, especially for $\sigma = 1$.

We first prove a weaker version of Theorem 1.3.

Proposition 4.1. *Suppose (1-1) is satisfied in Q_2. Then under the conditions of Theorem 1.3, for any $\gamma \in (0, \min\{\hat\alpha, 2-\sigma\})$ with $\hat\alpha$ being the one in Proposition 3.1, and any $\alpha \in (\gamma, \min\{\hat\alpha, 2-\sigma\})$, we have*

$$[u]_{1+\frac{\gamma}{\sigma},\sigma+\gamma;Q_{1/2}} \le C(\|u\|_{\frac{\alpha}{\sigma},\alpha} + C_f),$$

where $C > 0$ is a constant depending only on $d, \gamma, \alpha, \sigma, \lambda, \Lambda, N_0$, and C_b.

Proof. The proof is very similar to that of Propositions 3.2, 3.3, and 3.4. We fix an $\alpha \in (\gamma, \hat\alpha)$.

Case 1: $\sigma \in (0, 1)$. We start from (3-9). Let Q^l and ℓ_0 be as in the proof of Proposition 3.2. Multiplying $2^{(k+k_0)\gamma}$ to both sides of (3-9) and making use of the Hölder continuity of a and f, we have for all $l \ge \ell_0$

and $k \geq l+1$,

$$
\begin{aligned}
2^{(k+k_0)(\sigma+\gamma-\alpha)} &\sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k-k_0}}(t_0,x_0)} \\
&\leq C 2^{(k+k_0)(\gamma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \sum_{j=0}^{k} 2^{(k-j)\sigma} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;(t_0-2^{-k\sigma},t_0)\times B_{2^{j-k}}(x_0)} \\
&\quad + C 2^{(k+k_0)(\gamma-\alpha)} [u]_{\frac{\alpha}{\sigma},\alpha} \\
&\quad + C 2^{k_0(\sigma+\gamma-\alpha)} \Bigl[C_f + \sup_{(t_0,x_0)\in Q^{l+1}} \sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty} \Bigr].
\end{aligned}
$$

Taking the supremum in $k \geq \ell_0 + 1$ and using the fact that $\gamma < \alpha$, we have

$$
\begin{aligned}
\sup_{k\geq \ell_0+k_0+1} 2^{k(\sigma+\gamma-\alpha)} &\sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}(t_0,x_0)} \\
&\leq C 2^{k_0(\gamma-\alpha)} \sup_{k\geq 0} 2^{k(\sigma+\gamma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}(x_0)} \\
&\quad + C 2^{(\ell_0+k_0+1)(\gamma-\alpha)} [u]_{\frac{\alpha}{\sigma},\alpha} \\
&\quad + C 2^{k_0(\sigma+\gamma-\alpha)} \Bigl[C_f + \sup_{(t_0,x_0)\in Q^{l+1}} \sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty} \Bigr].
\end{aligned}
$$

By taking k_0 large, $l = \ell_0$, using (3-10), and noticing that

$$
\sup_{0\leq k\leq \ell_0+k_0} 2^{k(\sigma+\gamma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}(t_0,x_0)}} \leq C 2^{(\ell_0+k_0)(\sigma+\gamma-\alpha)} [u]_{\frac{\alpha}{\sigma},\alpha},
$$

we have

$$
\sup_{k\geq 0} 2^{k(\sigma+\gamma-\alpha)} \sup_{(t_0,x_0)\in Q_{1/2}} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}(t_0,x_0)} \leq C[C_f + \|u\|_{\frac{\alpha}{\sigma},\alpha}].
$$

Since

$$
[u]_{1+\frac{\gamma}{\sigma},\sigma+\gamma;Q_{1/2}} \leq C \sup_{k\geq 0} 2^{k(\sigma+\gamma-\alpha)} \sup_{(t_0,x_0)\in Q_{1/2}} \inf_{p\in\mathcal{P}_t} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}(t_0,x_0)} + C[u]_{\frac{\alpha}{\sigma},\alpha},
$$

we obtain

$$
[u]_{1+\frac{\gamma}{\sigma},\sigma+\gamma;Q_{1/2}} \leq C(\|u\|_{\frac{\alpha}{\sigma},\alpha} + C_f).
$$

Case 2: $\sigma \in (1,2)$. We start from (3-17). Let Q^l be as in the proof of Proposition 3.3. Multiplying $2^{(k+k_0)\gamma}$ to both sides of (3-17) and making use of the Hölder continuity of a, b, and f, we have for all $l \geq 1$ and $k \geq l+1$,

$$
\begin{aligned}
&2^{(k+k_0)(\sigma+\gamma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_1} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k-k_0}}(t_0,x_0)} \\
&\le C2^{(k+k_0)(\gamma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \sum_{j=0}^{k} 2^{(k-j)\sigma} \inf_{p\in\mathcal{P}_1} [u-p]_{\frac{\alpha}{\sigma},\alpha;(t_0-2^{-k\sigma},t_0)\times B_{2^{j-k}}(x_0)} \\
&\quad + C2^{(k+k_0)(\gamma-\alpha)}[u]_{\frac{\alpha}{\sigma},\alpha} + C2^{(k+k_0)(\gamma-\alpha)+k_0\sigma}[Du]_{\frac{\alpha}{\sigma},\alpha;Q^{l+1}} \\
&\quad + C2^{k_0(\sigma+\gamma-\alpha)}\Big[C_f + \|Du\|_{L_\infty(Q^{l+1})} + \sup_{(t_0,x_0)\in Q^{l+1}} \sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \inf_{p\in\mathcal{P}_1}[u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty}\Big].
\end{aligned}
$$

Note that this $\hat{\alpha}$ can be chosen very small, at least strictly smaller than $\sigma-1$. Taking the supremum in $k\ge 2$ and using the fact that $\gamma<\alpha$, we have

$$
\begin{aligned}
&\sup_{k\ge k_0+2} 2^{k(\sigma+\gamma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_1} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}(t_0,x_0)} \\
&\le C2^{k_0(\gamma-\alpha)} \sup_{k\ge 0} 2^{k(\sigma+\gamma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_1} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}(x_0)} \\
&\quad + C2^{(k_0+2)(\gamma-\alpha)}[u]_{\frac{\alpha}{\sigma},\alpha} + C2^{(2+k_0)(\gamma-\alpha)+k_0\sigma}[Du]_{\frac{\alpha}{\sigma},\alpha;Q^{l+1}} \\
&\quad + C2^{k_0(\sigma+\gamma-\alpha)}\Big[C_f + \|Du\|_{L_\infty(Q^{l+1})} + \sup_{(t_0,x_0)\in Q^{l+1}} \sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \inf_{p\in\mathcal{P}_1}[u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty}\Big].
\end{aligned}
$$

By taking k_0 large, $l=1$, using (3-19) and (2-4), and noticing that

$$
\sup_{0\le k\le 1+k_0} 2^{k(\sigma+\gamma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_1} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}(t_0,x_0)}} \le C2^{(1+k_0)(\sigma+\gamma-\alpha)}[u]_{\frac{\alpha}{\sigma},\alpha},
$$

we have

$$
\sup_{k\ge 0} 2^{k(\sigma+\gamma-\alpha)} \sup_{(t_0,x_0)\in Q_{1/2}} \inf_{p\in\mathcal{P}_1} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}(t_0,x_0)} \le C[C_f + \|u\|_{\frac{\alpha}{\sigma},\alpha}].
$$

Since

$$
[u]_{1+\frac{\gamma}{\sigma},\sigma+\gamma;Q_{1/2}} \le C \sup_{k\ge 0} 2^{k(\sigma+\gamma-\alpha)} \sup_{(t_0,x_0)\in Q_{1/2}} \inf_{p\in\mathcal{P}_1} [u-p]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}(t_0,x_0)} + C[u]_{\frac{\alpha}{\sigma},\alpha},
$$

we obtain

$$
[u]_{1+\frac{\gamma}{\sigma},\sigma+\gamma;Q_{1/2}} \le C(\|u\|_{\frac{\alpha}{\sigma},\alpha} + C_f).
$$

<u>Case 3</u>: $\sigma=1$. We start from (3-25). Multiplying $2^{(k+k_0)\gamma}$ to both sides of (3-25) and making use of the Hölder continuity of a, b, f, we have for all $l\ge 1$ and $k\ge l+1$,

$$
\begin{aligned}
&2^{(k+k_0)(1+\gamma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_1} [u-p]_{\alpha,\alpha;Q_{2^{-k-k_0}}(t_0,x_0)} \\
&\quad\le C2^{(k+k_0)(\gamma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \sum_{j=0}^{k} 2^{k-j} \inf_{p\in\mathcal{P}_1} [u-p]_{\alpha,\alpha;(t_0-2^{-k},t_0)\times B_{2^{j-k}}(x_0)} \\
&\quad\quad + C2^{k_0(1+\gamma-\alpha)}\Bigg[C_f + \|Du\|_{L_\infty(Q^{l+1})} + \sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \inf_{p\in\mathcal{P}_x} [u-p]^x_{\alpha,Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty}\Bigg] \\
&\quad\quad + C2^{(k+k_0)(\gamma-\alpha)}[u]_{\alpha,\alpha}.
\end{aligned}
$$

Taking the supremum in $k \ge 2$ and using the fact that $\gamma < \alpha$, we have

$$
\begin{aligned}
&\sup_{k\ge k_0+2} 2^{k(1+\gamma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_1} [u-p]_{\alpha,\alpha;Q_{2^{-k}}(t_0,x_0)} \\
&\quad\le C2^{(k_0-1)(\gamma-\alpha)} \sup_{k\ge 0} 2^{k(1+\gamma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_1} [u-p]_{\alpha,\alpha;Q_{2^{-k}}(t_0,x_0)} \\
&\quad\quad + C2^{k_0(1+\gamma-\alpha)}\Bigg[C_f + \|Du\|_{L_\infty(Q^{l+1})} + \sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \inf_{p\in\mathcal{P}_x} [u-p]^x_{\alpha,Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty}\Bigg] \\
&\quad\quad + C2^{k_0(\gamma-\alpha)}[u]_{\alpha,\alpha}.
\end{aligned}
$$

By taking k_0 large, $l = 1$, using (3-27), and noticing that

$$
\sup_{0\le k\le k_0+1} 2^{k(1+\gamma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_1} [u-p]_{\alpha,\alpha;Q_{2^{-k}}(t_0,x_0)} \le C2^{k_0(1+\gamma-\alpha)}[u]_{\alpha,\alpha},
$$

we have

$$
\sup_{k\ge 0} 2^{k(1+\gamma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \inf_{p\in\mathcal{P}_1} [u-p]_{\alpha,\alpha;Q_{2^{-k}}(t_0,x_0)} \le C[C_f + \|u\|_{\alpha,\alpha}].
$$

Since

$$
[Du]_{\gamma,\gamma;Q_{1/2}} + [\partial_t u]_{\gamma,\gamma;Q_{1/2}} \le C \sup_{k\ge 0} 2^{k(1+\gamma-\alpha)} \sup_{(t_0,x_0)\in Q_{1/2}} \inf_{p\in\mathcal{P}_1} [u-p]_{\alpha,\alpha;Q_{2^{-k}}(t_0,x_0)} + C[u]_{\alpha,\alpha},
$$

we obtain

$$
[Du]_{\gamma,\gamma;Q_{1/2}} + [\partial_t u]_{\gamma,\gamma;Q_{1/2}} \le C[C_f + \|u\|_{\alpha,\alpha}]. \qquad \square
$$

Proof of Theorem 1.3. The proof is the same as that of Theorem 1.1 using localizations. We sketch the proof here. We use the same notation as in the proof of Theorem 1.1. Without loss of generality, we assume (1-1) holds in Q_3.

Let $\eta_k \in C_0^\infty(\hat{Q}^{k+3})$ be a sequence of nonnegative smooth cutoff functions satisfying $\eta \equiv 1$ in Q^{k+2}, $|\eta| \le 1$ in Q^{k+3}, $\|\partial_t^j D^i \eta_k\|_{L_\infty} \le C2^{k(i+j)}$ for each $i, j \ge 0$. Set $v_k := u\eta_k \in C^{1+\frac{\gamma}{\sigma},\sigma+\gamma}$ and notice that in Q_1,

$$
\partial_t v_k = \inf_{\beta\in\mathcal{A}} (L_\beta v_k + b_\beta D v_k - b_\beta u D\eta_k + h_{k\beta} + \eta_k f_\beta + \partial_t \eta_k u),
$$

where

$$h_{k\beta}(t,x)=\int_{\mathbb{R}^d}\frac{\xi_k(t,x,y)a_\beta(t,x,y)}{|y|^{d+1}}\,dy,$$

and

$$\begin{aligned}\xi_k(t,x,y)&:=u(t,x+y)(\eta_k(t,x+y)-\eta_k(t,x))-u(t,x)y\cdot D\eta_k(t,x)(\chi_{\sigma=1}\chi_{B_1}+\chi_{\sigma>1})\\&=u(t,x+y)(\eta_k(t,x+y)-\eta_k(t,x))\quad\text{since } D\eta_k\equiv 0 \text{ in } Q^{k+1}.\end{aligned}$$

We will apply Proposition 4.1 to the equation of v_k in Q^{k+1} and obtain corresponding estimates for v_k in Q^k.

Obviously, in Q^{k+1} we have $\eta_k f_\beta\equiv f_\beta$, $buD\eta_k\equiv 0$, and $\partial_t\eta_k u\equiv 0$. Thus, we only need to estimate the modulus of continuity of $h_{k\beta}$ in Q^{k+1}. Since

$$\xi_k(t,x,y):=u(t,x+y)(\eta_k(t,x+y)-\eta_k(t,x)),$$

which is the same as in Theorem 1.1, we also have (3-31) here. Therefore,

$$[h_{k\beta}]_{\frac{\gamma}{\sigma},\gamma;Q^{k+1}}\le C\biggl(2^{k(\sigma+1)k}\|u\|_{L_\infty(Q_3)}+\sum_{j=0}^{\infty}2^{-j\sigma}\omega_u(2^j)\biggr)+C2^{k(\sigma+1)}[u]_{\frac{\gamma}{\sigma},\gamma}+C2^{k(\sigma+2)}\|u\|_{L_\infty(Q_3)}.$$

The rest is almost the same as (actually much simpler than) the proof of Theorem 1.1, by using Proposition 4.1 (recalling $\gamma<\alpha$), and we omit the details. □

In the following, we prove Theorem 1.4 using Theorem 1.3 and difference quotients.

Proof of Theorem 1.4. We only provide the proof for $\sigma+\gamma>2$. We know from Theorem 1.3 that there exists γ_0 such that $\sigma+\gamma_0<2$ is not an integer, and the theorem holds for $0<\gamma\le\gamma_0$. Below we will prove the theorem for all $\gamma\in(\gamma_0,\sigma)$ using difference quotients.

We suppose (1-7) holds in Q_4. We will consider the difference quotients in x first. For $h\in\bigl(0,\frac14\bigr)$, $e\in\mathbb{S}^{d-1}$, let

$$u^h(t,x)=\frac{u(t,x+he)-u(t,x)}{h^{\gamma-\gamma_0}},\qquad f^h(t,x)=\frac{f(t,x+he)-f(t,x)}{h^{\gamma-\gamma_0}},$$

and

$$a^h(t,x,y)=\frac{a(t,x+he,y)-a(t,x,y)}{h^{\gamma-\gamma_0}},\qquad b^h(t,x)=\frac{b(t,x+he)-b(t,x)}{h^{\gamma-\gamma_0}}.$$

Then u^h satisfies

$$\partial_t u^h(t,x)=L_h u^h+b(t,x+he)Du^h+f^h+b^hDu+g\quad\text{in } Q_1,$$

where

$$L_hu=\int_{\mathbb{R}^d}\frac{\delta u(t,x,y)a(t,x+he,y)}{|y|^{d+\sigma}}\,dy,\qquad g=\int_{\mathbb{R}^d}\frac{\delta u(t,x,y)a^h(t,x,y)}{|y|^{d+\sigma}}\,dy.$$

Applying the result for $\gamma=\gamma_0$ gives

$$[u^h]_{1+\frac{\gamma_0}{\sigma},\sigma+\gamma_0;Q_{3/4}}\le C\|u^h\|_{\frac{\gamma_0}{\sigma},\gamma_0}+C[f^h+b^hDu+g]_{\frac{\gamma_0}{\sigma},\gamma_0;Q_1}.$$

It follows from direct calculations that

$$[g]_{\frac{\gamma_0}{\sigma},\gamma_0;Q_1}\le C[u]_{1+\frac{\gamma_0}{\sigma},\sigma+\gamma_0;Q_{5/4}}+C\|u\|_{\frac{\gamma_0}{\sigma},\gamma_0}.$$

Applying the $C^{1+\frac{\gamma_0}{\sigma},\sigma+\gamma_0}$ estimate as mentioned at the beginning of this proof, we have

$$[g]_{\frac{\gamma_0}{\sigma},\gamma_0;Q_1} \le C\|u\|_{\frac{\gamma_0}{\sigma},\gamma_0} + C[f]_{\frac{\gamma_0}{\sigma},\gamma_0;Q_2}.$$

Similarly, we have

$$[b^h Du]_{\frac{\gamma_0}{\sigma},\gamma_0;Q_1} \le C\|Du\|_{\frac{\gamma_0}{\sigma},\gamma_0;Q_1} \le C\|u\|_{\frac{\gamma_0}{\sigma},\gamma_0} + C[f]_{\frac{\gamma_0}{\sigma},\gamma_0;Q_2}.$$

Therefore,

$$[u^h]_{1+\frac{\gamma_0}{\sigma},\sigma+\gamma_0;Q_{3/4}} \le C\|u\|_{\frac{\gamma}{\sigma},\gamma} + C[f]_{\frac{\gamma}{\sigma},\gamma;Q_2}.$$

Note that we assumed that $\sigma+\gamma>2$ and thus, $\sigma>1$. Also $1<\sigma+\gamma_0<2$. Then we have

$$[(Du)^h]^x_{\sigma+\gamma_0-1;Q_{3/4}} \le C\|u\|_{\frac{\gamma}{\sigma},\gamma} + C[f]_{\frac{\gamma}{\sigma},\gamma;Q_2},$$

that is,

$$\frac{|Du(t,x+2he)-2Du(t,x+he)+Du(t,x)|}{h^{\sigma+\gamma-1}} \le C\|u\|_{\frac{\gamma}{\sigma},\gamma} + C[f]_{\frac{\gamma}{\sigma},\gamma;Q_2}$$

for all $(t,x)\in Q_{\frac{1}{2}}$ and $h\le\frac{1}{20}$. Making use of (2-1) and sending $j\to\infty$ there, we have

$$\begin{aligned}|Du(t,x+he)-Du(t,x)-hD^2u(t,x)\cdot e| &\le Ch^{\sigma+\gamma-1}\sum_{k=1}^{\infty}2^{-k(\sigma+\gamma-2)}(\|u\|_{\frac{\gamma}{\sigma},\gamma}+[f]_{\frac{\gamma}{\sigma},\gamma;Q_2})\\ &\le C(\|u\|_{\frac{\gamma}{\sigma},\gamma}+[f]_{\frac{\gamma}{\sigma},\gamma;Q_2})h^{\sigma+\gamma-1},\end{aligned}$$

from which we have

$$[u]^x_{\sigma+\gamma;Q_{1/2}} \le C\|u\|_{\frac{\gamma}{\sigma},\gamma} + C[f]_{\frac{\gamma}{\sigma},\gamma;Q_2}. \tag{4-1}$$

Similarly, we can use the difference quotients in t. For $s\in\left(0,\frac{1}{10}\right)$, let

$$u^s(t,x)=\frac{u(t,x)-u(t-s,x)}{s^{\frac{\gamma-\gamma_0}{\sigma}}}. \tag{4-2}$$

By similar arguments, we have

$$[u^s]^t_{1+\frac{\gamma_0}{\sigma};Q_{1/2}} \le C\|u\|_{\frac{\gamma}{\sigma},\gamma} + C[f]_{\frac{\gamma}{\sigma},\gamma;Q_2},$$

that is

$$[(u_t)^s]^t_{\frac{\gamma_0}{\sigma};Q_{1/2}} \le C\|u\|_{\frac{\gamma}{\sigma},\gamma} + C[f]_{\frac{\gamma}{\sigma},\gamma;Q_2}.$$

The same arguments as the above (noticing $\sigma>\gamma$) will lead to

$$[u]^t_{1+\frac{\gamma}{\sigma};Q_{1/2}} \le C\|u\|_{\frac{\gamma}{\sigma},\gamma} + C[f]_{\frac{\gamma}{\sigma},\gamma;Q_2}.$$

This estimate, together with (4-1), implies

$$[u]_{1+\frac{\gamma}{\sigma},\sigma+\gamma;Q_{1/2}} \le C\|u\|_{\frac{\gamma}{\sigma},\gamma} + C[f]_{\frac{\gamma}{\sigma},\gamma;Q_2}.$$

We remark that actually the proof of the other situation $\sigma+\gamma\in(0,1)\cup(1,2)$ is exactly the same as above. □

5. Linear parabolic equations with measurable coefficient in *t*

We now consider the linear equation (1-7), where K, b, and f are Dini continuous in x but only measurable in the time variable t. We first need a proposition for the case that K does not depend on x, and $b \equiv 0$.

Proposition 5.1. *Let $\sigma \in (0,2)$ and $0 < \lambda \le \Lambda$. Assume K does not depend on x, and $b \equiv 0$. Let $\alpha \in (0,1)$ such that $\sigma+\alpha$ is not an integer. Suppose $u \in C_x^{\sigma+\alpha}(Q_1) \cap C^{\frac{\alpha}{\sigma},\alpha}((-2^\sigma,0)\times\mathbb{R}^d)$ is a solution of (1-7) in Q_1. Then,*

$$[u]^x_{\alpha+\sigma;Q_{1/2}} \le C\sum_{j=1}^{\infty} 2^{-j\sigma} M_j + C[f]^x_{\alpha;Q_1},$$

where

$$M_j = \sup_{\substack{(t,x),(t,x')\in(-1,0)\times B_{2^j}\\ 0<|x-x'|<2}} \frac{|u(t,x)-u(t,x')|}{|x-x'|^\alpha}$$

and $C>0$ is a constant depending only on d, σ, λ, Λ, and N_0, and is uniformly bounded as $\sigma \to 2^-$.

Proof. We only prove the case that $\sigma+\alpha>2$ as before. Let η be a cut-off function such that $\eta \in C_c^\infty(\hat{Q}_{\frac{3}{4}})$ and $\eta \equiv 1$ in $Q_{\frac{1}{2}}$. Let $w(t,x) = u(t,x)-u(t,0)$, $\tilde f(t,x) = f(t,x)-f(t,0)$ and $v = \eta w$. Then v satisfies

$$v_t = Lv + h + \eta_t w + \eta \tilde f - \eta g(t) \quad \text{in } (-2^\sigma,0)\times\mathbb{R}^d,$$

where

$$h = \eta L w - L(\eta w) = \int_{\mathbb{R}} \big((\eta(t,x)-\eta(t,x+y))w(t,x+y) + y^T D\eta(t,x) w(t,x)\big) K(t,y)\,dy$$

and

$$g(t) = (Lu)(t,0).$$

By Theorem 4 in [Mikulevičius and Pragarauskas 2014], we have

$$\|v\|^x_{\sigma+\alpha} \le C\|h + \eta_t w + \eta\tilde f - \eta g(t)\|^x_\alpha.$$

From (3.18) and (3.23) in [Dong and Zhang 2016a], we have

$$\|h\|^x_\alpha \le C(\|w\|^x_{\alpha;(-1,0)\times\mathbb{R}^d} + [\nabla w]^x_{\alpha;Q_{15/16}}) \le C(\|u\|^x_{\alpha;(-1,0)\times\mathbb{R}^d} + [\nabla u]^x_{\alpha;Q_{15/16}}).$$

It is clear that

$$\|\eta g(t)\|^x_\alpha \le C(\|D^2u\|_{L_\infty(Q_{7/8})} + \|Du\|_{L_\infty(Q_{7/8})} + \|u\|_{L_\infty((-1,0)\times\mathbb{R}^d)}),$$
$$\|\eta\tilde f\|^x_\alpha \le C[f]^x_{\alpha,Q_{3/4}}.$$

Therefore, we have

$$[u]^x_{\sigma+\alpha;Q_{1/2}} \le C(\|D^2u\|_{L_\infty(Q_{7/8})} + \|Du\|_{L_\infty(Q_{7/8})} + [\nabla u]^x_{\alpha;Q_{15/16}} + \|u\|^x_{\alpha;(-1,0)\times\mathbb{R}^d} + [f]^x_{\alpha;Q_{3/4}}).$$

The same interpolation arguments of the proof of Theorem 1.1 in [Dong and Zhang 2016b] lead to

$$[u]^x_{\sigma+\alpha;Q_{1/2}} \le C(\|u\|^x_{\alpha;(-1,0)\times\mathbb{R}^d} + [f]^x_{\alpha;Q_{3/4}}).$$

Then as in the proof of Proposition 3.1, see also [Dong and Zhang 2016b, Corollary 4.6], applying this estimate to the equation of $\tilde{v} := \tilde{\eta}(u(t,x)-u(t,0))$, where $\tilde{\eta} \in C_0^\infty(\hat{Q}_{\frac{15}{16}})$ satisfying $\tilde{\eta}=1$ in $Q_{\frac{3}{4}}$, we have the desired estimates for $[u]^x_{\alpha+\sigma;Q_{1/2}}$. □

Proposition 5.2. *Suppose (1-7) is satisfied in Q_2. Then under the conditions of Theorem 1.5, we have*

$$[u]^x_{\sigma;Q_{1/2}} \le C\|u\|^x_\alpha + C\sum_{k=1}^{\infty}\omega_f(2^{-k}), \tag{5-1}$$

where $C>0$ is a constant depending only on d, λ, Λ, ω_a, ω_b, N_0 and σ.

Proof. We will consider three cases separately.

Case 1: $\sigma \in (0,1)$. For $k \in \mathbb{N}$, let v be the solution of

$$\begin{cases} \partial_t v = L(t,0)v + f(t,0) & \text{for } x \in B_{2^{-k}}, \text{and almost every } t \in (-2^{-\sigma k},0], \\ \quad v = u & \text{in } ((-2^{-k\sigma},0)\times B^c_{2^{-k}}) \cup (\{t=-2^{-k\sigma}\}\times B_{2^{-k}}). \end{cases} \tag{5-2}$$

We sketch the proof of the existence of such v as follows. Let $K^\varepsilon(t,0,y)$ and $f^\varepsilon(t,0)$ be the mollifications of $K(t,0,y)$ and $f(t,0)$ in t. Then there exists v^ε satisfying

$$\begin{cases} \partial_t v^\varepsilon = L^\varepsilon(t,0)v^\varepsilon + f^\varepsilon(t,0) & \text{in } Q_{2^{-k}}, \\ \quad v^\varepsilon = u & \text{in } ((-2^{-k\sigma},0)\times B^c_{2^{-k}}) \cup (\{t=-2^{-k\sigma}\}\times B_{2^{-k}}). \end{cases} \tag{5-3}$$

Since this equation is uniformly elliptic, we have the global uniform Hölder estimate of v^ε, which is independent of ε. Thus, there exists a subsequence converging locally uniformly to a global Hölder continuous function v. On the other hand, by Proposition 5.1, we can reselect a subsequence such that for almost every time, they converge to v locally uniformly in $C_x^{\sigma+\alpha}(B_{2^{-k}})$. Since we have from (5-3) that for all $t \in (-2^{-k\sigma},0]$,

$$\begin{cases} v^\varepsilon(t,x) = u(-2^{-k\sigma},x) + \int_{-2^{-k\sigma}}^t L^\varepsilon(\tau,0)v^\varepsilon(\tau,x)\,d\tau + \int_{-2^{-k\sigma}}^t f^\varepsilon(\tau,0)\,d\tau & \text{in } Q_{2^{-k}}, \\ \qquad v^\varepsilon = u & \text{in } ((-2^{-k\sigma},0)\times B^c_{2^{-k}}) \cup (\{t=-2^{-k\sigma}\}\times B_{2^{-k}}), \end{cases}$$

we can send $\varepsilon \to 0$, using the dominated convergence theorem, to obtain

$$\begin{cases} v(t,x) = u(-2^{-k\sigma},x) + \int_{-2^{-k\sigma}}^t L(\tau,0)v(\tau,x)\,d\tau + \int_{-2^{-k\sigma}}^t f(\tau,0)\,d\tau & \text{in } Q_{2^{-k}}, \\ \qquad v = u & \text{in } ((-2^{-k\sigma},0)\times B^c_{2^{-k}}) \cup (\{t=-2^{-k\sigma}\}\times B_{2^{-k}}). \end{cases}$$

This proves (5-2). Moreover, we have from the estimates of v^ε in Proposition 5.1 by sending $\varepsilon \to 0$, that

$$[v]^x_{\alpha+\sigma;Q_{2^{-k-1}}} \le C\sum_{j=1}^{\infty} 2^{(k-j)\sigma}M_j + C2^{k\sigma}[v]^x_{\alpha;Q_{2^{-k}}}, \tag{5-4}$$

where $\alpha \in (0,1)$ satisfying $\sigma+\alpha<1$ and

$$M_j = \sup_{\substack{(t,x),(t,x')\in(-2^{-k\sigma},0)\times B_{2^{j-k}} \\ 0<|x-x'|<2^{-k+1}}} \frac{|u(t,x)-u(t,x')|}{|x-x'|^\alpha}.$$

Let $k_0 \geq 1$ be an integer to be specified. From (5-4), we have

$$[v]^x_{\alpha;Q_{2^{-k-k_0}}} \leq C2^{-(k+k_0)\sigma} \sum_{j=1}^{k} 2^{(k-j)\sigma} M_j + C2^{-(k+k_0)\sigma}[u]^x_\alpha + C2^{-k_0\sigma}[v]^x_{\alpha;Q_{2^{-k}}}. \tag{5-5}$$

Let $w := u - v$ which satisfies

$$\begin{cases} \partial_t w = L(t,0)w + C_k & \text{in } Q_{2^{-k}}, \\ \quad w = 0 & \text{in } ((-2^{-k\sigma}, 0) \times B^c_{2^{-k}}) \cup (\{t = -2^{-k\sigma}\} \times B_{2^{-k}}), \end{cases}$$

where

$$C_k(t,x) = f(t,x) - f(t,0) + (L(t,x) - L(t,0))u.$$

It is easily seen that

$$\|C_k\|_{L_\infty(Q_{2^{-k}})} \leq \omega_f(2^{-k}) + C\omega_a(2^{-k})\Bigg(\sup_{(t_0,x_0)\in Q_{2^{-k}}} \sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)}[u]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty}\Bigg).$$

Then by the Hölder estimate [Dong and Zhang 2016b, Lemma 2.5], we have

$$\begin{aligned} [w]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}} &\leq C2^{-k(\sigma-\alpha)} C_k \\ &\leq C2^{-k(\sigma-\alpha)}\Bigg[\omega_f(2^{-k}) + \omega_a(2^{-k})\Bigg(\sup_{(t_0,x_0)\in Q_{2^{-k}}} \sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)}[u]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty}\Bigg)\Bigg]. \end{aligned} \tag{5-6}$$

Combining (5-5) and (5-6) yields

$$\begin{aligned} &2^{(k+k_0)(\sigma-\alpha)}[u]^x_{\alpha;Q_{2^{-k-k_0}}} \\ &\quad\leq C2^{-(k+k_0)\alpha} \sum_{j=1}^{k} 2^{(k-j)\sigma}[u]^x_{\alpha;(-2^{-k\sigma},0)\times B_{2^{j-k}}} \\ &\qquad + C2^{-(k+k_0)\alpha}[u]^x_\alpha + C2^{-k_0\alpha+k(\sigma-\alpha)}[u]^x_{\alpha;(-2^{-k\sigma},0)\times B_{2^{-k}}} + C2^{k_0(\sigma-\alpha)}\omega_f(2^{-k}) \\ &\qquad + C2^{k_0(\sigma-\alpha)}\omega_a(2^{-k})\Bigg(\sup_{(t_0,x_0)\in Q_{2^{-k}}} \sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)}[u]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty}\Bigg). \end{aligned} \tag{5-7}$$

Let Q^l, $l = \ell_0, \ell_0+1, \dots,$ be those in the proof of Proposition 3.2. By translation of the coordinates, from (5-7) we have for $l \geq \ell_0$, $k \geq l+1$,

$$\begin{aligned} &2^{(k+k_0)(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} [u]^x_{\alpha;Q_{2^{-k-k_0}}(t_0,x_0)} \\ &\quad\leq C2^{-(k+k_0)\alpha} \sup_{(t_0,x_0)\in Q^l} \sum_{j=0}^{k} 2^{(k-j)\sigma}[u]^x_{\alpha;(t_0-2^{-k\sigma},t_0)\times B_{2^{j-k}}(x_0)} + C2^{-(k+k_0)\alpha}[u]^x_\alpha \\ &\qquad + C2^{k_0(\sigma-\alpha)}\Bigg[\omega_f(2^{-k}) + \omega_a(2^{-k})\Bigg(\sup_{(t_0,x_0)\in Q^{l+1}} \sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)}[u]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty}\Bigg)\Bigg]. \end{aligned} \tag{5-8}$$

Then we take the sum (5-8) in $k = l+1, l+2, \dots$ to obtain

$$\begin{aligned}\sum_{k=l+1}^{\infty} 2^{(k+k_0)(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} [u]^x_{\alpha;Q_{2^{-k-k_0}}(t_0,x_0)} &\le C\sum_{k=l+1}^{\infty} 2^{-(k+k_0)\alpha} \sup_{(t_0,x_0)\in Q^l} \sum_{j=0}^{k} 2^{(k-j)\sigma}[u]^x_{\alpha;(t_0-2^{-k\sigma},t_0)\times B_{2^{j-k}}(x_0)} \\ &\quad + C2^{-(l+k_0)\alpha}[u]^x_\alpha + C2^{k_0(\sigma-\alpha)} \sum_{k=l+1}^{\infty} \omega_f(2^{-k}) \\ &\quad + C2^{k_0(\sigma-\alpha)} \sum_{k=l+1}^{\infty} \omega_a(2^{-k}) \left(\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} [u]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty} \right).\end{aligned}$$

As before, by switching the order of summations and then replacing k by $k+j$, the first term on the right-hand side is bounded by

$$C2^{-k_0\alpha} \sum_{k=0}^{\infty} 2^{k(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} [u]^x_{\alpha;Q_{2^{-k}}(t_0,x_0)}.$$

With the above inequality, we have

$$\begin{aligned}\sum_{k=l+1}^{\infty} 2^{(k+k_0)(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} [u]^x_{\alpha;Q_{2^{-k-k_0}}(t_0,x_0)} &\le C2^{-k_0\alpha} \sum_{k=0}^{\infty} 2^{k(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} [u]^x_{\alpha;Q_{2^{-k}}(t_0,x_0)} + C2^{-(l+k_0)\alpha}[u]^x_\alpha + C2^{k_0(\sigma-\alpha)} \sum_{k=l+1}^{\infty} \omega_f(2^{-k}) \\ &\quad + C2^{k_0(\sigma-\alpha)} \sum_{k=l+1}^{\infty} \omega_a(2^{-k}) \left(\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} [u]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty} \right).\end{aligned}$$

The bound above, together with the obvious inequality

$$\sum_{j=0}^{l+k_0} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} [u]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)} \le C2^{(l+k_0)(\sigma-\alpha)}[u]^x_\alpha,$$

implies

$$\begin{aligned}\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} [u]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)} &\le C2^{-k_0\alpha} \sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} [u]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)} \\ &\quad + C2^{-(l+k_0)\alpha}[u]^x_\alpha + C2^{(l+k_0)(\sigma-\alpha)}[u]^x_\alpha + C2^{k_0(\sigma-\alpha)} \sum_{k=l+1}^{\infty} \omega_f(2^{-k}) \\ &\quad + C2^{k_0(\sigma-\alpha)} \sum_{k=l+1}^{\infty} \omega_a(2^{-k}) \left(\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} [u]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)} + \|u\|_{L_\infty} \right).\end{aligned}$$

By first choosing k_0 sufficiently large and then ℓ_0 sufficiently large (recalling $l \ge \ell_0$), we get

$$\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} [u]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)} \le \frac{1}{4}\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} [u]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)} + C2^{(l+k_0)(\sigma-\alpha)}\|u\|^x_\alpha + C\sum_{k=1}^{\infty}\omega_f(2^{-k}).$$

This implies

$$\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} [u]^x_{\alpha;Q_{2^{-j}}(t_0,x_0)} \le C\|u\|^x_\alpha + C\sum_{k=1}^{\infty}\omega_f(2^{-k}),$$

which together with Lemma 2.1(i) gives (5-1).

Case 2: $\sigma \in (1,2)$. For $k \in \mathbb{N}$, let v_M be the solution of

$$\begin{cases} \partial_t v_M = L(t,0)v_M + f(t,0) + b(t,0)Du(t,0) - \partial_t p_0 & \text{in } Q_{2^{-k}}, \\ v_M = g_M & \text{in } ((-2^{-k\sigma},0)\times B^c_{2^{-k}}) \cup (\{t=-2^{-k\sigma}\}\times B_{2^{-k}}), \end{cases}$$

where $M \ge 2\|u-p_0\|_{L_\infty(Q_{2^{-k}})}$ is a constant to be specified later,

$$g_M = \max(\min(u-p_0, M), -M),$$

and $p_0 = p_0(t,x)$ is the first-order Taylor's expansion of $u^{(2^{-k})}$ in x at $(t,0)$, and $u^{(2^{-k})}$ is the mollification of u in the x-variable only:

$$u^{(R)}(t,x) = \int_{\mathbb{R}^d} u(t,x-Ry)\zeta(y)\,dy$$

with $\zeta \in C_0^\infty(B_1)$ being a radial nonnegative function with unit integral.

By Proposition 5.1, we have

$$[v_M]^x_{\alpha+\sigma;Q_{2^{-k-1}}} \le C\sum_{j=1}^{\infty} 2^{(k-j)\sigma} M_j + C2^{k\sigma}[v_M]^x_{\alpha;Q_{2^{-k}}},$$

where $\alpha \in (0, \min\{2-\sigma, (\sigma-1)/2\})$ and

$$M_j = \sup_{\substack{(t,x),(t,x')\in(-2^{-k\sigma},0)\times B_{2^{j-k}}\\ 0<|x-x'|<2^{-k+1}}} \frac{|u(t,x)-p_0(t,x)-u(t,x')+p_0(t,x')|}{|x-x'|^\alpha}.$$

From Lemma 2.4 with $\sigma \in (1,2)$, it follows that for $j > k$, we have

$$M_j \le C[u]^x_{\alpha;(-2^{-k\sigma},0)\times\mathbb{R}^d},$$

and thus,

$$\begin{aligned}[v_M]^x_{\alpha+\sigma;Q_{2^{-k-1}}} &\le C\sum_{j=1}^{\infty} 2^{(k-j)\sigma} M_j + C2^{k\sigma}[v_M]^x_{\alpha;Q_{2^{-k}}} \\ &\le C\sum_{j=1}^{k} 2^{(k-j)\sigma} M_j + C[u]^x_{\alpha;(-2^{-k\sigma},0)\times\mathbb{R}^d} + C2^{k\sigma}[v_M]^x_{\alpha;Q_{2^{-k}}}. \end{aligned} \tag{5-9}$$

From (5-9), and the mean value formula (recalling that $\alpha < 2-\sigma$),

$$\|v_M - p_1\|_{L_\infty(Q_{2^{-k-k_0}})} \le C2^{-(k+k_0)(\sigma+\alpha)} \sum_{j=1}^{k} 2^{(k-j)\sigma} M_j + C2^{-(k+k_0)(\sigma+\alpha)}[u]^x_{\alpha;(-2^{-k\sigma},0)\times\mathbb{R}^d} + C2^{-k\alpha-k_0(\sigma+\alpha)}[v_M]^x_{\alpha;Q_{2^{-k}}},$$

where p_1 is the first-order Taylor's expansion of v_M in x at $(t,0)$. The above inequality, (5-9), and the interpolation inequality imply

$$[v_M - p_1]^x_{\alpha;Q_{2^{-k-k_0}}} \le C2^{-(k+k_0)\sigma} \sum_{j=1}^{k} 2^{(k-j)\sigma} M_j + C2^{-(k+k_0)\sigma}[u]^x_{\alpha;(-2^{-k\sigma},0)\times\mathbb{R}^d} + C2^{-k_0\sigma}[v_M]^x_{\alpha;Q_{2^{-k}}}. \tag{5-10}$$

Next $w_M := g_M - v_M$, which equals $u - p_0 - v_M$ in $Q_{2^{-k}}$, satisfies

$$\begin{cases} \partial_t w_M = L(t,0)w_M + h_M + C_k & \text{in } Q_{2^{-k}}, \\ w_M = 0 & \text{in } ((-2^{-k\sigma},0)\times B^c_{2^{-k}}) \cup (\{t=-2^{-k\sigma}\}\times B_{2^{-k}}), \end{cases}$$

where

$$h_M := L(t,0)(u - p_0 - g_M)$$

and

$$C_k = f - f(t,0) + bDu - b(t,0)Du(t,0) + (L - L(t,0))u.$$

It follows easily that

$$\begin{aligned} |C_k| \le{}& \omega_f(2^{-k}) + \omega_b(2^{-k})\|Du\|_{L_\infty(Q_{2^{-k}})} + \|b\|_{L_\infty} 2^{-k\alpha}[Du]^x_{\alpha;Q_{2^{-k}}} \\ &+ C\omega_a(2^{-k}) \sup_{(t_0,x_0)\in Q_{2^{-k}}} \sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{t\in(t_0-2^{-j\sigma},t_0)} [u(t,\cdot) - p_{t,x_0}]^x_{\alpha;B_{2^{-j}}(x_0)} \\ &+ C\omega_a(2^{-k})(\|Du\|_{L_\infty(Q_{2^{-k}})} + \|u\|_{L_\infty}), \end{aligned}$$

where $p_{t,x_0} = p_{t,x_0}(x)$ is the first-order Taylor's expansion of u with respect to x at (t,x_0). From Lemma 2.3, we obtain

$$\begin{aligned} |C_k| \le{}& \omega_f(2^{-k}) + \omega_b(2^{-k})\|Du\|_{L_\infty(Q_{2^{-k}})} + \|b\|_{L_\infty} 2^{-k\alpha}[Du]^x_{\alpha;Q_{2^{-k}}} \\ &+ C\omega_a(2^{-k}) \sup_{(t_0,x_0)\in Q_{2^{-k}}} \sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{t\in(t_0-2^{-j\sigma},t_0)} \inf_{p\in\mathcal{P}_x} [u(t,\cdot) - p]^x_{\alpha;B_{2^{-j}}(x_0)} \\ &+ C\omega_a(2^{-k})(\|Du\|_{L_\infty(Q_{2^{-k}})} + \|u\|_{L_\infty}). \end{aligned}$$

By the dominated convergence theorem, it is easy to see that

$$\|h_M\|_{L_\infty(Q_{2^{-k}})} \to 0 \quad \text{as } M \to \infty.$$

Thus, similar to (3-6), choosing M sufficiently large so that

$$\|h_M\|_{L_\infty(Q_{2^{-k}})} \le \tfrac{1}{2} C_k,$$

we have

$$\begin{aligned}
&[w_M]_{\frac{\alpha}{\sigma},\alpha;Q_{2^{-k}}} \\
&\le C 2^{-k(\sigma-\alpha)} \Big[\omega_f(2^{-k}) + (\omega_b(2^{-k}) + \omega_a(2^{-k})) \|Du\|_{L_\infty(Q_{2^{-k}})} + 2^{-k\alpha}[Du]^x_{\alpha;Q_{2^{-k}}} \\
&\qquad + \omega_a(2^{-k}) \Big(\sup_{(t_0,x_0)\in Q_{2^{-k}}} \sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{t\in(t_0-2^{-j\sigma},t_0)} \inf_{p\in\mathcal{P}_x} [u(t,\cdot)-p]^x_{\alpha;B_{2^{-j}}(x_0)} + \|u\|_{L_\infty} \Big) \Big]. \qquad (5\text{-}11)
\end{aligned}$$

From Lemma 2.4 (more precisely, its proof)

$$M_j \le C \sup_{t\in(-2^{-k\sigma},0)} \inf_{p\in\mathcal{P}_x} [u(t,\cdot) - p]^x_{\alpha;B_{2^{j-k}}}. \qquad (5\text{-}12)$$

From the triangle inequality and Lemma 2.4 with $j = 0$,

$$\begin{aligned}
[v_M]^x_{\alpha;Q_{2^{-k}}} &\le [w_M]^x_{\alpha;Q_{2^{-k}}} + [u - p_0]^x_{\alpha;Q_{2^{-k}}} \\
&\le [w_M]^x_{\alpha;Q_{2^{-k}}} + C \sup_{t\in(-2^{-k\sigma},0)} \inf_{p\in\mathcal{P}_x} [u(t,\cdot) - p]^x_{\alpha;B_{2^{-k}}}.
\end{aligned}$$

For $l = 1, 2, \ldots$, let $Q^l = Q_{1-2^{-l}}$. Combining (5-10), (5-11), and (5-12), similar to (5-8), we then get

$$\begin{aligned}
&2^{(k+k_0)(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \sup_{t\in(t_0-2^{-(k_0+k)\sigma},t_0)} \inf_{p\in\mathcal{P}_x} [u(t,\cdot) - p]^x_{\alpha;B_{2^{-(k_0+k)}}(x_0)} \\
&\quad\le 2^{(k+k_0)(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} [u - p_0 - p_1]^x_{\alpha;Q_{2^{-(k_0+k)}}(t_0,x_0)} \\
&\quad\le C 2^{-(k+k_0)\alpha} \sup_{(t_0,x_0)\in Q^l} \sum_{j=1}^{k} 2^{(k-j)\sigma} \sup_{t\in(t_0-2^{-k\sigma},t_0)} \inf_{p\in\mathcal{P}_x} [u(t,\cdot) - p]^x_{\alpha;B_{2^{j-k}}(x_0)} \\
&\qquad + C 2^{-(k+k_0)\alpha}[u]^x_\alpha + \sup_{(t_0,x_0)\in Q^l} 2^{-k\alpha+k_0(\sigma-\alpha)}[Du]^x_{\alpha;Q_{2^{-k}}(t_0,x_0)} \\
&\qquad + C 2^{k_0(\sigma-\alpha)} \Big[\omega_f(2^{-k}) + (\omega_b(2^{-k}) + \omega_a(2^{-k})) \|Du\|_{L_\infty(Q^{l+1})} \\
&\qquad + \omega_a(2^{-k}) \Big(\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \sup_{t\in(t_0-2^{-j\sigma},t_0)} \inf_{p\in\mathcal{P}_x} [u(t,\cdot) - p]^x_{\alpha;B_{2^{-j}}(x_0)} + \|u\|_{L_\infty} \Big) \Big].
\end{aligned}$$

Summing the above inequality in $k = l + 1, l + 2, \ldots$ as before, we obtain

$$\begin{aligned}
&\sum_{k=l+1}^{\infty} 2^{(k+k_0)(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \sup_{t\in(t_0-2^{-(k_0+k)\sigma},t_0)} \inf_{p\in\mathcal{P}_x} [u(t,\cdot) - p]^x_{\alpha;B_{2^{-(k_0+k)}}(x_0)} \\
&\le C 2^{-k_0\alpha} \sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \sup_{t\in(t_0-2^{-j\sigma},t_0)} \inf_{p\in\mathcal{P}_x} [u(t,\cdot) - p]^x_{\alpha;B_{2^{-j}}(x_0)}
\end{aligned}$$

$$+C2^{-(k_0+l)\alpha}[u]^x_\alpha + C2^{k_0(\sigma-\alpha)}\sum_{k=l+1}^{\infty} 2^{-k\alpha}\sup_{(t_0,x_0)\in Q^l}[Du]^x_{\alpha;Q_{2^{-k}}(t_0,x_0)}$$
$$+C2^{k_0(\sigma-\alpha)}\bigg[\sum_{k=l+1}^{\infty}\big(\omega_f(2^{-k})+(\omega_b(2^{-k})+\omega_a(2^{-k}))\|Du\|_{L_\infty(Q^{l+1})}\big)$$
$$+\sum_{k=l+1}^{\infty}\omega_a(2^{-k})\bigg(\|u\|_{L_\infty}+\sum_{j=0}^{\infty}2^{j(\sigma-\alpha)}\sup_{(t_0,x_0)\in Q^{l+1}}\sup_{t\in(t_0-2^{-j\sigma},t_0)}\inf_{p\in\mathcal{P}_x}[u(t,\cdot)-p]^x_{\alpha;B_{2^{-j}}(x_0)}\bigg)\bigg], \tag{5-13}$$

and

$$\sum_{j=0}^{\infty}2^{j(\sigma-\alpha)}\sup_{(t_0,x_0)\in Q^{l}}\sup_{t\in(t_0-2^{-j\sigma},t_0)}\inf_{p\in\mathcal{P}_x}[u(t,\cdot)-p]^x_{\alpha;B_{2^{-j}}(x_0)}$$
$$\le C2^{-k_0\alpha}\sum_{j=0}^{\infty}2^{j(\sigma-\alpha)}\sup_{(t_0,x_0)\in Q^{l+1}}\sup_{t\in(t_0-2^{-j\sigma},t_0)}\inf_{p\in\mathcal{P}_x}[u(t,\cdot)-p]^x_{\alpha;B_{2^{-j}}(x_0)}$$
$$+C2^{k_0(\sigma-\alpha)}\|u\|_{L_\infty}+C2^{(l+k_0)(\sigma-\alpha)}[u]^x_\alpha+C2^{k_0(\sigma-\alpha)-l\alpha}[Du]^x_{\alpha;Q^{l+1}}$$
$$+C2^{k_0(\sigma-\alpha)}\sum_{k=l+1}^{\infty}\big(\omega_f(2^{-k})+(\omega_b(2^{-k})+\omega_a(2^{-k}))\|Du\|_{L_\infty(Q^{l+1})}\big)$$
$$+C2^{k_0(\sigma-\alpha)}\sum_{k=l+1}^{\infty}\omega_a(2^{-k})\sum_{j=0}^{\infty}2^{j(\sigma-\alpha)}\sup_{(t_0,x_0)\in Q^{l+1}}\sup_{t\in(t_0-2^{-j\sigma},t_0)}\inf_{p\in\mathcal{P}_x}[u(t,\cdot)-p]^x_{\alpha;B_{2^{-j}}(x_0)}.$$

By choosing k_0 and l sufficiently large, and using (2-4) and interpolation inequalities (recalling that $\alpha<(\sigma-1)/2$), we obtain

$$\sum_{j=0}^{\infty}2^{j(\sigma-\alpha)}\sup_{(t_0,x_0)\in Q^{l}}\sup_{t\in(t_0-2^{-j\sigma},t_0)}\inf_{p\in\mathcal{P}_x}[u(t,\cdot)-p]^x_{\alpha;B_{2^{-j}}(x_0)}$$
$$\le\frac14\sum_{j=0}^{\infty}2^{j(\sigma-\alpha)}\sup_{(t_0,x_0)\in Q^{l+1}}\sup_{t\in(t_0-2^{-j\sigma},t_0)}\inf_{p\in\mathcal{P}_x}[u(t,\cdot)-p]^x_{\alpha;B_{2^{-j}}(x_0)}$$
$$+C2^{(k_0+l)(\sigma-\alpha)}\|u\|^x_\alpha+C\sum_{k=1}^{\infty}\omega_f(2^{-k}).$$

Therefore,

$$\sum_{j=0}^{\infty}2^{j(\sigma-\alpha)}\sup_{(t_0,x_0)\in Q^{l}}\sup_{t\in(t_0-2^{-j\sigma},t_0)}\inf_{p\in\mathcal{P}_x}[u(t,\cdot)-p]^x_{\alpha;B_{2^{-j}}(x_0)}\le C\|u\|^x_\alpha+C\sum_{k=1}^{\infty}\omega_f(2^{-k}), \tag{5-14}$$

which together with Lemma 2.1(ii) (actually the proof of it) gives (5-1).

Case 3: $\sigma=1$. Set

$$B_0(t)=\int_t^0 b(s,0)\,ds$$

and we define

$$\hat{u}(t,x)=u(t,x+B_0(t)),\quad \hat{f}_\beta(t,x)=f_\beta(t,x+B_0(t)),\quad \text{and}\quad \hat{b}(t,x)=b(t,x+B_0(t)).$$

It is easy to see that in Q_δ for some $\delta>0$,

$$\begin{aligned}\partial_t\hat{u}(t,x)&=(\partial_t u)(t,x+B_0(t))-b(t,0)\nabla u(t,x+B_0(t))\\&=\hat{L}_\beta\hat{u}+\hat{f}_\beta+(\hat{b}-b(t,0))\nabla\hat{u},\end{aligned}\tag{5-15}$$

where $\hat{L}$ is the operator with kernel $a(t,x+B_0(t),y)|y|^{-d-\sigma}$. For $(t,x)\in Q_{2^{-k}}$,

$$|\hat{f}_\beta(t,x)-\hat{f}_\beta(t,0)|\le\omega_f(2^{-k}),$$
$$|\hat{b}-b(t,0)|\le\omega_b((1+N_0)2^{-k}).$$

Furthermore,

$$\|Du\|_{L_\infty}+\|\partial_t u\|_{L_\infty}\le(1+N_0)(\|D\hat{u}\|_{L_\infty}+\|\partial_t\hat{u}\|_{L_\infty}).$$

Therefore, it is sufficient to bound $\hat{u}$. In the rest of the proof, we estimate the solution to (5-15) and abuse the notation to use u instead of $\hat{u}$ for simplicity. By scaling, translation and covering arguments, we also assume u satisfies the equation in Q_2.

The proof is similar to the case $\sigma\in(1,2)$ and we indeed proceed as in the previous case. Take p_0 to be the first-order Taylor's expansion of $u^{(2^{-k})}$ in x at $(t,0)$, where $u^{(2^{-k})}$ is the mollification of u in the x-variable only, as in Case 2. We also assume that the solution v to the equations

$$\begin{cases}\partial_t v=\hat{L}(t,0)v+\hat{f}(t,0)-\partial_t p_0 & \text{in } Q_{2^{-k}},\\ v=u-p_0 & \text{in } ((-2^{-k\sigma},0)\times B^c_{2^{-k}})\cup(\{t=-2^{-k\sigma}\}\times B_{2^{-k}}),\end{cases}$$

exists without carrying out another approximation argument. By Proposition 5.1 with $\sigma=1$ and Lemma 2.4 in [Dong and Zhang 2016a],

$$\begin{aligned}[v]^x_{1+\alpha;Q_{2^{-k-1}}}&\le C\sum_{j=1}^{\infty}2^{k-j}M_j+C2^k[v]_{\alpha,\alpha;Q_{2^{-k}}}\\&\le C\sum_{j=1}^{\infty}2^{k-j}\sup_{t\in(-2^{-k\sigma},0)}\inf_{p\in\mathcal{P}_x}[u(t,\cdot)-p]^x_{\alpha;B_{2^{j-k}}}+C2^k[v]^x_{\alpha;Q_{2^{-k}}}\\&\le C\sum_{j=1}^{k}2^{k-j}\sup_{t\in(-2^{-k\sigma},0)}\inf_{p\in\mathcal{P}_x}[u(t,\cdot)-p]^x_{\alpha;B_{2^{j-k}}}+C[u]^x_\alpha+C2^k[v]^x_{\alpha;Q_{2^{-k}}}.\end{aligned}\tag{5-16}$$

From (5-16) and the interpolation inequality, we obtain

$$\begin{aligned}&[v-p_1]^x_{\alpha;Q_{2^{-k-k_0}}}\\&\le C2^{-(k+k_0)}\sum_{j=1}^{k}2^{k-j}\sup_{t\in(-2^{-k\sigma},0)}\inf_{p\in\mathcal{P}_x}[u(t,\cdot)-p]^x_{\alpha;B_{2^{j-k}}}+C2^{-k_0}[v]^x_{\alpha;Q_{2^{-k}}}+C2^{-(k+k_0)}[u]^x_\alpha,\end{aligned}\tag{5-17}$$

where p_1 is the first-order Taylor's expansion of v in x at $(t,0)$. Next $w := u - p_0 - v$ satisfies

$$\begin{cases} \partial_t w = \hat{L}(t,0)w + C_k & \text{in } Q_{2^{-k}}, \\ \quad w = 0 & \text{in } ((-2^{-k},0)\times B^c_{2^{-k}}) \cup (\{t=-2^{-k}\}\times B_{2^{-k}}), \end{cases}$$

where by the cancellation property,

$$C_k = \hat{f} - \hat{f}(t,0) + (\hat{b} - b(t,0))\nabla u + (\hat{L} - \hat{L}(t,0))u,$$

so that

$$\begin{aligned} |C_k| \le{}& \omega_f(2^{-k}) + \omega_b((1+N_0)2^{-k})\|Du\|_{L_\infty(Q_{2^{-k}})} \\ &+ C\omega_a((1+N_0)2^{-k})\Bigg(\sup_{(t_0,x_0)\in Q_{2^{-k}}}\sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{t\in(t_0-2^{-j\sigma},t_0)} \inf_{p\in\mathcal{P}_x}[u(t,\cdot)-p]^x_{\alpha;B_{2^{-j}}(x_0)} + \|u\|_{L_\infty}\Bigg). \end{aligned}$$

Clearly, for any $r \ge 0$,

$$\omega_\bullet((1+N_0)r) \le (2+N_0)\omega_\bullet(r).$$

Therefore, similar to (5-11), we have

$$\begin{aligned} &[w]_{\alpha,\alpha;Q_{2^{-k}}} \\ &\le C2^{-k(1-\alpha)}\Bigg[\omega_f(2^{-k}) + \omega_b(2^{-k})\|Du\|_{L_\infty(Q_{2^{-k}})} \\ &\qquad + \omega_a(2^{-k})\Bigg(\sup_{(t_0,x_0)\in Q_{2^{-k}}}\sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{t\in(t_0-2^{-j\sigma},t_0)} \inf_{p\in\mathcal{P}_x}[u(t,\cdot)-p]^x_{\alpha;B_{2^{-j}}(x_0)} + \|u\|_{L_\infty}\Bigg)\Bigg]. \end{aligned} \tag{5-18}$$

From the proof of (2-16) and the triangle inequality,

$$[v]^x_{\alpha;Q_{2^{-k}}} \le [w]^x_{\alpha;Q_{2^{-k}}} + [u-p_0]^x_{\alpha;Q_{2^{-k}}} \le [w]_{\alpha;Q_{2^{-k}}} + C \sup_{t\in(-2^{-k},0)} \inf_{p\in\mathcal{P}_x}[u(t,\cdot)-p]^x_{\alpha;B_{2^{-k}}}.$$

For $l = 1,2,\dots$, let $Q^l = Q_{1-2^{-l}}$. Similar to (5-8), by combining (5-17) and (5-18), shifting the coordinates, and using the above inequality, we obtain for $l \ge 1$ and $k \ge l+1$,

$$\begin{aligned} &2^{(k+k_0)(1-\alpha)} \sup_{(t_0,x_0)\in Q^l} \sup_{t\in(t_0-2^{-(k+k_0)\sigma},t_0)} \inf_{p\in\mathcal{P}_x}[u(t,\cdot)-p]^x_{\alpha;B_{2^{-k-k_0}}(x_0)} \\ &\le C2^{-(k+k_0)\alpha}\sum_{j=0}^{k} 2^{k-j} \sup_{(t_0,x_0)\in Q^l} \sup_{t\in(t_0-2^{-k\sigma},t_0)} \inf_{p\in\mathcal{P}_x}[u(t,\cdot)-p]^x_{\alpha;B_{2^{j-k}}(x_0)} \\ &\quad + C2^{k_0(1-\alpha)}\Bigg[\omega_f(2^{-k}) + \omega_b(2^{-k})\|Du\|_{L_\infty(Q^{l+1})} \\ &\qquad + \omega_a(2^{-k})\Bigg(\sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \sup_{t\in(t_0-2^{-j\sigma},t_0)} \inf_{p\in\mathcal{P}_x}[u(t,\cdot)-p]^x_{\alpha,B_{2^{-j}}(x_0)} + \|u\|_{L_\infty}\Bigg)\Bigg] \\ &\quad + C2^{-(k+k_0)\alpha}[u]^x_\alpha, \end{aligned} \tag{5-19}$$

which by summing in $k = l+1, l+2, \dots$ implies

$$\sum_{k=l+1}^{\infty} 2^{(k+k_0)(1-\alpha)} \sup_{(t_0,x_0)\in Q^l} \sup_{t\in(t_0-2^{-(k+k_0)\sigma},t_0)} \inf_{p\in\mathcal{P}_x} [u(t,\cdot)-p]^x_{\alpha;B_{2^{-k-k_0}}(x_0)}$$
$$\le C2^{-k_0\alpha}\sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \sup_{t\in(t_0-2^{-j\sigma},t_0)} \inf_{p\in\mathcal{P}_x} [u(t,\cdot)-p]^x_{\alpha;B_{2^{-j}}(x_0)}$$
$$+C2^{-(k_0+l)\alpha}[u]^x_\alpha + C2^{k_0(1-\alpha)} \sum_{k=l+1}^{\infty} \omega_f(2^{-k})$$
$$+C2^{k_0(1-\alpha)} \sum_{k=l+1}^{\infty} \bigg[\omega_b(2^{-k})\|Du\|_{L_\infty(Q^{l+1})}$$
$$+\omega_a(2^{-k})\bigg(\sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \sup_{t\in(t_0-2^{-j\sigma},t_0)} \inf_{p\in\mathcal{P}_x} [u(t,\cdot)-p]^x_{\alpha;B_{2^{-j}}(x_0)} + \|u\|_{L_\infty}\bigg)\bigg],$$

where for the first term on the right-hand side of (5-19), we replaced j by $k-j$ and switched the order of the summation as before. Therefore,

$$\sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{(t_0,x_0)\in Q^l} \sup_{t\in(t_0-2^{-j\sigma},t_0)} \inf_{p\in\mathcal{P}_x} [u(t,\cdot)-p]^x_{\alpha;B_{2^{-j}}(x_0)}$$
$$\le C2^{-k_0\alpha}\sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \sup_{t\in(t_0-2^{-j\sigma},t_0)} \inf_{p\in\mathcal{P}_x} [u(t,\cdot)-p]^x_{\alpha;B_{2^{-j}}(x_0)}$$
$$+C2^{(l+k_0)(1-\alpha)}[u]^x_\alpha + C2^{k_0(1-\alpha)} \sum_{k=l+1}^{\infty} \omega_f(2^{-k})$$
$$+C2^{k_0(1-\alpha)} \sum_{k=l+1}^{\infty} \omega_b(2^{-k})\|Du\|_{L_\infty(Q^{l+1})}$$
$$+C2^{k_0(1-\alpha)} \sum_{k=l+1}^{\infty} \omega_a(2^{-k})\bigg(\sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \sup_{t\in(t_0-2^{-j\sigma},t_0)} \inf_{p\in\mathcal{P}_x} [u(t,\cdot)-p]^x_{\alpha;B_{2^{-j}}(x_0)} + \|u\|_{L_\infty}\bigg).$$

Then we choose k_0 and then l sufficiently large, and apply Lemma 2.1(iii) (actually the proof it) to get

$$\sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{(t_0,x_0)\in Q^l} \sup_{t\in(t_0-2^{-j\sigma},t_0)} \inf_{p\in\mathcal{P}_x} [u(t,\cdot)-p]^x_{\alpha;B_{2^{-j}}(x_0)}$$
$$\le \frac{1}{4}\sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{(t_0,x_0)\in Q^{l+1}} \sup_{t\in(t_0-2^{-j\sigma},t_0)} \inf_{p\in\mathcal{P}_x} [u(t,\cdot)-p]^x_{\alpha;B_{2^{-j}}(x_0)}$$
$$+ C2^{(k_0+l)(1-\alpha)}\|u\|^x_\alpha + C\sum_{k=1}^{\infty} \omega_f(2^{-k}).$$

This implies

$$\sum_{j=0}^{\infty} 2^{j(1-\alpha)} \sup_{(t_0,x_0)\in Q^l} \sup_{t\in(t_0-2^{-j\sigma},t_0)} \inf_{p\in\mathcal{P}_x} [u(t,\cdot)-p]^x_{\alpha;B_{2^{-j}}(x_0)} \le C\|u\|^x_\alpha + C\sum_{k=1}^{\infty}\omega_f(2^{-k}),$$

from which (5-1) follows. □

Proof of Theorem 1.5. As before, Theorem 1.5 follows from Proposition 5.2 using the argument of freezing the coefficients. We only present the detailed proof of Theorem 1.5 for $\sigma \in (1,2)$. We omit the proof of the case $\sigma \in (0,1]$ since it is almost the same and actually is simpler.

Indeed, the proof here for $\sigma \in (1,2)$ is almost identical to that of Theorem 1.1, so we just sketch it.

Without loss of generality, we assume the equation holds in Q_3.

Step 1: For $k = 1, 2, \dots$, define $Q^k := Q_{1-2^{-k}}$. Let $\eta_k \in C_0^\infty(\hat{Q}^{k+3})$ be a sequence of nonnegative smooth cutoff functions satisfying $\eta \equiv 1$ in Q^{k+2}, $|\eta| \le 1$ in Q^{k+3}, and $\|\partial_t^j D^i \eta_k\|_{L_\infty} \le C 2^{k(i+j)}$ for each $i, j \ge 0$. Set $v_k := u\eta_k \in C^{1,\sigma+}$ and notice that in Q^{k+1},

$$\begin{aligned}\partial_t v_k &= \eta_k \partial_t u + \partial_t \eta_k u = \eta_k L u + \eta_k b D u + \eta_k f + \partial_t \eta_k u \\ &= L v_k + b D v_k - b u D\eta_k + h_k + \eta_k f + \partial_t \eta_k u,\end{aligned}$$

where

$$h_k = \eta_k L u - L v_k = \int_{\mathbb{R}^d} \frac{\xi_k(t,x,y)a(t,x,y)}{|y|^{d+\sigma}}\,dy,$$

and

$$\begin{aligned}\xi_k(t,x,y) &= u(t,x+y)(\eta_k(t,x+y)-\eta_k(t,x)) - y\cdot D\eta_k(t,x)u(t,x) \\ &= u(t,x+y)(\eta_k(t,x+y)-\eta_k(t,x)) \quad \text{since } D\eta_k \equiv 0 \text{ in } Q^{k+1}.\end{aligned}$$

We will apply Proposition 5.2 to the equation of v_k in Q^{k+1} and obtain corresponding estimates for v_k in Q^k.

As before, we have $\eta_k f \equiv f$, $\partial_t \eta_k u \equiv 0$, and $b u D\eta_k \equiv 0$ in Q^{k+1}. Thus, we only need to estimate the moduli of continuity of h_k in Q^{k+1} with respect to x. The same proof of (3-31) shows that

$$\omega_h(r) := C\left(2^{\sigma k}\|u\|_{L_\infty(Q_3)} + \sum_{j=0}^{\infty} 2^{-j\sigma}\omega_u(2^j)\right)\omega_a(r) + C 2^{k\sigma}\omega_u(r) + C 2^{k(\sigma+1)}\|u\|_{L_\infty(Q_3)} r. \tag{5-20}$$

As in the proof of Theorem 1.1, by making use of Proposition 5.2 to v_k and interpolation inequalities, an iteration procedure will lead to

$$[u]^x_{\sigma;Q^4} \le C\|u\|_{L_\infty(Q_3)} + C\sum_{j=0}^{\infty}(2^{-j\sigma}\omega_u(2^j) + \omega_u(2^{-j}) + \omega_f(2^{-j})). \tag{5-21}$$

Applying this to the equation of $u(t,x) - u(t,0)$ gives to (1-10).

Finally, since $\|v_1\|^x_\alpha$ is bounded by the right-hand side of (5-21), from (5-14), we see that

$$\sum_{j=0}^{\infty} 2^{j(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^l} \sup_{t\in(t_0-2^{-j\sigma},t_0)} \inf_{p\in\mathcal{P}_x} [u(t,\cdot)-p]^x_{\alpha;B_{2^{-j}}(x_0)} \le C$$

for some large l.

This and (5-13) with u replaced by v_1 and f replaced by $h_1+\eta_1 f+\partial_t\eta_1 u-buD\eta_1$ give

$$\sum_{j=k_1+1}^{\infty} 2^{(j+k_0)(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^{k_1}} \sup_{t\in(t_0-2^{-(j+k_0)\sigma},\,t_0)} \inf_{p\in\mathcal{P}_x} [v_1-p]^x_{\alpha;B_{2^{-j-k_0}}(x_0)} \le C2^{-k_0\alpha}+C2^{k_0(\sigma-\alpha)}\sum_{j=k_1}^{\infty}(\omega_f(2^{-j})+\omega_a(2^{-j})+\omega_u(2^{-j})+\omega_b(2^{-j})+2^{-j\alpha}).$$

Here we also used (5-20) with $k=1$. Therefore, for any small $\varepsilon>0$, we can find k_0 sufficiently large, then k_1 sufficiently large, depending only on C, σ, N_0, α, ω_a, ω_f, ω_b, and ω_u, such that

$$\sum_{j=k_1+1}^{\infty} 2^{(j+k_0)(\sigma-\alpha)} \sup_{(t_0,x_0)\in Q^{k_1}} \sup_{t\in(t_0-2^{-(j+k_0)\sigma},\,t_0)} \inf_{p\in\mathcal{P}_x} [v_1-p]^x_{\alpha;B_{2^{-j-k_0}}(x_0)} < \varepsilon.$$

This, together with the fact that $v_1=u$ in $Q_{\frac{1}{2}}$ and the proof of Lemma 2.1(ii), indicates that

$$\sup_{(t_0,x_0)\in Q_{1/2}} [u]^x_{\sigma;Q_r(t_0,x_0)} \to 0 \quad \text{as } r\to 0$$

with a decay rate depending only on d, λ, N_0, Λ, ω_a, ω_f, ω_b, ω_u, and σ. Also, by evaluating (1-7) on both sides and making use of the dominated convergence theorem, we have that $\partial_t u$ is uniformly continuous in x in $Q_{\frac{1}{2}}$ with a modulus of continuity controlled by d, σ, λ, Λ, ω_a, ω_f, ω_u, N_0, ω_b, and $\|u\|_{L_\infty}$.

Hence, the proof of the case when $\sigma\in(1,2)$ is completed. □

Proof of Theorem 1.6. Given the proofs of Theorems 1.3 and 1.4, Theorem 1.6 can be similarly proved (actually simpler), and we omit the details. □

Acknowledgments

Part of this work was completed while H. Dong was visiting the Department of Mathematics at the Hong Kong University of Science and Technology under the HKUST-ICERM visiting fellow program, to which he is grateful for providing the very stimulating research environment and support. T. Jin would like to thank Roman Shvydkoy for inspiring discussions. The authors thank the referees for their helpful comments and Luis Silvestre for pointing out an omission in the original manuscript.

References

[Bao 2002] J. Bao, "Fully nonlinear elliptic equations on general domains", *Canad. J. Math.* **54**:6 (2002), 1121–1141. MR Zbl

[Burch 1978] C. C. Burch, "The Dini condition and regularity of weak solutions of elliptic equations", *J. Differential Equations* **30**:3 (1978), 308–323. MR Zbl

[Caffarelli and Silvestre 2009] L. Caffarelli and L. Silvestre, "Regularity theory for fully nonlinear integro-differential equations", *Comm. Pure Appl. Math.* **62**:5 (2009), 597–638. MR Zbl

[Caffarelli and Silvestre 2011] L. Caffarelli and L. Silvestre, "The Evans–Krylov theorem for nonlocal fully nonlinear equations", *Ann. of Math.* (2) **174**:2 (2011), 1163–1187. MR Zbl

[Campanato 1966] S. Campanato, "Equazioni paraboliche del secondo ordine e spazi $\mathcal{L}^{2,\theta}(\Omega,\delta)$", *Ann. Mat. Pura Appl.* (4) **73** (1966), 55–102. MR Zbl

[Chang-Lara and Kriventsov 2017] H. A. Chang-Lara and D. Kriventsov, "Further time regularity for nonlocal, fully nonlinear parabolic equations", *Comm. Pure Appl. Math.* **70**:5 (2017), 950–977. MR Zbl

[Dong and Kim 2012] H. Dong and D. Kim, "On L_p-estimates for a class of non-local elliptic equations", *J. Funct. Anal.* **262**:3 (2012), 1166–1199. MR Zbl

[Dong and Kim 2013] H. Dong and D. Kim, "Schauder estimates for a class of non-local elliptic equations", *Discrete Contin. Dyn. Syst.* **33**:6 (2013), 2319–2347. MR Zbl

[Dong and Zhang 2016a] H. Dong and H. Zhang, "Dini estimates for nonlocal fully nonlinear elliptic equations", preprint, 2016. To appear in *Ann. Inst. H. Poincaré Anal. Non Linéaire*. arXiv

[Dong and Zhang 2016b] H. Dong and H. Zhang, "On Schauder estimates for a class of nonlocal fully nonlinear parabolic equations", preprint, 2016. To appear in *Calc. Var. Partial Differential Equations*. arXiv

[Duzaar and Gastel 2002] F. Duzaar and A. Gastel, "Nonlinear elliptic systems with Dini continuous coefficients", *Arch. Math.* (*Basel*) **78**:1 (2002), 58–73. MR Zbl

[Imbert et al. 2016] C. Imbert, T. Jin, and R. Shvydkoy, "Schauder estimates for an integro-differential equation with applications to a nonlocal Burgers equation", preprint, 2016. To appear in *Ann. Fac. Sci. Toulouse Math*. arXiv

[Jin and Xiong 2015] T. Jin and J. Xiong, "Schauder estimates for solutions of linear parabolic integro-differential equations", *Discrete Contin. Dyn. Syst.* **35**:12 (2015), 5977–5998. MR Zbl

[Jin and Xiong 2016] T. Jin and J. Xiong, "Schauder estimates for nonlocal fully nonlinear equations", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **33**:5 (2016), 1375–1407. MR Zbl

[Kim and Lee 2013] Y.-C. Kim and K.-A. Lee, "Regularity results for fully nonlinear parabolic integro-differential operators", *Math. Ann.* **357**:4 (2013), 1541–1576. MR Zbl

[Knerr 1980/81] B. F. Knerr, "Parabolic interior Schauder estimates by the maximum principle", *Arch. Rational Mech. Anal.* **75**:1 (1980/81), 51–58. MR Zbl

[Komatsu 1984] T. Komatsu, "Pseudodifferential operators and Markov processes", *J. Math. Soc. Japan* **36**:3 (1984), 387–418. MR Zbl

[Kovats 1997] J. Kovats, "Fully nonlinear elliptic equations and the Dini condition", *Comm. Partial Differential Equations* **22**:11-12 (1997), 1911–1927. MR Zbl

[Krylov 1996] N. V. Krylov, *Lectures on elliptic and parabolic equations in Hölder spaces*, Graduate Studies in Mathematics **12**, American Mathematical Society, Providence, RI, 1996. MR Zbl

[Lara and Dávila 2014] H. C. Lara and G. Dávila, "Regularity for solutions of non local parabolic equations", *Calc. Var. Partial Differential Equations* **49**:1-2 (2014), 139–172. MR Zbl

[Li 2017] Y. Li, "On the C^1 regularity of solutions to divergence form elliptic systems with Dini-continuous coefficients", *Chin. Ann. Math. Ser. B* **38**:2 (2017), 489–496. MR Zbl

[Lieberman 1987] G. M. Lieberman, "Hölder continuity of the gradient of solutions of uniformly parabolic equations with conormal boundary conditions", *Ann. Mat. Pura Appl.* (4) **148** (1987), 77–99. MR Zbl

[Lieberman 1992] G. M. Lieberman, "Intermediate Schauder theory for second order parabolic equations, IV: Time irregularity and regularity", *Differential Integral Equations* **5**:6 (1992), 1219–1236. MR Zbl

[Maz'ya and McOwen 2011] V. Maz'ya and R. McOwen, "Differentiability of solutions to second-order elliptic equations via dynamical systems", *J. Differential Equations* **250**:2 (2011), 1137–1168. MR Zbl

[Mikulevičius and Pragarauskas 1992] R. Mikulevičius and H. Pragarauskas, "On the Cauchy problem for certain integro-differential operators in Sobolev and Hölder spaces", *Liet. Mat. Rink.* **32**:2 (1992), 299–331. MR Zbl

[Mikulevičius and Pragarauskas 2014] R. Mikulevičius and H. Pragarauskas, "On the Cauchy problem for integro-differential operators in Hölder classes and the uniqueness of the martingale problem", *Potential Anal.* **40**:4 (2014), 539–563. MR Zbl

[Mou 2016] C. Mou, "Interior regularity for nonlocal fully nonlinear equations with Dini continuous terms", *J. Differential Equations* **260**:11 (2016), 7892–7922. MR Zbl

[Nirenberg 1954] L. Nirenberg, "On a generalization of quasi-conformal mappings and its application to elliptic partial differential equations", pp. 95–100 in *Contributions to the theory of partial differential equations*, Annals of Mathematics Studies **33**, Princeton University Press, 1954. MR Zbl

[Safonov 1988] M. V. Safonov, "Classical solution of second-order nonlinear elliptic equations", *Izv. Akad. Nauk SSSR Ser. Mat.* **52**:6 (1988), 1272–1287. In Russian; translated in *Math. USSR Izv.* **33**:3 (1989), 597–612. MR Zbl

[Schwab and Silvestre 2016] R. W. Schwab and L. Silvestre, "Regularity for parabolic integro-differential equations with very irregular kernels", *Anal. PDE* **9**:3 (2016), 727–772. MR Zbl

[Serra 2015] J. Serra, "$C^{\sigma+\alpha}$ regularity for concave nonlocal fully nonlinear elliptic equations with rough kernels", *Calc. Var. Partial Differential Equations* **54**:4 (2015), 3571–3601. MR Zbl

[Sperner 1981] E. Sperner, Jr., "Schauder's existence theorem for α-Dini continuous data", *Ark. Mat.* **19**:2 (1981), 193–216. MR Zbl

[Wang 2006] X.-J. Wang, "Schauder estimates for elliptic and parabolic equations", *Chinese Ann. Math. Ser. B* **27**:6 (2006), 637–642. MR Zbl

Received 30 Sep 2017. Revised 14 Jan 2018. Accepted 14 Feb 2018.

HONGJIE DONG: hongjie_dong@brown.edu
Division of Applied Mathematics, Brown University, Providence, RI, United States

TIANLING JIN: tianlingjin@ust.hk
Department of Mathematics, The Hong Kong University of Science and Technology, Kowloon, Hong Kong

HONG ZHANG: hong_zhang@alumni.brown.edu
Division of Applied Mathematics, Brown University, Providence, RI, United States

ANALYSIS AND PDE
Vol. 11, No. 6, 2018
dx.doi.org/10.2140/apde.2018.11.1535

msp

SQUARE FUNCTION ESTIMATES FOR THE BOCHNER–RIESZ MEANS

SANGHYUK LEE

We consider the square-function (known as Stein's square function) estimate associated with the Bochner–Riesz means. The previously known range of the sharp estimate is improved. Our results are based on vector-valued extensions of Bennett, Carbery and Tao's multilinear (adjoint) restriction estimate and an adaptation of an induction argument due to Bourgain and Guth. Unlike the previous work by Bourgain and Guth on L^p boundedness of the Bochner–Riesz means in which oscillatory operators associated to the kernel were studied, we take more direct approach by working on the Fourier transform side. This enables us to obtain the correct order of smoothing, which is essential for obtaining the sharp estimates for the square functions.

1. Introduction

We consider the Bochner–Riesz mean of order α, which is defined by

$$\widehat{\mathcal{R}_t^\alpha f}(\xi) = \left(1 - \frac{|\xi|^2}{t^2}\right)_+^\alpha \hat{f}(\xi), \quad t > 0,\ \xi \in \mathbb{R}^d,\ d \geq 2.$$

Let $1 \leq p \leq \infty$. The Bochner–Riesz conjecture is that the estimate

$$\|\mathcal{R}_t^\alpha f\|_p \leq C\|f\|_p \tag{1}$$

holds (except for $p = 2$) if and only if

$$\alpha > \alpha(p) = \max\left(d\left|\frac{1}{2} - \frac{1}{p}\right| - \frac{1}{2}, 0\right). \tag{2}$$

The Bochner–Riesz mean, which is a kind of summability method, has been studied in order to understand convergence properties of Fourier series and integrals. In fact, for $1 \leq p < \infty$, L^p boundedness of $\mathcal{R}_t^\alpha$ implies $\mathcal{R}_t^\alpha f \to f$ in L^p as $t \to \infty$. The necessary condition (2) has been known for a long time [Fefferman 1971; Stein 1993, p. 389].

When $d = 2$, the conjecture was verified by Carleson and Sjölin [1972]; also see [Fefferman 1971]. In higher dimensions $d \geq 3$, the problem is still open and partial results are known. The conjecture was shown to be true for $\max(p, p') \geq 2(d+1)/(d-1)$ by an argument due to Stein [Fefferman 1970], also see [Stein 1993, Chapter 9], and the sharp $L^2 \to L^{\frac{2(d+1)}{d-1}}$ restriction estimate (the Stein–Tomas theorem) for the sphere [Tomas 1975; Stein 1986]. It was Bourgain [1991a; 1991b] who first made progress beyond this result when $d = 3$. Since then, subsequent progress has been tied to that of the restriction problem. Bilinear or multilinear generalizations under transversality assumptions have turned out to be the most effective and

MSC2010: primary 42B15; secondary 35B65.
Keywords: square function, Bochner–Riesz means.

fruitful tools. These results have propelled progress in this area and there is a large body of literature on restriction estimates and related problems. See [Tao et al. 1998; Tao and Vargas 2000a; Wolff 2001; Tao 2003; Lee and Vargas 2008; 2010; Lee 2004; 2006a; 2006b] for bilinear restriction estimates and related results, [Bennett et al. 2006; 2015; Bourgain and Guth 2011; Lee and Vargas 2012; Bourgain 2013; Temur 2014; Bourgain and Demeter 2015; Bennett 2014; Bejenaru 2016; 2017; Ramos 2016] for multilinear restriction estimates and their applications, and [Guth 2016a; 2016b; Shayya 2017; Du et al. 2017; Zhang 2015; Ou and Wang 2017] for the most recent developments related to the polynomial partitioning method.

Concerning improved L^p boundedness of the Bochner–Riesz means in higher dimensions, the sharp L^p bounds for the Bochner–Riesz operator on the range $\max(p, p') \ge 2(d+2)/d$ were established by the author [Lee 2004], making use of the sharp bilinear restriction estimate due to Tao [2003]. When $d \ge 5$, further progress was recently made by Bourgain and Guth [2011]. They improved the range of the sharp (linear) estimates for the oscillatory integral operators of Carleson–Sjölin-type with phases that additionally satisfy an elliptic condition (see [Stein 1986; Bourgain 1991c; Lee 2006a] for earlier results) by using the multilinear estimates for oscillatory integral operators due to Bennett, Carbery and Tao [Bennett et al. 2006] and a factorization theorem. Also see [Carleson and Sjölin 1972; Hörmander 1973; Stein 1986; 1993, Chapter 11] for the relationship between the Bochner–Riesz problem and the oscillatory integral operators of Carleson–Sjölin-type.

The following is currently the best known result for the sharp L^p boundedness of the Bochner–Riesz operator.

Theorem 1.1 [Carleson and Sjölin 1972; Lee 2004; Bourgain and Guth 2011]. *Let $d \ge 2$, $p \in [1, \infty]$, and $p_\circ$ be defined by*

$$p_\circ = p_\circ(d) = 2 + \frac{12}{4d-3-k} \quad \text{if } d \equiv k \pmod 3,\ k = -1, 0, 1. \tag{3}$$

If $\max(p, p') \ge p_\circ$, then (1) holds for $\alpha > \alpha(p)$.[†]

There are also results concerning the endpoint estimates at the critical exponent $\alpha = \alpha(p)$; for example, see [Christ 1987; 1988; Seeger 1996; Tao 1996]. It was shown by Tao [1998] that the sharp L^p bounds of $\mathcal{R}_t^\alpha$ for $1 < p < p_\circ < 2d/(d-1)$ imply the weak-type bounds of $\mathcal{R}_t^{\alpha(p)}$ for $1 < p < p_\circ$. We refer interested readers to [Lee et al. 2014] for variants and related problems.

Square function estimate. We now consider the square function $\mathcal{G}^\alpha f$ which is defined by

$$\mathcal{G}^\alpha f(x) = \left(\int_0^\infty \left| \frac{\partial}{\partial t} \mathcal{R}_t^\alpha f(x) \right|^2 t\, dt \right)^{\frac{1}{2}}.$$

It was introduced by Stein [1958] to study the almost-everywhere summability of Fourier series. Due to the derivative in t, the square function behaves as if it is a multiplier of order $\alpha - 1$ and the derivative $\partial/\partial t$ makes an L^p estimate possible by mitigating bad behavior near the origin. In this paper we are concerned

[†]For the sharp bound for $\max(p, p') \ge p_*$, we have the following relationships: bilinear, $p_* = 2 + 4/d$; multilinear, $p_* = 3 + 3/d + O(d^{-2})$; conjecture, $p_* = 2 + 2/d + O(d^{-2})$.

with the estimate

$$\|\mathcal{G}^\alpha f\|_p \le C\|f\|_p. \tag{4}$$

The L^p estimate for the square function has various consequences and applications. First of all, it is related to smoothing estimates for solutions to dispersive equations associated to radial symbols such as wave and Schrödinger operators. See [Lee et al. 2012; 2013] for the details; see also Remark 3.3. The sharp square-function estimate implies the sharp maximal bounds for Bochner–Riesz means, which will be discussed below in connection to pointwise convergence. It also gives L^p and maximal L^p boundedness of general radial Fourier multipliers, especially the sharp L^p boundedness result of Hörmander–Mikhlin-type; see Corollary 1.3 below and [Carbery et al. 1984; Carbery 1985; Lee et al. 2012].

For $1 < p \le 2$, the inequality (4) is well understood. In this range of p, we know $\mathcal{G}^\alpha$ is bounded on L^p if and only if $\alpha > d\big(\frac{1}{p} - \frac{1}{2}\big) + \frac{1}{2}$; see [Sunouchi 1967; Lee et al. 2014]. Sufficiency can be shown by using vector-valued Calderón–Zygmund theory. In contrast with the case $1 < p \le 2$, if $p > 2$, due to a smoothing effect resulting from averaging in time, the problem has more interesting features and may be considered as a vector-valued extension of the Bochner–Riesz conjecture in that its sharp L^p bound also implies that of the Bochner–Riesz operator. The condition $\alpha > \max\big\{\frac{1}{2}, d\big(\frac{1}{2} - \frac{1}{p}\big)\big\}$ is known to be necessary for (4), see, for example [Lee et al. 2014], and it is natural to conjecture that this is also sufficient for $p > 2$. This conjecture in two dimensions was proven by Carbery [1983], and in higher dimensions $d \ge 3$, sharp estimates for $p > 2(d+1)/(d-1)$ were obtained by Christ [1985] and Seeger [1986] and it was later improved to the range of $p \ge 2(d+2)/d$ by the author, Rogers, and Seeger [Lee et al. 2012]. There are also endpoint estimates at the critical exponent $\alpha = d/2 - d/p$ and weaker $L^{p,2} \to L^p$ endpoint estimates were obtained in [Lee et al. 2014] for $2(d+1)/(d-1) < p < \infty$.

There are two notable approaches for the study of the Bochner–Riesz problem. One, which may be called the *spatial-side approach*, is to prove the sharp estimates for the oscillatory integral operators of Carleson–Sjölin-type [Carleson and Sjölin 1972; Hörmander 1973; Stein 1986]. These operators are natural variable coefficient generalizations of the adjoint restriction operators [Bourgain 1991c; Lee 2006a; Wisewell 2005] for hypersurfaces with nonvanishing Gaussian curvature such as spheres, paraboloids, and hyperboloids. The other, which we may call the *frequency-side approach*, is more related to the Fourier transform side, based on a suitable decomposition in the frequency side and orthogonality between the decomposed pieces [Fefferman 1973; Carbery 1983; Christ 1985; 1987; Seeger 1996; Tao 1998; Lee 2004]. As has been demonstrated in related works, the latter approach makes it possible to carry out finer analysis and to obtain refined results such as the sharp maximal bounds, square-function estimates, and various endpoint estimates.

The recently improved bound for the Bochner–Riesz operator in [Bourgain and Guth 2011] was obtained from the sharp estimate for the oscillatory integral operators of Carleson–Sjölin-type with an additional elliptic assumption. However, this approach doesn't seem appropriate for the study of the square function. In particular, there is an obvious difficulty when one tries to make use of the disjointness of the singularity of the Fourier transform of $\mathcal{R}_t^\alpha f$ which occurs as t varies; for example, see (76). This is where the extra smoothing of order $\frac{1}{2}$ for the square-function estimate comes in, which is most important for the sharp estimates for $\mathcal{G}^\alpha f$ [Carbery 1983; Christ 1985; Lee 2004; Lee et al. 2012]. This kind of smoothing can be seen clearly in the Fourier transforms of Bochner–Riesz means but is not easy to exploit in the oscillatory kernel side. As is already known [Bourgain 1991c; Wisewell 2005; Lee 2006a; Bourgain and Guth 2011], the behavior of the oscillatory integral operators of Carleson–Sjölin-type are more subtle and generally considered to be difficult to analyze when compared to their constant-coefficient counterparts, the adjoint restriction operators. So, we take the frequency-side approach, in which we directly handle the associated multiplier by working in the frequency space rather than dealing with the oscillatory integral operator given by the kernel of the Bochner–Riesz operator.

In this paper, we obtain the sharp square-function estimates which are new when $d \ge 9$.

Theorem 1.2. *Let us define $p_s = p_s(d)$ by*

$$p_s = 2 + \frac{12}{4d-6-k}, \quad d \equiv k \pmod 3,\ k = 0, 1, 2. \tag{5}$$

Then, if $p \ge \min(p_s, 2(d+2)/d)$ and $\alpha > d/2 - d/p$, the estimate (4) holds.

The range here does not match that of Theorem 1.2. This results from an additional time average which increases the number of decomposed frequency pieces. (See Section 3F.)

Maximal estimate and pointwise convergence. A straightforward consequence of the estimate (4) is the maximal estimate

$$\Big\| \sup_{t>0} |\mathcal{R}_t^\alpha f| \Big\|_p \le C \|f\|_p \tag{6}$$

for $\alpha > \alpha(p)$, which follows from Sobolev imbedding and (4). Hence, Theorem 1.2 yields the sharp maximal bounds for $p \ge p_s(d)$. When $p \ge 2$, it has been conjectured that (6) holds as long as (2) is satisfied. The sharp L^2 bound goes back to Stein [1958]. The conjecture in $\mathbb{R}^2$ and the sharp bounds for $p > 2(d+1)/(d-1)$, $d \ge 3$, were verified by square-function estimates [Christ 1985; Seeger 1986]. The bounds were later improved to the range $p > 2(d+2)/d$ by the author [Lee 2004] using an $L^p \to L^p(L_t^4)$ estimate. The inequality (6) has been studied in connection with almost-everywhere convergence of Bochner–Riesz means. However, the problem of showing $\mathcal{R}_t^\alpha f \to f$ a.e. for $f \in L^p$, $p > 2$, $\alpha > \alpha(p)$, was settled by Carbery, Rubio de Francia and Vega [Carbery et al. 1988]. Their result relies on weighted L^2 estimates. There are also results on pointwise convergence at the critical $\alpha = \alpha(p)$. See [Lee and Seeger 2015; Annoni 2017]. When $1 < p < 2$, by Stein's maximal theorem almost-everywhere convergence of $\mathcal{R}_t^\alpha f \to f$ for $f \in L^p$ is equivalent to the $L^p \to L^{p,\infty}$ estimate for the maximal operator and it was shown by Tao [1998] that the stronger condition $\alpha \ge (2d-1)/(2p) - d/2$ is necessary for (6). Except for $d = 2$ [Tao 2002], little is known beyond the classical result which follows from interpolation between L^2 ($\alpha > 0$) and L^1 ($\alpha > (d-1)/2$) estimates.

Radial multiplier. Let m be a function defined on $\mathbb{R}_+$. Combining the inequality due to Carbery, Gasper and Trebels [Carbery et al. 1984] and Theorem 1.2, we obtain the following L^p boundedness result of Hörmander–Mikhlin-type, which is sharp in that the regularity assumption cannot be improved. A similar result for the maximal function $f \to \sup_{t>0} |\mathcal{F}^{-1}(m(t|\cdot|)\hat{f})|$ is also possible thanks to the inequality due to Carbery [1985].

Corollary 1.3. *Let $d \ge 2$, and φ be a nontrivial smooth function with compact support contained in $(0,\infty)$. If $\min(p_s, 2(d+2)/d) \le \max(p, p') < \infty$ and $\alpha > d\left|\frac{1}{p}-\frac{1}{2}\right|$, then*

$$\left\|\mathcal{F}^{-1}[m(|\cdot|)\hat{f}]\right\|_p \lesssim \sup_{t>0} \|\varphi m(t\,\cdot)\|_{L^2_\alpha(\mathbb{R})} \|f\|_p.$$

About the paper. In Section 2, by working in the frequency side we provide an alternative proof of Theorem 1.1. Although, this doesn't give an improvement over the current range, we include this because it has some new consequences, clarifies several issues which were not clearly presented in [Bourgain and Guth 2011], and provides preparation for Section 3, in which we work in a vector-valued setting. The proof in that paper is sketchy and doesn't look readily accessible. Also the heuristic that a function with Fourier support in a ball of radius σ behaves as if it is constant on balls of radius $\frac{1}{\sigma}$ is now widely accepted and has important role in the induction argument but it doesn't seem justified at high level of rigor. We provide a rigorous argument by making use of Fourier series (see Lemmas 2.13 and 3.14). Another problem of the induction argument is that the primary object (the associated surfaces or phase functions) changes in the course of induction. However, these issues are not properly addressed in literature. We handle this matter by introducing a stronger induction assumption (see Remark 2.4) and carefully handling the stability of various estimates. We also use a different type of multilinear decomposition which is more systematic, easier and more efficient for dealing with multiplier operators (see Section 2E, especially the discussion at the beginning of Section 2E).

Section 3 is very much built on the frequency-side analysis in Section 2, as it may be regarded a vector-valued extension of Section 2. Consequently, the structure of Section 3 is similar to that of Section 2 and some of the arguments commonly work in both sections. In such cases we try to minimize repetition while keeping readability as much as possible. We first obtain vector-valued extensions of multilinear estimates (Propositions 3.6 and 3.10) which serve as basic estimates for the sharp-square function estimate. Then, to derive the linear estimate (Theorem 1.2) we adapt the frequency side approach in Section 2 to the vector-valued setting and prove our main theorem.

Finally, the oscillatory integral approach has its own limits for proving Bochner–Riesz conjecture. As is now well known [Bourgain 1991c; Wisewell 2005; Lee 2006a; Bourgain and Guth 2011], the sharp L^p–L^q estimates for the oscillatory operators of Carleson–Sjölin-type fail for $q < q_\circ$, $q_\circ > 2d/(d-1)$, even under the elliptic condition on the phase [Wisewell 2005; Lee 2006a; Bourgain and Guth 2011]. The Fourier-transform-side approach may help further development in a different direction and thanks to its flexibility may have applications to related problems.

Notation. The following is a list of notation we frequently use in the rest of the paper:

- C, c are constants which depend only on d and may differ at each occurrence.
- For $A, B \ge 0$, we say $A \lesssim B$ if there is a constant C such that $A \le CB$.
- $I = [-1,1]$ and $I^d = [-1,1]^d \subset \mathbb{R}^d$.
- $\tau_h f(x) = f(x-h)$ and $\tau_i f$ denotes $\tau_{h_i} f$ for some $h_i \in \mathbb{R}^d$, $i = 1, \dots, m$.

- We denote by $\mathfrak{q}(a,\ell) \subset \mathbb{R}^d$ the closed cube centered at a with side length 2ℓ, namely, $a + \ell I^d$. If $\mathfrak{q} = \mathfrak{q}(a,\ell)$, denote a, the center of $\mathfrak{q}$, by $\boldsymbol{c}(\mathfrak{q})$.
- For $r > 0$ and a given cube or rectangle Q, we denote by rQ the cube or rectangle which is the r-times dilation of Q from the center of Q.
- Let $\rho \in \mathcal{S}(\mathbb{R}^d)$ be a function with Fourier support in $\mathfrak{q}(0,1)$ and $\rho \geq 1$ on $\mathfrak{q}(0,1)$. And we also set $\rho_{B(z,r)}(x) := \rho((\cdot - z)/r)$.
- For a given set $A \subset \mathbb{R}^d$, we define the set $A + O(\delta)$ by
$$A + O(\delta) := \{x \in \mathbb{R}^d : \operatorname{dist}(x, A) < C\delta\}.$$
- For a given dyadic cube $\mathfrak{q}$ and function f, we define $f_{\mathfrak{q}}$ by $\hat{f}_{\mathfrak{q}} = \chi_{\mathfrak{q}} \hat{f}$.
- Besides $\hat{}$ and $\check{}$, we also denote by $\mathcal{F}(\,\cdot\,)$ and $\mathcal{F}^{-1}(\,\cdot\,)$ the Fourier transform and the inverse Fourier transform, respectively.
- For a smooth function G on I^k, we set $\|G\|_{C^N(I^k)} := \max_{|\alpha|\leq N} \max_{x\in I^k} |\partial^\alpha G(x)|$.

2. Estimates for multiplier operators

In this section we consider the multiplier operators of Bochner–Riesz-type which are associated with elliptic-type surfaces. They are natural generalizations of the Bochner–Riesz operator $\mathcal{R}_1^\alpha$. We prove the sharp L^p boundedness of these of operators and this provides an alternative proof of Theorem 1.1. Basically we adapt the induction argument in [Bourgain and Guth 2011]. However, compared to the (adjoint) restriction counterpart, the induction argument becomes less obvious when we consider it for the Fourier multiplier operator. However, exploiting sharpness of bounds for the frequency-localized operator T_δ, see (9)–(10), we manage to carry out a similar argument. See Section 2F.

From now on we write
$$\xi = (\zeta, \tau) \in \mathbb{R}^{d-1} \times \mathbb{R}.$$
Let ψ be a smooth function defined on I^d and $\chi_\circ$ be a smooth function supported in a small neighborhood of the origin. We consider the multiplier operator $T^\alpha = T^\alpha(\psi)$ which is defined by
$$\mathcal{F}(T^\alpha f)(\xi) = (\tau - \psi(\zeta))_+^\alpha \chi_\circ(\xi) \hat{f}(\xi).$$
By a finite decomposition, rotation and translation and by discarding the harmless smooth multiplier, it is easy to see that the L^p boundedness of $\mathcal{R}_1^\alpha$ is equivalent to that of T^α, which is given by $\psi(\zeta) = 1 - (1 - |\zeta|^2)^{\frac{1}{2}}$. A natural generalization of the Bochner–Riesz problem is as follows: If $\det H\psi \neq 0$ on the support of $\chi_\circ$ (here, $H\psi$ is the Hessian matrix of ψ), we may conjecture that, for $1 \leq p \leq \infty,\ p \neq 2$,
$$\|T^\alpha f\|_p \leq C\|f\|_p \tag{7}$$
if and only if $\alpha > \alpha(p)$. From explicit computation of the kernel of T^α it is easy to see that the condition $\alpha > \alpha(p)$ is necessary for (7). However, in this paper we only work with specific choices of ψ.

2A. ***Elliptic function.*** Let us set

$$\psi_\circ(\zeta) = \tfrac{1}{2}|\zeta|^2.$$

For $0 < \varepsilon_\circ \ll \frac{1}{2}$ and an integer $N \ge 100d$ we denote by $\mathfrak{G}(\varepsilon_\circ, N)$ the collection of smooth functions which is given by

$$\mathfrak{G}(\varepsilon_\circ, N) = \{\psi : \|\psi - \psi_\circ\|_{C^N(I^{d-1})} \le \varepsilon_\circ\}.$$

If $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$ and $a \in \frac{1}{2}I^{d-1}$, then $H\psi(a)$ has eigenvalues $\lambda_1, \dots, \lambda_{d-1}$ close to 1 and we may write $H\psi(a) = P^{-1}DP$ for an orthogonal matrix P, while D is a diagonal matrix with diagonal entries $\lambda_1, \dots, \lambda_{d-1}$. We denote by $\sqrt{H\psi(a)}$ the matrix $P^{-1}D'P$, where D' is the diagonal matrix with diagonal entries $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_{d-1}}$. So, $(\sqrt{H\psi(a)})^2 = H\psi(a)$.

For $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, $a \in \frac{1}{2}I^{d-1}$, and $0 < \varepsilon \le \frac{1}{2}$, we define

$$\psi_a^\varepsilon(\zeta) = \frac{1}{\varepsilon^2}\Big(\psi\big(\varepsilon[\sqrt{H\psi(a)}]^{-1}\zeta + a\big) - \psi(a) - \varepsilon\nabla\psi(a)\cdot[\sqrt{H\psi(a)}]^{-1}\zeta\Big). \tag{8}$$

Since $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, by Taylor's theorem it is easy to see that $\|\psi_a^\varepsilon - \psi_\circ\|_{C^N(I^{d-1})} \le C\varepsilon$ for $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$.[†] Hence we get the following.

Lemma 2.1. *Let $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$ and $a \in \frac{1}{2}I^{d-1}$. Then there is a constant $\kappa = \kappa(\varepsilon_\circ, N)$, independent of a, ψ, such that $\psi_a^\varepsilon \in \mathfrak{G}(\varepsilon_\circ, N)$ provided that $0 < \varepsilon \le \kappa$.*

Remark 2.2. If ψ is smooth and $H\psi(a)$ has $d-1$ positive eigenvalues, after finite decomposition and affine transformations we may assume $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$ for arbitrarily small $\varepsilon_\circ$ and large N. Indeed, for given $\varepsilon > 0$, decomposing the multiplier $(\tau - \psi(\zeta))_+^\alpha \chi_\circ(\xi)$ to multipliers supported in balls of small radius ε/C with some large C, one may assume that $\mathcal{F}f$ is supported in $B((a, \psi(a)), \varepsilon/C)$. Then, the change of variables (12) transforms $\psi \to \psi_a^\varepsilon$ and gives rise to a new multiplier operator $T^\alpha(\psi_a^\varepsilon)$ and, as can be easily seen by a simple change of variables, the operator norm $\|T^\alpha(\psi_a^\varepsilon)\|_{p\to p}$ remains same. (See the proof of Proposition 2.5.) By Lemma 2.1 we see $\psi_a^\varepsilon \in \mathfrak{G}(\varepsilon_\circ, N)$ if ε is small enough.

2B. ***Multiplier operator with localized frequency.*** Let ϕ be a smooth function supported in $2I$. For $\delta > 0$, $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, and f with Fourier transform supported in $\frac{1}{2}I^d$ we define the (frequency-localized) multiplier operator $T_\delta = T_\delta(\psi)$ by

$$\widehat{T_\delta f}(\xi) = \phi\left(\frac{\tau - \psi(\zeta)}{\delta}\right)\hat{f}(\xi). \tag{9}$$

As is well known, the L^p bound for T_δ largely depends on the curvature of the surface $\tau = \psi(\zeta)$. By decomposing the multiplier dyadically away from the singularity $\tau = \psi(\zeta)$, in order to prove (7) for $p > 2d/(d-1)$ and $\alpha > \alpha(p)$, it is enough to show that, for any $\varepsilon > 0$,

$$\|T_\delta f\|_p \le C\delta^{\frac{d}{p} - \frac{d-1}{2} - \varepsilon}\|f\|_p \tag{10}$$

whenever $\hat{f}$ is supported in $\frac{1}{2}I^d$. The following recovers the sharp L^p bound up to the currently best known range in [Bourgain and Guth 2011].

[†] Indeed, since $|\partial^\alpha(\psi_a^\varepsilon - \psi_\circ)| \lesssim \varepsilon^{|\alpha|-2}$ for any multi-index α, we need only to show $|\partial^\alpha(\psi_a^\varepsilon - \psi_\circ)| \lesssim \varepsilon$ for $|\alpha| = 0, 1, 2$. This follows by Taylor's theorem since $N \ge 100d$.

Proposition 2.3. *Let $\varepsilon > 0$. If $p \ge p_\circ(d)$ and $\varepsilon_\circ$ is small enough, there is an $N = N(\varepsilon)$ such that (10) holds uniformly provided that $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$ and* supp $\hat f \subset \frac{1}{2} I^d$.

It is possible to remove the loss of $\delta^{-\varepsilon}$ in (10) by the ε-removal argument in [Tao 1998, Section 4].

Induction quantity. To control the L^p norm of T_δ, for $0 < \delta$, we define $A(\delta) = A_p(\delta)$ by

$$A(\delta) = \sup\big\{\|T_\delta(\psi) f\|_{L^p} : \psi \in \mathfrak{G}(\varepsilon_\circ, N),\ \|f\|_p \le 1,\ \operatorname{supp} \hat f \subset \tfrac{1}{2} I^d\big\}.$$

Remark 2.4. Though the induction argument in [Bourgain and Guth 2011] heavily relies on the stability of the multilinear estimates, such issue doesn't seem properly addressed. In particular, after (multiscale) decomposition and rescaling, the associated phase functions (or surfaces) are no longer fixed-phase functions (or surfaces).† This requires the induction quantity defined over a class of phase functions or surfaces. This leads us to consider $A(\delta)$.

From the estimate for the kernel of T_δ (see Lemma 2.9), it is easy to see that $A(\delta) \le C$ uniformly in $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$ if $\delta \ge 1$ and $A(\delta) \le C\delta^{-\frac{d-1}{2}}$ if $0 < \delta \le 1$, because the L^1-norm of the kernel is uniformly $O(\delta^{-\frac{d-1}{2}})$. To prove Proposition 2.3, we need to show $A(\delta) \lesssim \delta^{\frac{d}{p} - \frac{d-1}{2} - \varepsilon}$ for any $\varepsilon > 0$. However, due to the lack of monotonicity $A(\delta)$ is not suitable for closing the induction. So, we need to modify $A(\delta)$. For $\beta, \delta > 0$, we define

$$\mathcal{A}^\beta(\delta) = \mathcal{A}^\beta_p(\delta) := \sup_{\delta < s \le 1} s^{\frac{d-1}{2} - \frac{d}{p} + \beta} A_p(s).$$

Hence, Proposition 2.3 follows if we show $\mathcal{A}^\beta(\delta) \le C$ for any $\beta > 0$.

The following lemma makes precise the heuristic that the bound of T_δ improves if it acts on functions with Fourier transforms supported in a smaller set. However, this becomes less obvious for the multiplier operator when it is compared to the restriction (adjoint) operator; see [Bourgain and Guth 2011]. This type of improvement is basically due to the parabolic rescaling structure of the operator, and generally appears in $L^p - L^q$ estimates for p, q satisfying $(d+1)/q < (d-1)(1-1/p)$, $p \le q$, which are not invariant under the parabolic rescaling. The following is important for the induction argument to work.

Proposition 2.5. *Let $0 < \delta \ll 1$, $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, and $(a, \mu) \in \mathbb{R}^{d-1} \times \mathbb{R}$. Suppose that* supp $\hat f \subset \mathfrak{q}((a,\mu), \varepsilon) \subset \frac{1}{2} I^d$, *$0 < \varepsilon < \frac{1}{2}$ and $\delta \le (10)^{-2}\varepsilon^2$. Then, there is a $\kappa = \kappa(\varepsilon_\circ, N)$ such that for $0 < \varepsilon \le \kappa$*

$$\|T_\delta f\|_p \le C A(\varepsilon^{-2}\delta) \|f\|_p \tag{11}$$

holds with C independent of ψ and ε.

Proof. Decomposing $\mathfrak{q}(a, \varepsilon)$ into as many as $O(d^d)$, we may assume $\hat f$ is supported in $\mathfrak{q}((a,\mu), \varepsilon/(10d))$. Since $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$ and supp $\hat f \subset \mathfrak{q}((a,\mu), \varepsilon/(10d))$, by Taylor's theorem we have $\phi((\tau - \psi(\zeta))/\delta)\hat f(\xi)$, which is supported in the set

$$\Big\{(\xi, \tau) : |\tau - \psi(a) - \nabla\psi(a)\cdot(\zeta - a)| \le \frac{(1+\varepsilon_0)\varepsilon^2}{2 \times 10^2}\Big\}.$$

†It is only true for the paraboloid.

Hence, we may write

$$\phi\Big(\frac{\tau-\psi(\zeta)}{\delta}\Big)\hat f(\xi)=\phi\Big(\frac{\tau-\psi(\zeta)}{\delta}\Big)\tilde\chi\Big(\frac{\tau-\psi(a)-\nabla\psi(a)\cdot(\zeta-a)}{\varepsilon^2}\Big)\hat f(\xi),$$

where $\tilde\chi$ is a smooth function supported in $\frac{1}{2}I$ such that $\tilde\chi=1$ on $\frac{1}{4}I$. Let us set $M=(\sqrt{H\psi(a)})^{-1}$ and make the change of variables in the frequency domain

$$(\zeta,\tau)\to L(\zeta,\tau)=(\varepsilon M\zeta+a,\varepsilon^2\tau+\psi(a)+\varepsilon\nabla\psi(a)\cdot M\zeta). \tag{12}$$

Then it follows that

$$\mathcal{F}(T_\delta(\psi)f)(L\xi)=\phi\Big(\frac{\tau-\psi_a^\varepsilon(\zeta)}{\varepsilon^{-2}\delta}\Big)\tilde\chi(\tau)\hat f(L\xi).$$

Since L is an invertible affine transformation, it is easy to see

$$\|\mathcal{F}^{-1}(\hat g(L\cdot))\|_p=\varepsilon^{(d+1)(\frac{1}{p}-1)}\|g\|_p$$

for any g. We also note that $\mathrm{supp}(\tilde\chi(\tau)\hat f(L\cdot))\subset\frac{1}{2}I^d$ and by Lemma 2.1 there exists a $\kappa>0$ such that $\psi_a^\varepsilon\in\mathfrak{G}(\varepsilon_\circ,N)$ if $0<\varepsilon\le\kappa$ whenever $\psi\in\mathfrak{G}(\varepsilon_\circ,N)$. So, by the definition of $A(\delta)$ it follows that, for $0<\varepsilon\le\kappa$,

$$\begin{aligned}\|T_\delta(\psi)f\|_p&=\varepsilon^{(d+1)(1-\frac{1}{p})}\Big\|\mathcal{F}^{-1}\Big(\phi\Big(\frac{\tau-\psi_a^\varepsilon(\zeta)}{\varepsilon^{-2}\delta}\Big)\tilde\chi(\tau)\hat f(L\xi)\Big)\Big\|_p\\&\le\varepsilon^{(d+1)(1-\frac{1}{p})}A(\varepsilon^{-2}\delta)\|\mathcal{F}^{-1}(\tilde\chi(\tau)\hat f(L\xi))\|_p\le CA(\varepsilon^{-2}\delta)\|f\|_p.\end{aligned}$$

For the last inequality we also use the trivial bound $\|\mathcal{F}^{-1}(\tilde\chi(\tau)\hat g)\|_p\le C\|g\|_p$ for any $1\le p\le\infty$. The inequality is valid for any $\psi\in\mathfrak{G}(\varepsilon_\circ,N)$. This gives the desired bound. □

We will need the following estimate which is easy to show by making use of Rubio de Francia's one-dimensional inequality [Rubio de Francia 1985].

Lemma 2.6. *Let $\{\mathfrak{q}\}$ be a collection of (distinct) dyadic cubes of the same side length σ. Let $2\le p<\infty$. Then, there is a constant C, independent of the collection $\{\mathfrak{q}\}$, such that*

$$\Big(\sum_{\mathfrak{q}}\|\mathcal{F}^{-1}(\hat f\chi_{\mathfrak{q}})\|_p^p\Big)^{\frac{1}{p}}\le C\|f\|_p.$$

2C. *Multilinear estimates.* In this subsection we consider various multilinear estimates which are basically consequences of multilinear restriction and Kakeya estimates in [Bennett et al. 2006].

For $\psi\in\mathfrak{G}(\varepsilon_\circ,N)$ let us set

$$\Gamma=\Gamma(\psi)=\big\{(\zeta,\psi(\zeta)):\zeta\in\tfrac{1}{2}I^{d-1}\big\}.$$

Let $2\le k\le d$, and let $U_1,U_2,\dots,U_k$ be compact subsets of I^{d-1}. For $i=1,\dots,k$, and $\lambda>0$, set

$$\Gamma_i=\{(\zeta,\psi(\zeta)):\zeta\in U_i\},\quad \Gamma_i(\lambda)=\Gamma_i+O(\lambda).$$

For $\xi=(\zeta,\psi(\zeta))\in\Gamma(\psi)$, let $\mathrm{N}(\xi)$ be the upward unit normal vector at $(\zeta,\psi(\zeta))$.

For $v_1, \dots, v_k \in \mathbb{R}^d$, denote by $\mathrm{Vol}(v_1, \dots, v_k)$ the k-dimensional volume of the parallelepiped given by $\{s_1 v_1 + \cdots + s_k v_k : s_i \in [0,1],\ 1 \le i \le k\}$. Transversality among the surfaces $\Gamma_1, \dots, \Gamma_k$ is important for the multilinear estimates. The degree of transversality is quantitatively stated as follows:

$$\mathrm{Vol}(\mathrm{N}(\xi_1), \mathrm{N}(\xi_2), \dots, \mathrm{N}(\xi_k)) \ge \sigma \tag{13}$$

for some $\sigma > 0$ whenever $\xi_i \in \Gamma_i$, $i = 1, \dots, k$. Since $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, we know $\nabla\psi$ is a diffeomorphism which is close to the identity map. The condition (13) may be replaced by a simpler one: $\mathrm{Vol}(\zeta_1, \zeta_2, \dots, \zeta_k|) \gtrsim \sigma$ whenever $\zeta_i \in U_i$, $i = 1, \dots, k$. The following is due to Bennett, Carbery and Tao [Bennett et al. 2006].

Theorem 2.7. *Let $0 < \delta \ll \sigma \ll 1$ and $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$. Suppose that $\Gamma_1, \dots, \Gamma_k$ are given as in the above and (13) is satisfied whenever $\xi_i \in \Gamma_i$, $i = 1, \dots, k$, and suppose that $\widehat{F_i} \subset \Gamma_i(\delta)$, $i = 1, \dots, k$. Then, if $p \ge 2k/(k-1)$ and $\varepsilon_\circ$ is sufficiently small, for $\varepsilon > 0$ there are constants $N = N(\varepsilon)$ such that, for $x \in \mathbb{R}^d$,*

$$\Big\| \prod_{i=1}^k F_i \Big\|_{L^{p/k}(B(x,\delta^{-1}))} \le C\sigma^{-C_\varepsilon}\delta^{-\varepsilon} \prod_{i=1}^k \delta^{\frac12} \|F_i\|_2$$

holds with C, C_ε, independent of ψ.

Besides the stability issue, this estimate is essentially the same as the multilinear restriction estimate in [Bennett et al. 2006]; see Theorem 1.16 of that paper, as well as Lemma 2.2, for the case $k = d$ and see Section 5 for the case of lower linearity $2 \le k < d$. Though we are considering only the surfaces which are the graphs of $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, the theorem remains true for surfaces even with vanishing curvature as long as the transversality condition is satisfied. Uniformity of the estimate follows from the fact that the multilinear Kakeya and restriction estimates are stable under perturbation of the associated surfaces. The estimate is conjectured to be true without $\delta^{-\varepsilon}$ loss (this is equivalent with the endpoint k-linear restriction estimate) but it remains open when $k \ge 3$ even though the corresponding endpoint case for the multilinear Kakeya estimate was obtained by Guth [2010].

Remark 2.8. The proof of Theorem 2.7 is based on the multilinear Kakeya estimate and an induction-on-scale argument, which involves iteration of the induction assumption to reduce the exponent of δ^{-1}. Such an improvement of exponent is possible at the expense of extra loss of bounds in terms of σ^{-c}. By following the argument in [Bennett et al. 2006], one can easily see that one may take $C_\varepsilon \lesssim C \log \frac{1}{\varepsilon}$; see the paragraph below (20). Hence, the bound becomes less efficient when σ gets as small as δ^c for some $c > 0$. In $\mathbb{R}^3$ the sharp bound depending on σ was recently obtained by Ramos [2016]. However, the argument of Bourgain and Guth avoids such problems by keeping the Fourier supports of functions largely separated while being decomposed. In contrast with the conventional approach in which functions are usually decomposed into finer frequency pieces, this was achieved by decomposing the input functions into those of relatively large frequency supports.

Lemma 2.9. *Let $\varphi \in C_c^\infty(2I)$ and $\eta \in C_c^\infty(I^d)$, where $\frac12 \le \eta \le 2$. Let $0 < \delta \ll \sigma \le 1$. Set*

$$K_\delta = \mathcal{F}^{-1}\bigg(\varphi\Big(\frac{\eta(\xi)(\tau - \psi(\zeta))}{C\delta}\Big)\tilde\chi(\xi)\bigg),$$

and $\mathfrak{K}_M(x) = (1+\delta|x|)^{-M}$. *Suppose* $\tilde\chi$ *is supported in a cube of side length* $C\sigma$ *and* $|\partial_\xi^\alpha \tilde\chi| \lesssim \sigma^{-|\alpha|}$ *for any* α. *Then, for any* M, *there is an* $N = N(M)$ *such that*

$$|K_\delta(x)| \le C\delta\sigma^{d-1}\mathfrak{K}_M(x) \tag{14}$$

with C *depending only on* $\|\psi\|_{C^N(I^{d-1})}$.

Proof. Changing variables $\tau \to \delta\tau + \psi(\zeta)$, we write

$$K_\delta(x) = (2\pi)^{-d}\delta \int e^{i\delta\tau x_d} \int e^{i(x'\cdot\zeta + x_d\psi(\zeta))}\tilde\varphi(\xi)\, d\zeta\, d\tau,$$

where

$$\tilde\varphi(\xi) = \varphi\Big(\frac{\eta(\zeta, \delta\tau+\psi(\zeta))\tau}{C}\Big)\tilde\chi(\zeta, \delta\tau + \psi(\zeta)).$$

We note that $|\partial_\zeta^\alpha\tilde\varphi| \lesssim \sigma^{-|\alpha|}(\|\psi\|_{C^{|\alpha|}} + \|\eta\|_{C^{|\alpha|}})$. Then, if $|x'|/100 \ge |x_d|$, by integration by parts it follows that

$$\Big|\int e^{i(x'\cdot\zeta + x_d\psi(\zeta))}\tilde\phi(\xi)\, d\zeta\Big| \le C\sigma^{d-1}\big(\|\psi\|_{C^M(I^{d-1})} + \|\eta\|_{C^M(I^d)}\big)(1+\sigma|x'|)^{-M}.$$

Note that $\tilde\phi(\xi) = 0$ if $|\tau| \ge 5C$ since $\frac12 \le \eta \le 1$. This gives the desired inequality (14) by taking integration in τ since $\delta \ll \sigma$. On the other hand, if $|x'|/100 < |x_d|$, we integrate in τ first. Since $|\partial_\tau^l\tilde\varphi| \lesssim (\|\psi\|_{C^l} + \|\eta\|_{C^l})$, by integration by parts again we have

$$\Big|\int e^{i\delta\tau x_d}\tilde\phi(\xi)\, d\tau\Big| \le C\big(\|\psi\|_{C^M(I^{d-1})} + \|\eta\|_{C^M(I^d)}\big)(1+|\delta x_d|)^{-M}.$$

This and taking integration in ζ yield (14). □

From Theorem 2.7 and Lemma 2.9 we can obtain the sharp multilinear L^p estimate for T_δ under the transversality condition without localizing the multilinear operator on a ball of radius $\frac{1}{\delta}$. In fact, since $T_\delta f = K_\delta * f$ and the kernel K_δ (from Lemma 2.9) is rapidly decaying outside of $B(0, C/\delta)$, one may handle f as if it were supported in a ball B of radius $\delta^{-1-\varepsilon}$. This type of localization and Hölder's inequality make it possible to lift the L^2 estimate to that of L^p, $p \ge 2$, with sharp bound. Such an idea of deducing L^p estimates from L^2 ones goes back to Stein [1993, pp. 442–443], see also [Fefferman 1970; 1973], and in [Lee 2004; Lee et al. 2012] a similar idea was used to make use of the L^2 bilinear restriction estimate. The same argument also works with the multilinear estimates with a little modification. We make it precise in what follows.

Proposition 2.10. *Let* $0<\delta \ll \sigma \ll \tilde\sigma \ll 1$ *and* $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, *and let* $Q_1, \dots, Q_k \in \frac12 I^d$ *be dyadic cubes of side length* $\tilde\sigma$. *Suppose that* (13) *is satisfied whenever* $\xi_i \in \Gamma \cap Q_i$, $i = 1, \dots, k$, *and* supp $\hat f_i \subset Q_i$, $i = 1, \dots, k$. *Then, if* $p \ge 2k/(k-1)$ *and* $\varepsilon_\circ$ *is small enough, for* $\varepsilon > 0$ *there is an* $N = N(\varepsilon)$ *such that*

$$\Big\|\prod_{i=1}^k T_\delta f_i\Big\|_{L^{p/k}(\mathbb{R}^d)} \le C\sigma^{-C_\varepsilon}\delta^{-\varepsilon}\prod_{i=1}^k \delta^{\frac{d}{p}-\frac{d-1}{2}}\|f_i\|_p \tag{15}$$

holds with C, C_ε *independent of* ψ.

Proof. Set $\widetilde{Q}_i = \{\xi : \mathrm{dist}(\xi, Q_i) \le \tilde{c}\sigma\}$, and let $\tilde{\chi}_i$ be a smooth function supported in $\widetilde{Q}_i$ which satisfies $\tilde{\chi}_i = 1$ on Q_i and $|\partial_\xi^\alpha \tilde{\chi}_i| \lesssim \sigma^{-|\alpha|}$. Let us define K_i by

$$\mathcal{F}(K_i)(\xi) = \phi\Big(\frac{\tau - \psi(\zeta)}{\delta}\Big)\tilde{\chi}_i(\xi).$$

Since $\hat{f}_i$ is supported in Q_i, we have $T_\delta f_i = K_i * f_i$.

Let $\{\mathcal{B}\}$ be the collection of boundedly overlapping balls of radius δ^{-1} which cover $\mathbb{R}^d$. For $\varepsilon > 0$ we denote by $\widetilde{\mathcal{B}}$ the balls $B(a, \delta^{-1-\varepsilon})$ if $\mathcal{B} = B(a, \delta^{-1})$. By decomposing $f_i = \chi_{\widetilde{\mathcal{B}}} f_i + \chi_{\widetilde{\mathcal{B}}^c} f_i$, we bound the (p/k)-th power of the left-hand side of (15) by

$$\sum_{\mathcal{B}} \int_{\mathcal{B}} \prod_{i=1}^{k} |T_\delta f_i(x)|^{\frac{p}{k}}\, dx = \sum_{\mathcal{B}} \int_{\mathcal{B}} \prod_{i=1}^{k} |K_i * f_i(x)|^{\frac{p}{k}}\, dx \lesssim I + II,$$

where

$$I = \sum_{\mathcal{B}} \int_{\mathcal{B}} \prod_{i=1}^{k} |K_i * (\chi_{\widetilde{\mathcal{B}}} f_i)(x)|^{\frac{p}{k}}, \quad II = \sum_{\mathcal{B}} \Big(\sum_{\substack{g_i = \chi_{\widetilde{\mathcal{B}}^c} f_i \\ \text{for some } i}} \int_{\mathcal{B}} \prod_{i=1}^{k} |K_\delta * g_i(x)|^{\frac{p}{k}}\, dx \Big).$$

The second sum in II is summation over all possible choices of g_i with $g_i = \chi_{\widetilde{\mathcal{B}}} f_i$ or $\chi_{\widetilde{\mathcal{B}}^c} f_i$, and $g_i = \chi_{\widetilde{\mathcal{B}}^c} f_i$ for some i. So, in $\prod_{i=1}^k K_\delta * g_i(x)$ there is at least one g_i which satisfies $g_i = \chi_{\widetilde{\mathcal{B}}^c} f_i$.

Since $\mathcal{F}(K_i * (\chi_{\widetilde{\mathcal{B}}} f_i)) \subset \Gamma(\delta) \cap \widetilde{Q}_i$, taking a sufficiently small $\tilde{c} > 0$, from continuity it is easy to see that $F_1 = K_1 * (\chi_{\widetilde{\mathcal{B}}} f_1), \dots, F_k = K_k * (\chi_{\widetilde{\mathcal{B}}} f_k)$ satisfy the assumption of Theorem 2.7. So, by Theorem 2.7 and Plancherel's theorem we see

$$I \lesssim \sigma^{-C_\varepsilon} \Big(\frac{1}{\delta}\Big)^{\varepsilon} \sum_{\mathcal{B}} \prod_{i=1}^{k} \delta^{\frac{p}{2k}} \|K_i * (\chi_{\widetilde{\mathcal{B}}} f_i)\|_2^{\frac{p}{k}} \le \sigma^{-C_\varepsilon} \Big(\frac{1}{\delta}\Big)^{\varepsilon} \sum_{\mathcal{B}} \prod_{i=1}^{k} \delta^{\frac{p}{2k}} \|\chi_{\widetilde{\mathcal{B}}} f_i\|_2^{\frac{p}{k}}$$

for $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$ and $\varepsilon_\circ$ small enough. Since $p > 2$, by applying Hölder's inequality twice we have

$$I \lesssim \sigma^{-C_\varepsilon} \Big(\frac{1}{\delta}\Big)^{\varepsilon} \prod_{i=1}^{k} \delta^{\frac{p}{k}(\frac{1}{2} + d(1+\varepsilon)(\frac{1}{p} - \frac{1}{2}))} \Big(\sum_{\mathcal{B}} \|\chi_{\widetilde{\mathcal{B}}} f_i\|_p^p \Big)^{\frac{1}{k}} \lesssim \sigma^{-C_\varepsilon} \Big(\frac{1}{\delta}\Big)^{c\varepsilon} \Big(\prod_{i=1}^{k} \delta^{\frac{d}{p} - \frac{d-1}{2}} \|f_i\|_p \Big)^{\frac{p}{k}}.$$

For II, we use Lemma 2.9. There is a constant $C = C(\|\psi\|_{C^N(I^{d-1})})$ such that $|K_i * (\chi_{\widetilde{\mathcal{B}}^c} f_i)(x)| \le C\delta\delta^{\varepsilon(M-d-1)} \mathfrak{K}_{d+1} * |f_i|(x)$ if $x \in B$, and $|K_i * g_i(x)| \le C\delta \mathfrak{K}_{d+1} * |f_i|(x)$. Thus, we get

$$II \lesssim \delta^{\frac{(k-1)p}{k}} \delta^{\varepsilon(N-d-1)\frac{p}{k}} \int \prod_{i=1}^{k} (\mathfrak{K}_{d+1} * |f_i|(x))^{\frac{p}{k}}\, dx \lesssim \delta^{c_2 N\varepsilon - c_1} \prod_{i=1}^{k} \|f_i\|_p^{\frac{p}{k}}$$

for some $c_1, c_2 > 0$ because $\|\mathfrak{K}_{d+1} * f\|_p \le C\delta^{-d}\|f\|_p$ for $1 \le p \le \infty$ by Young's convolution inequality. Combining the two estimates for I and II with N large enough, we see that for $\varepsilon > 0$ there is an N such that

$$\Big\| \prod_{i=1}^{k} T_\delta f_i \Big\|_{L^{p/k}(\mathbb{R}^d)} \le C\sigma^{-C_\varepsilon} \Big(\frac{1}{\delta}\Big)^{c\varepsilon} \prod_{i=1}^{k} \delta^{\frac{d}{p} - \frac{d-1}{2}} \|f_i\|_p$$

for $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$ and $\varepsilon_\circ$ small enough. Therefore, choosing $\varepsilon = \varepsilon/c$, we get the desired bound (15). □

In what follows we show that if the normal vectors of the surfaces are confined in a $C\delta$-neighborhood of a k-plane in Proposition 2.11, then the associated multilinear restriction estimate has an improved bound. In particular, if one takes $p = 2k/(k-1)$, the bound in (17) is $\sim \delta^{-\varepsilon}\delta^{\frac{d}{2}}$, which is better than the corresponding bound $\sim \delta^{-\varepsilon}\delta^{\frac{k}{2}}$ in Proposition 2.10. However, it seems difficult to make use of such an improvement to get a better linear bound without using the square sum function (see Corollary 2.12 below).

Proposition 2.11. *Let $0 < \delta \ll \sigma \ll 1$, $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, and Π be a k-plane containing the origin. Suppose that $\Gamma(\psi)$, $\Gamma_1, \dots, \Gamma_k$ are given as in the above and (13) is satisfied whenever $\xi_i \in \Gamma_i$, $i = 1, \dots, k$. Suppose that*

$$\operatorname{supp} \widehat{F}_i \subset \Gamma_i(\delta) \cap \mathrm{N}^{-1}(\Pi + O(\delta)), \quad i = 1, \dots, k. \tag{16}$$

Then, if $2 \le p \le 2k/(k-1)$ and $\varepsilon_\circ$ is sufficiently small, for $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that

$$\Big\| \prod_{i=1}^{k} F_i \Big\|_{L^{p/k}(B(x,\delta^{-1}))} \le C\sigma^{-C_\varepsilon}\delta^{-\varepsilon}\delta^{dk(\frac{1}{2}-\frac{1}{p})} \prod_{i=1}^{k} \|F_i\|_2 \tag{17}$$

holds with C, C_ε, independent of ψ.

If $p/k \ge 1$, the inequality could be shown by using Hölder's inequality and the k-linear multilinear restriction estimate in [Bennett et al. 2006]. However, this is not true in general and we prove Proposition 2.11 by making use of the induction-on-scale argument and the multilinear Kakeya estimate. The following is a consequence of Proposition 2.11.

Corollary 2.12. *Suppose that the same assumptions in Proposition 2.11 hold. Let $\{\mathfrak{q}\}$, $\mathfrak{q} \subset \frac{1}{2}I^d$, be the collection of dyadic cubes of side length ℓ, $2^{-2}\delta < \ell \le 2^{-1}\delta$. Then, if $2 \le p \le 2k/(k-1)$, for $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that, for $x \in \mathbb{R}^d$,*

$$\Big\| \prod_{i=1}^{k} F_i \Big\|_{L^{p/k}(B(x,\delta^{-1}))} \le C\sigma^{-C_\varepsilon}\delta^{-\varepsilon} \prod_{i=1}^{k} \Big\| \Big(\sum_{\mathfrak{q}} |F_{i\mathfrak{q}}|^2 \Big)^{\frac{1}{2}} \rho_{B(x,\delta^{-1})} \Big\|_p \tag{18}$$

holds with C, C_ε, independent of $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$.

This may be compared with a discrete formulation of the multilinear inequality in [Bourgain and Guth 2011, (1.1), p. 1250]. The inequality (18) can be easily deduced from Proposition 2.11 by the standard argument using Plancherel's theorem and orthogonality; see the proof of Corollary 3.11. So, we omit the proof.

Proof of Proposition 2.11. For $p = 2$ the estimate (17) follows from Hölder's inequality and Plancherel's theorem. Hence, in view of interpolation, it is enough to show (17) for $p = 2k/(k-1)$.

We prove (17) by adapting the proof of the multilinear restriction estimate in [Bennett et al. 2006]. By translation we may assume $x = 0$. We make the following assumption that, for $0 < \delta \ll \sigma$ and some $\alpha > 0$,

$$\Big\| \prod_{i=1}^{k} F_i \Big\|_{L^{2/(k-1)}(B(0,\delta^{-1}))} \lesssim \delta^{-\alpha}\delta^{\frac{d}{2}} \prod_{i=1}^{k} \|F_i\|_2 \tag{19}$$

holds uniformly for $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$ whenever (16) holds and (13) is satisfied for $\xi_i \in \Gamma_i$, $i = 1, \dots, k$. It is clearly true with a large $\alpha > 0$, as can be seen by making use of Lemma 2.9. We show (19) implies that, for $\varepsilon > 0$, there is an N such that

$$\Big\|\prod_{i=1}^k F_i\Big\|_{L^{2/(k-1)}(B(0,\delta^{-1}))} \lesssim C_\varepsilon \sigma^{-\kappa} \delta^{-\frac{\alpha}{2} - c\varepsilon} \delta^{\frac{d}{2}} \prod_{i=1}^k \|F_i\|_2 \tag{20}$$

holds uniformly for $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$. In what follows we set $R = \delta^{-1}$.

Iteration of the implication from (19) to (20) allows us to suppress α as small as $\sim \varepsilon$. In fact, since the implication remains valid as long as $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, by fixing an ε and iterating the implication (19) to (20) l times, we have the bound

$$C_\varepsilon^l \sigma^{-\kappa l} R^{2^{-l}\alpha + c\varepsilon(1+2^{-1}\varepsilon+\dots+2^{-l+1})} \le C_\varepsilon^l \sigma^{-\kappa l} R^{2^{-l}\alpha + 2c\varepsilon}.$$

Choosing l such that $2^{-l}\alpha \sim \varepsilon$ gives the bound $\widetilde{C}_\varepsilon \sigma^{Ck \log\frac{\alpha}{\varepsilon}} R^{C\varepsilon}$. Hence, taking $\varepsilon = \varepsilon/C$, we get the desired bound.

Let $\{\boldsymbol{q}\}$ be the collection of dyadic cubes (hence essentially disjoint) of side length ℓ, $\ell < R^{-\frac{1}{2}} \le 2\ell$, such that $\mathbb{R}^d = \bigcup \boldsymbol{q}$. Since the Fourier transform of $\rho_{B(z,\sqrt{R})} F_i$ is supported in $\Gamma(\delta^{\frac{1}{2}}) \cap \mathrm{N}^{-1}(\Pi + O(\delta^{\frac{1}{2}}))$, by the assumption it follows that

$$\begin{aligned}\Big\|\prod_{i=1}^k F_i\Big\|_{L^{2/(k-1)}(B(z,R^{1/2}))} &\lesssim \Big\|\prod_{i=1}^k \rho_{B(z,\sqrt{R})} F_i\Big\|_{L^{2/(k-1)}(B(z,R^{1/2}))} \\ &\lesssim \delta^{-\frac{\alpha}{2}} \delta^{\frac{d}{4}} \prod_{i=1}^k \|\rho_{B(z,\sqrt{R})} F_i\|_2 \lesssim \delta^{-\frac{\alpha}{2}} \delta^{\frac{d}{4}} \prod_{i=1}^k \Big\|\rho_{B(z,\sqrt{R})} \Big(\sum_{\boldsymbol{q}} |F_{i\boldsymbol{q}}|^2\Big)^{\frac{1}{2}}\Big\|_2.\end{aligned}$$

Here $F_{i\boldsymbol{q}}$ is given by $\mathcal{F}(F_{i\boldsymbol{q}}) = \widehat{F}_i \chi_{\boldsymbol{q}}$. Since the supports of $\mathcal{F}(\rho_{B(z,\sqrt{R})} F_{i\boldsymbol{q}})$ are boundedly overlapping, the last inequality follows from Plancherel's theorem. By the rapid decay of ρ we have, for a large $M > 0$,

$$\Big\|\prod_{i=1}^k F_i\Big\|_{L^{2/(k-1)}(B(z,\sqrt{R}))} \lesssim \delta^{-\frac{\alpha}{2}} \delta^{\frac{d}{4}} \prod_{i=1}^k \Big\|\chi_{B(z,R^{1/2+\varepsilon})} \Big(\sum_{\boldsymbol{q}} |F_{i\boldsymbol{q}}|^2\Big)^{\frac{1}{2}}\Big\|_2 + \delta^M \prod_{i=1}^k \|F_i\|_2. \tag{21}$$

For a given $\xi \in \mathrm{N}^{-1}(\Pi)$, let $\{v_1, \dots, v_{k-1}\}$ be an orthonormal basis for the tangent space $T_\xi(\mathrm{N}^{-1}(\Pi))$ at ξ, $v_k = \mathrm{N}(\xi)$, and let $v_{k+1}, \dots, v_d$ form an orthonormal basis for $(\operatorname{span}\{v_1, \dots, v_{k-1}, v_k\})^\perp$. (So, the vectors $v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_d$ depend on $\xi \in \mathrm{N}^{-1}(\Pi)$.) Then, we define $\boldsymbol{p}(\xi)$ and $\boldsymbol{P}(\xi)$ by

$$\begin{aligned}\boldsymbol{p}(\xi) &= \xi + \big\{x : |x \cdot v_j| \le C_1\sqrt{\delta},\ j = 1, \dots, k-1, \text{ and } |x \cdot v_j| \le C_1\delta,\ j = k+1, \dots, d\big\}, \\ \boldsymbol{P}(\xi) &= \big\{x : |x \cdot v_j| \le C\sqrt{\delta},\ j = 1, \dots, k-1, \text{ and } |x \cdot v_j| \le C,\ j = k+1, \dots, d\big\}.\end{aligned}$$

Since $\mathrm{N}^{-1}(\Pi)$ is smooth, $\mathrm{N}^{-1}(\Pi) + O(\delta)$ can be covered by a collection of boundedly overlapping $\{\boldsymbol{p}(\xi_\alpha)\}$, $\xi_\alpha \in \mathrm{N}^{-1}(\Pi)$ (here, we are seeing $\mathrm{N}^{-1}(\Pi)$ as a subset of $\mathbb{R}^d$), such that for any $\boldsymbol{q}$ there exists ξ_α satisfying

$$\operatorname{supp} \widehat{F}_i \cap \boldsymbol{q} \subset \tfrac{1}{2}\boldsymbol{p}(\xi_\alpha) \tag{22}$$

with a sufficiently large $C_1 > 0$.

For $(i, \boldsymbol{q})$ satisfying $\operatorname{supp} \widehat{F}_i \cap \boldsymbol{q} \neq \varnothing$ let us denote by $\xi_{i,\boldsymbol{q}}$ the ξ_α which satisfies (22) (if there are more than one, we simply choose one of them). We also denote by $L(i, \boldsymbol{q})$ the bijective affine map from $\frac{1}{2}\boldsymbol{p}(\xi_{i,\boldsymbol{q}})$ to $\mathfrak{q}(0, 1)$. Then we define $\widetilde{F}_{i\boldsymbol{q}}$ by

$$\mathcal{F}(\widetilde{F}_{i\boldsymbol{q}})(\xi) = \frac{1}{\rho(L(i, \boldsymbol{q})\xi)} \widehat{F}_{i\boldsymbol{q}}(\xi).$$

We also set $\boldsymbol{P}_{i,\boldsymbol{q}} = \boldsymbol{P}(\xi_{i,\boldsymbol{q}})$ and $\boldsymbol{K}_{i,\boldsymbol{q}} = \mathcal{F}^{-1}(\rho(L(i, \boldsymbol{q})\,\cdot))$. By $R\boldsymbol{P}_{i,\boldsymbol{q}}$ we denote the rectangle which is the R-times dilation of $\boldsymbol{P}_{i,\boldsymbol{q}}$ from the center of $\boldsymbol{P}_{i,\boldsymbol{q}}$. Also denote by $\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}$ the set $R^{1+\varepsilon}\boldsymbol{P}_{i,\boldsymbol{q}}$ which is the $R^{1+\varepsilon}$-times dilation of $\boldsymbol{P}_{i,\boldsymbol{q}}$ from its center. Since $\boldsymbol{K}_{i,\boldsymbol{q}} * \widetilde{F}_{i\boldsymbol{q}} = F_{i\boldsymbol{q}}$ and $|\boldsymbol{K}_{i,\boldsymbol{q}}| \lesssim \chi_{R\boldsymbol{P}_{i,\boldsymbol{q}}}/|R\boldsymbol{P}_{i,\boldsymbol{q}}|$, we have, for $y \in B(x, 2R^{\frac{1}{2}+\varepsilon})$ and some $c > 0$,

$$|F_{i\boldsymbol{q}}(y)|^2 = |\boldsymbol{K}_{i,\boldsymbol{q}}| * |\widetilde{F}_{i\boldsymbol{q}}|^2(y) \lesssim \frac{\chi_{R\boldsymbol{P}_{i,\boldsymbol{q}}}}{|R\boldsymbol{P}_{i,\boldsymbol{q}}|} * |\widetilde{F}_{i\boldsymbol{q}}|^2(y) \lesssim R^{c\varepsilon} \frac{\chi_{\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}}}{|\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}|} * |\widetilde{F}_{i\boldsymbol{q}}|^2(x).$$

The last inequality is trivial since $|\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}| \sim R^{c\varepsilon}|R\boldsymbol{P}_{i,\boldsymbol{q}}|$ for some $c > 0$. Hence, for $x, y \in B(z, R^{\frac{1}{2}+\varepsilon})$ we have

$$\sum_{\boldsymbol{q}} |F_{i\boldsymbol{q}}|^2(y) \lesssim R^{c\varepsilon} \sum_{\boldsymbol{q}} \frac{\chi_{\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}}}{|\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}|} * |\widetilde{F}_{i\boldsymbol{q}}|^2(x). \tag{23}$$

Integrating in y over $B(z, R^{\frac{1}{2}+\varepsilon})$ for each $1 \le i \le k$, we see that, for $x \in B(z, R^{\frac{1}{2}+\varepsilon})$,

$$\prod_{i=1}^{k} \bigg\| \chi_{B(z,R^{1/2+\varepsilon})} \Big(\sum_{\boldsymbol{q}} |F_{i\boldsymbol{q}}|^2 \Big)^{\frac{1}{2}} \bigg\|_2 \lesssim R^{c\varepsilon} R^{\frac{dk}{4}} \prod_{i=1}^{k} \Big(\sum_{\boldsymbol{q}} \frac{\chi_{\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}}}{|\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}|} * |\widetilde{F}_{i\boldsymbol{q}}|^2 \Big)^{\frac{1}{2}}(x). \tag{24}$$

Now, integration in x over $B(z, R^{\frac{1}{2}+\varepsilon})$ yields

$$\prod_{i=1}^{k} \bigg\| \chi_{B(z,R^{1/2+\varepsilon})} \Big(\sum_{\boldsymbol{q}} |F_{i\boldsymbol{q}}|^2 \Big)^{\frac{1}{2}} \bigg\|_2 \lesssim R^{c\varepsilon} R^{\frac{d}{4}} \bigg\| \prod_{i=1}^{k} \Big(\sum_{\boldsymbol{q}} \frac{\chi_{\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}}}{|\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}|} * |\widetilde{F}_{i\boldsymbol{q}}|^2 \Big)^{\frac{1}{2}} \bigg\|_{L^{2/(k-1)}(B(z,R^{1/2+\varepsilon}))}. \tag{25}$$

Combining this with (21) we have, for any large $M > 0$,

$$\bigg\| \prod_{i=1}^{k} F_i \bigg\|_{L^{2/(k-1)}(B(z,\sqrt{R}))} \lesssim \delta^{-\frac{\alpha}{2}-c\varepsilon} \bigg\| \prod_{i=1}^{k} \Big(\sum_{\boldsymbol{q}} \frac{\chi_{\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}}}{|\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}|} * |\widetilde{F}_{i\boldsymbol{q}}|^2 \Big)^{\frac{1}{2}} \bigg\|_{L^{2/(k-1)}(B(z,R^{1/2+\varepsilon}))} + \delta^M \prod_{i=1}^{k} \|F_i\|_2. \tag{26}$$

We now cover $B(0, R)$ with boundedly overlapping balls $B(z, \sqrt{R})$ and use the above inequality for each of them. Then we get

$$\bigg\| \prod_{i=1}^{k} F_i \bigg\|_{L^{2/(k-1)}(B(0,R))} \lesssim \delta^{-\frac{\alpha}{2}-c\varepsilon} \bigg\| \prod_{i=1}^{k} \Big(\sum_{\boldsymbol{q}} \frac{\chi_{\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}}}{|\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}|} * |\widetilde{F}_{i\boldsymbol{q}}|^2 \Big)^{\frac{1}{2}} \bigg\|_{L^{2/(k-1)}(B(0,2R))} + \delta^{M-C} \prod_{i=1}^{k} \|F_i\|_2.$$

Here we have an increased c because of the overlapping of the balls $B(z, R^{\frac{1}{2}+\varepsilon})$ in the right-hand side. Since $\sum_{\boldsymbol{q}} \|\widetilde{F}_{i\boldsymbol{q}}\|_2^2 \sim \|F_i\|_2^2$, for (20) it is sufficient to show

$$\Big\|\prod_{i=1}^{k}\Big(\sum_{\boldsymbol{q}} \frac{\chi_{\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}}}{|\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}|} * |\widetilde{F}_{i\boldsymbol{q}}|^2\Big)\Big\|_{L^{1/(k-1)}(B(0,2R))} \lesssim \sigma^{-\kappa}\delta^{\frac{d}{2}-c\varepsilon}\prod_{i=1}^{k}\Big(\sum_{\boldsymbol{q}}\|\widetilde{F}_{i\boldsymbol{q}}\|_2^2\Big).$$

By rescaling this is equivalent to

$$\Big\|\prod_{i=1}^{k}\Big(\sum_{\boldsymbol{q}} \frac{\chi_{\boldsymbol{P}_{i,\boldsymbol{q}}}}{|\boldsymbol{P}_{i,\boldsymbol{q}}|} * f_{i,\boldsymbol{q}}\Big)\Big\|_{L^{1/(k-1)}(B(0,2))} \lesssim \sigma^{-\kappa}R^{c\varepsilon}\prod_{i=1}^{k}\Big(\sum_{\boldsymbol{q}}\|f_{i,\boldsymbol{q}}\|_1\Big). \tag{27}$$

Let $\mathcal{I}_i = \{\boldsymbol{q} : \operatorname{supp} \widehat{F}_i \cap \boldsymbol{q} \neq \varnothing\}$, $I_i \subset \mathcal{I}_i$ and $\mathcal{T}_{i,\boldsymbol{q}}$ be a finite subset of $\mathbb{R}^d$. Allowing the loss of $(\log R)^C$ in bound, by a standard reduction with pigeon-holing it suffices to show

$$\Big\|\prod_{i=1}^{k}\Big(\sum_{\boldsymbol{q}\in I_i}\sum_{\tau\in\mathcal{T}_{i,\boldsymbol{q}}} \chi_{\boldsymbol{P}_{i,\boldsymbol{q}}+\tau}\Big)\Big\|_{L^{1/(k-1)}(B(0,2))} \lesssim \sigma^{-\frac{\kappa}{2}}R^{c\varepsilon}\prod_{i=1}^{k}\Big(\sum_{\boldsymbol{q}\in I_i}\sum_{\tau\in\mathcal{T}_{i,\boldsymbol{q}}}|\boldsymbol{P}_{i,\boldsymbol{q}}+\tau|\Big). \tag{28}$$

We write $x = (u, v) \in \Pi \times \Pi^{\perp} (= \mathbb{R}^d)$. Then the left-hand side is clearly bounded by

$$\sup_{v\in\Pi^{\perp}}\Big\|\prod_{i=1}^{k}\Big(\sum_{\boldsymbol{q}\in I_i}\sum_{\tau\in\mathcal{T}_{i,\boldsymbol{q}}} \chi_{\boldsymbol{P}_{i,\boldsymbol{q}}+\tau}(\,\cdot\,,v)\Big)\Big\|_{L^{1/(k-1)}(\widetilde{B}(0,2))},$$

where $\widetilde{B}(0,\rho) \subset \mathbb{R}^k$ is the ball of radius ρ which is centered at the origin.

For $v \in \Pi^{\perp}$ let us set

$$(\boldsymbol{P}_{i,\boldsymbol{q}}+\tau)^v = \{u : (u,v) \in \boldsymbol{P}_{i,\boldsymbol{q}}+\tau\}.$$

Then $(\boldsymbol{P}_{i,\boldsymbol{q}}+\tau)^v$ is contained in a tube of length ~ 1 and width $CR^{-\frac{1}{2}}$, with axis parallel to $\mathrm{N}(\xi_{i,\boldsymbol{q}})$. This is because the longer sides of $\boldsymbol{P}_{i,\boldsymbol{q}}$, except the one parallel to $\mathrm{N}(\xi_{i,\boldsymbol{q}})$, are transversal to Π. More precisely, we can show that if $\varepsilon_\circ$ is sufficiently small and N is large enough, there a constant $c > 0$, independent of $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, such that, for $w \in \big(T_{\xi_{i,\boldsymbol{q}}}(\mathrm{N}^{-1}(\Pi)) \oplus \operatorname{span}\{\mathrm{N}(\xi_{i,\boldsymbol{q}})\}\big)^{\perp}$,

$$\measuredangle(w, \Pi) \geq c > 0. \tag{29}$$

Since (13) is satisfied whenever $\xi_i \in \Gamma_i$, $i = 1, \dots, k$, we know $\mathrm{N}(\xi_{1,\boldsymbol{q}}), \dots, \mathrm{N}(\xi_{k,\boldsymbol{q}})$ which are, respectively, parallel to the axes of tubes $(\boldsymbol{P}_{1,\boldsymbol{q}}+\tau)^v, \dots, (\boldsymbol{P}_{k,\boldsymbol{q}}+\tau)^v$, satisfy $|\operatorname{Vol}(\mathrm{N}(\xi_{1,\boldsymbol{q}}), \dots, \mathrm{N}(\xi_{k,\boldsymbol{q}})) \gtrsim \sigma$. Also note that $|\boldsymbol{P}_{i,\boldsymbol{q}}^v + \tilde{\tau}| \sim |\boldsymbol{P}_{i,\boldsymbol{q}}|$. Hence, by the multilinear Kakeya estimate in $\mathbb{R}^k$ (Theorem 3.7) it follows that

$$\Big\|\prod_{i=1}^{k}\Big(\sum_{\boldsymbol{q},\tau} \chi_{\boldsymbol{P}_{i,\boldsymbol{q}}+\tau}(\,\cdot\,,v)\Big)\Big\|_{L^{1/(k-1)}(\widetilde{B}(0,2))} \lesssim \sigma^{-1}\prod_{i=1}^{k}\Big(\sum_{\boldsymbol{q},\tau}|\boldsymbol{P}_{i,\boldsymbol{q}}+\tau|\Big).$$

This gives the desired inequality (28).

Now it remains to show (29). By continuity, taking sufficiently small $\varepsilon_\circ$, we only need to show (29) when $\psi = \psi_\circ$ since $\|\psi - \psi_\circ\|_{C^N(I^{d-1})} \leq \varepsilon_\circ$. Though it is easy to show and intuitively obvious, we include a proof for clarity. By rotation we may assume $\Pi \cap \{x_d = -1\} = \{(\boldsymbol{y}, \boldsymbol{a}, -1) : \boldsymbol{y} \in \mathbb{R}^{k-1}\}$ for some $\boldsymbol{a} \in \mathbb{R}^{d-k}$. Since Π contains the origin, Π can parametrized (except for $\Pi \cap \{x_d = 0\}$) as follows:

$$s(\boldsymbol{y}, \boldsymbol{a}, -1), \quad s \in \mathbb{R}, \ \boldsymbol{y} \in \mathbb{R}^{k-1}. \tag{30}$$

We may assume $\Gamma_i(\delta) \cap (\mathrm{N}^{-1}(\Pi) + O(\delta)) \neq \varnothing$ because otherwise $F_i = 0$ and there is nothing to prove. Since $\mathrm{N}(\Gamma) \cap \Pi = \varnothing$ if $|\boldsymbol{a}|$ is large, we may assume that $|\boldsymbol{a}| \le C$ for some $C > 0$ and note that $\xi_{i,\boldsymbol{q}} \in \Gamma(\psi)$. Furthermore, it suffices to show that

$$\Pi \cap \big(T_{\xi_{i,\boldsymbol{q}}}(\mathrm{N}^{-1}(\Pi)) \oplus \mathrm{span}\{\mathrm{N}(\xi_{i,\boldsymbol{q}})\}\big)^{\perp} = \{0\}, \tag{31}$$

which implies $\measuredangle(w, \Pi) > 0$ if $w \in \big(T_{\xi_{i,\boldsymbol{q}}}(\mathrm{N}^{-1}(\Pi)) \oplus \mathrm{span}\{\mathrm{N}(\xi_{i,\boldsymbol{q}})\}\big)^{\perp}$. Then, by continuity and compactness (29) follows. We now verify (29) with $\psi = \psi_\circ$. By rotation we may assume $\boldsymbol{a} = (0, \dots, 0, a) =: (\boldsymbol{0}, a) \in \mathbb{R}^{d-k-1} \times \mathbb{R}$. Using the above parametrization of Π, we see that

$$\Pi = \mathrm{span}\{e_1, \dots, e_{k-1}, (0, \dots, 0, \boldsymbol{0}, a, -1)\}.$$

The normal vector at $(x', |x'|^2/2) \in \mathbb{R}^{d-1} \times \mathbb{R}$ is parallel to $(x', -1)$. Hence, if $(x', |x'|^2/2) \in \mathrm{N}^{-1}(\Pi)$, that is, $(x', -1) \in \Pi$, then x' takes the form $x' = (\boldsymbol{y}, \boldsymbol{a})$ because of (30). Hence, it follows that $\mathrm{N}^{-1}(\Pi) = \big\{\big(\boldsymbol{y}, \boldsymbol{0}, a, \tfrac{1}{2}(|\boldsymbol{y}|^2 + |a|^2)\big)\big\}$. Then, if $\xi_{i,\boldsymbol{q}} = \big(\boldsymbol{y}, \boldsymbol{0}, a, \tfrac{1}{2}(|\boldsymbol{y}|^2 + a^2)\big)$, we have $T_{\xi_{i,\boldsymbol{q}}}(\mathrm{N}^{-1}(\Pi))$ is spanned by $\boldsymbol{y}_1 = (1, 0, \dots, \boldsymbol{0}, 0, y_1)$, $\boldsymbol{y}_2 = (0, 1, \dots, 0, \boldsymbol{0}, 0, y_2)$, $\dots$, $\boldsymbol{y}_{k-1} = (0, 0, \dots, 1, \boldsymbol{0}, 0, y_{k-1})$. For (31) it is sufficient to show that $\mathfrak{P} := \Pi \cap (\mathrm{span}\{(\boldsymbol{y}, \boldsymbol{0}, a, -1), \boldsymbol{y}_1, \dots, \boldsymbol{y}_{k-1}\})^{\perp} = \{0\}$. Let $w \in \mathfrak{P}$. Then, since $w \in \mathrm{span}\{e_1, \dots, e_{k-1}, (0, \dots, 0, \boldsymbol{0}, a, -1)\}$, we may write $w = (c_1, \dots, c_{k-1}, \boldsymbol{0}, c_k a, -c_k)$. Also, $w \cdot \boldsymbol{y}_1 = \dots = w \cdot \boldsymbol{y}_{k-1} = w \cdot (\boldsymbol{y}, \boldsymbol{0}, a, -1) = 0$ gives $c_1 = \dots = c_k = 0$. So, $v = 0$ and, hence, we get (31). ☐

2D. *Scattered modulation sum of scale σ*. When the Fourier transform of a given function f is supported in a ball of radius σ, then f behaves as though it were constant on balls of radius σ^{-1}. This observation has important role in Bourgain and Guth's argument [2011] and is widely taken for granted without being made rigorous. There seems to be several ways which make this heuristic rigorous; see [Tao and Vargas 2000b; Tao 1999]. For this purpose we make use of the Fourier series expansion.

Fix $\sigma > 0$ and large positive constants $M = M(d) \ge 100d$ and C_M which are to be chosen to be large. For $l \in \sigma^{-1}\mathbb{Z}^d$ we set

$$A_l = A_l(\sigma) = C_M(1 + |\sigma l|)^{-M}, \quad \tau_l f(x) = f(x - l). \tag{32}$$

For $\sigma > 0$, we define $[F]_\sigma$, $|[F]|_\sigma$ (the scattered modulation sum of σ-scale) by

$$[F]_\sigma(x) = \sum_{l \in \sigma^{-1}\mathbb{Z}^d} A_l |\tau_l F(x)|, \quad |[F]|_\sigma(x) = \sum_{l_1, l_2 \in \sigma^{-1}\mathbb{Z}^d} A_{l_1} A_{l_2} |\tau_{l_1 + l_2} F(x)|. \tag{33}$$

Lemma 2.13. *Let $\xi_0, x_0 \in \mathbb{R}^d$. Suppose that F is a function with $\widehat{F}$ supported in $\mathfrak{q}(\xi_0, \sigma)$. Then, if $x \in \mathfrak{q}\big(x_0, \frac{1}{\sigma}\big)$,*

$$|F(x)| \le [F]_\sigma(x_0) \le |[F]|_\sigma(x).$$

It should be noted that the inequality holds regardless of ξ_0, x_0, and σ.

Proof. Let a be a smooth function supported in $[-\pi,\pi]^d$ and $a(x)=1$ if $|x_i|\le 1$, $i=1,\dots,d$. Set

$$A(x,\xi)=a(x)a(\xi)e^{ix\cdot\xi}.$$

Since $|\partial_\xi^\alpha A|\le C_\alpha$ for any multi-indices α, by expanding into Fourier series in ξ we have

$$a(x)a(\xi)e^{ix\cdot\xi}=\sum_{l\in\mathbb{Z}^d}a_l(x)e^{-i\xi\cdot l},\quad x,\xi\in[-\pi,\pi]^d, \tag{34}$$

while a_l satisfies $|a_l(x)|\le C_M(1+|l|)^{-M}$ for any large $M>0$. On the other hand, from the inversion formula we have

$$F(x)=(2\pi)^{-d}\int e^{i(x-x_0)\cdot\xi_0}e^{i(x-x_0)\cdot(\xi-\xi_0)}e^{ix_0\cdot\xi}\widehat{F}(\xi)\,d\xi.$$

Hence, since $x\in\mathfrak{q}\big(x_0,\frac{1}{\sigma}\big)$, inserting the harmless bump function a, we may write

$$F(x)=(2\pi)^{-d}e^{i(x-x_0)\cdot\xi_0}\int A\bigg(\sigma(x-x_0),\frac{\xi-\xi_0}{\sigma}\bigg)e^{ix_0\cdot\xi}\widehat{F}(\xi)\,d\xi.$$

Using (34) we have

$$F(x)=(2\pi)^{-d}e^{i(x-x_0)\cdot\xi_0}\sum_{l\in\mathbb{Z}^d}a_l(\sigma(x-x_0))\int e^{-i\frac{(\xi-\xi_0)}{\sigma}\cdot l}e^{ix_0\cdot\xi}\widehat{F}(\xi)\,d\xi.$$

Then it follows that

$$|F(x)|\le\sum_{l\in\sigma^{-1}\mathbb{Z}^d}A_l|\tau_lF(x_0)|\le\sum_{l_1,l_2\in\sigma^{-1}\mathbb{Z}^d}A_{l_1}A_{l_2}|\tau_{(l_1+l_2)}F(x)|. \tag{35}$$

The second inequality follows by applying the first one to each τ_lF with the roles of x, x_0 interchanged. □

2E. *Multiscale decomposition.* We now attempt to bound part of $T_\delta f$ with a sum of products which satisfy the transversality assumption, while the remaining parts are given by a sum of functions which have relatively small Fourier supports. The first is rather directly estimated by making use of the multilinear estimates and the latter is to be handled by Proposition 2.5, the induction assumption and Lemma 2.6.

In what follows, we basically adapt the idea in [Bourgain and Guth 2011]. However, concerning the decomposition in that paper, reappearance of many small-scale functions in large-scale decomposition becomes problematic when one attempts to sum up resulting estimates. For the adjoint restriction estimates this can be overcome by using L^∞-functions, as was done in [Bourgain and Guth 2011]. But such an argument doesn't work for the multiplier operators and leads to a loss in its bound. To get over this, unlike the decomposition in [Bourgain and Guth 2011] where one starts to decompose with d-linear products and proceeds by reducing the degree multilinearity based on dichotomy, we decompose the multiplier operator by increasing the degree of multilinearity in order to avoid small-scale functions appearing inside of large-scale ones. This has a couple of advantages. First, this allows us to keep the function relatively intact in the course of decomposition so that we can easily add up decomposed pieces to obtain the sharp L^p bound. Secondly, the decomposition makes it possible to obtain directly obtain the L^p-L^p estimate. Hence we don't need to rely on the factorization theorem to deduce L^p-L^p from $L^\infty-L^p$. (The same is also true for the adjoint restriction operators.) Hence, we can obtain the sharp L^p bounds for multiplier operators of Bochner–Riesz-type, which lack symmetry.

2E1. *Spatial and frequency dyadic cubes.* Let $0 < \varepsilon_\circ \ll 1$, $1 \ll N$, $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, and T_δ be given by (9). Let $\kappa = \kappa(\varepsilon_\circ, N)$ be the number given in Proposition 2.5 so that (11) holds whenever $0 < \varepsilon \le \kappa$ and $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$. Let m be an integer such that $2 \le m \le d-1$, and $\sigma_1, \dots, \sigma_m$ be dyadic numbers such that

$$\delta \ll \sigma_m \ll \cdots \ll \sigma_1 \ll \min(\kappa, 1). \tag{36}$$

These numbers will be specified to terminate induction. We call σ_i the *i-th scale*.

Let us denote by $\{\mathfrak{q}^i\}$ the collection of the dyadic cubes $\mathfrak{q}^i$ of side length $2\sigma_i$ which are contained in I^d (so, $\mathfrak{q}^i$ denotes the member of $\{\mathfrak{q}^i\}$ and the cubes $\mathfrak{q}^i$ are essentially disjoint). Rather than introducing a new notation to denote each collection of $\mathfrak{q}^i$, we take the convention that $\{\mathfrak{q}^i\}$ denotes the collection of all dyadic cubes of side length $2\sigma_i$ contained in I^d if it is not specified otherwise. For each i-th scale there is a unique collection so that there will be no ambiguity, and we also use $\mathfrak{q}^i$ as indices which run over the set $\{\mathfrak{q}^i\}$. Thus, we may write

$$\bigcup_{\mathfrak{q}^i} \mathfrak{q}^i = I^d. \tag{37}$$

For the rest of this section, we assume that

$$\operatorname{supp} \hat{f} \subset \tfrac{1}{2} I^d.$$

Since $f = \sum_{\mathfrak{q}^i} f_{\mathfrak{q}^i}$, for $i = 1, \dots, m$, we write

$$T_\delta f = \sum_{\mathfrak{q}^i} T_\delta f_{\mathfrak{q}^i}. \tag{38}$$

Clearly, we may assume that $\mathfrak{q}^i$ is contained in a $C\sigma_i$-neighborhood of the surface $\Gamma(\psi)$ because $T_\delta f_{\mathfrak{q}^i} = 0$ otherwise. So, in what follows, $\mathfrak{q}^i, \mathfrak{q}^i_1, \dots, \mathfrak{q}^i_{i+1}$ and $\mathfrak{q}^i_*$ denote the elements of $\{\mathfrak{q}^i\}$.

For convenience we extend in a trivial way the map N defined on $\Gamma(\psi)$ to the cube I^d by setting, for $\xi = (\zeta, \tau) \in I^d$,

$$\boldsymbol{n}(\zeta, \tau) = \mathrm{N}(\zeta, \psi(\zeta)).$$

This extension is not necessarily needed in what follows because we only consider a small neighborhood of $\Gamma(\psi)$. However, this allows us to define a normal vector for any point in I^d and makes exposition simpler. This definition of $\boldsymbol{n}$ agrees with the one given in the next section.

Definition 2.14. Let k be an integer such that $1 \le k \le m$ and fix a constant $c > 0$. Let $\mathfrak{q}^k_1, \dots, \mathfrak{q}^k_{k+1} \in \{\mathfrak{q}^k\}$ (k-th scale cubes). We say $\mathfrak{q}^k_1, \mathfrak{q}^k_2, \dots, \mathfrak{q}^k_{k+1}$ *are* $(\sigma_1, \sigma_2, \dots, \sigma_k)$*-transversal* if

$$\mathrm{Vol}(\boldsymbol{n}(\xi_1), \boldsymbol{n}(\xi_2), \dots, \boldsymbol{n}(\xi_{k+1})) \ge c\sigma_1\sigma_2 \cdots \sigma_k, \tag{39}$$

whenever $\xi_i \in \mathfrak{q}^k_i$, $i = 1, \dots, k+1$. And we simply denote this by "$\mathfrak{q}^k_1, \mathfrak{q}^k_2, \dots, \mathfrak{q}^k_{k+1}$ trans" and say $\mathfrak{q}^k_1, \mathfrak{q}^k_2, \dots, \mathfrak{q}^k_{k+1}$ are transversal, omitting dependence on $\sigma_1, \sigma_2, \dots, \sigma_k$.

Let us set

$$M_i = \frac{1}{\sigma_i}, \quad i = 1, \dots, m.$$

We denote by $\{\mathfrak{Q}^i\}$ the collection of the dyadic cubes of side length $2M_i$, which covers $\mathbb{R}^d$ (so, $\mathfrak{Q}^i$ again denotes a member of the sets $\{\mathfrak{Q}^i\}$). We write[†]

$$\bigcup_{\mathfrak{Q}^i} \mathfrak{Q}^i = \mathbb{R}^d. \tag{40}$$

Since the Fourier support of $T_\delta f_{\mathfrak{q}^i}$ is contained $\mathfrak{q}^i$, it may be thought of as a constant on $\mathfrak{Q}^i$ by invoking Lemma 2.13 with $\sigma = \sigma_i$. Since the scale σ_i is clear from the side length of the cube $\mathfrak{q}^i$, we simply set

$$[T_\delta f_{\mathfrak{q}^i}] := [T_\delta f_{\mathfrak{q}^i}]_{\sigma_i}, \quad \|[T_\delta f_{\mathfrak{q}^i}]\| := \|[T_\delta f_{\mathfrak{q}^i}]\|_{\sigma_i}.$$

2E2. *σ_1-scale decomposition.* Bilinear decomposition is rather elementary. Fix $x \in \mathbb{R}^d$. From (38),

$$|T_\delta f(x)| \le \sum_{\mathfrak{q}^1} |T_\delta f_{\mathfrak{q}^1}(x)|.$$

We denote by $\mathfrak{q}^1_* = \mathfrak{q}^1_*(x)$ a cube $\mathfrak{q}^1 \in \{\mathfrak{q}^1\}$ such that $|T_\delta f_{\mathfrak{q}^1_*}(x)| = \max_{\mathfrak{q}^1} |T_\delta f_{\mathfrak{q}^1_*}(x)|$. (There may be many such cubes but $\mathfrak{q}^1_*$ denotes just one of them.) Then we consider the following two cases separately:

$$\sum_{\mathfrak{q}^1} |T_\delta f_{\mathfrak{q}^1}(x)| \le 100^d |T_\delta f_{\mathfrak{q}^1_*}(x)|, \quad \sum_{\mathfrak{q}^1} |T_\delta f_{\mathfrak{q}^1}(x)| > 100^d |T_\delta f_{\mathfrak{q}^1_*}(x)|.$$

For the second case

$$\sum_{\mathrm{dist}(\mathfrak{q}^1, \mathfrak{q}^1_*) < 10\sigma_1} |T_\delta f_{\mathfrak{q}^1}(x)| < 50^d |T_\delta f_{\mathfrak{q}^1_*}(x)| \le 2^{-d} \sum_{\mathfrak{q}^1} |T_\delta f_{\mathfrak{q}^1}(x)|.$$

Hence there is $\mathfrak{q}^1_1 \in \{\mathfrak{q}^1\}$ such that $\mathrm{dist}(\mathfrak{q}^1_1, \mathfrak{q}^1_*) \ge 10\sigma_1$ and

$$\sum_{\mathfrak{q}^1} |T_\delta f_{\mathfrak{q}^1}(x)| \lesssim \sigma_1^{-(d-1)} |T_\delta f_{\mathfrak{q}^1_1}(x)| \le \sigma_1^{-(d-1)} |T_\delta f_{\mathfrak{q}^1_1}(x) T_\delta f_{\mathfrak{q}^1_*}(x)|^{\frac12}.$$

From these two cases we get

$$\sum_{\mathfrak{q}^1} |T_\delta f_{\mathfrak{q}^1}(x)| \lesssim \max_{\mathfrak{q}^1} |T_\delta f_{\mathfrak{q}^1}(x)| + C\sigma_1^{-\frac{d-1}{2}} \max_{\mathrm{dist}(\mathfrak{q}^1_1, \mathfrak{q}^1_2) \gtrsim \sigma_1} |T_\delta f_{\mathfrak{q}^1_1}(x) T_\delta f_{\mathfrak{q}^1_2}(x)|^{\frac12}. \tag{41}$$

Using the imbedding $\ell^p \subset \ell^\infty$, Proposition 2.5 and Lemma 2.6 give

$$\|\max_{\mathfrak{q}^1} |T_\delta f_{\mathfrak{q}^1}\|_p \le \Big(\sum_{\mathfrak{q}^1} \|T_\delta f_{\mathfrak{q}^1}\|_p^p\Big)^{\frac1p} \le \Big(\sum_{\mathfrak{q}^1} A(\sigma_1^{-2}\delta)^p \|f_{\mathfrak{q}^1}\|_p^p\Big)^{\frac1p} \lesssim A(\sigma_1^{-2}\delta)\|f\|_p. \tag{42}$$

Hence, combining this with (41), we have

$$\|T_\delta f\|_p \lesssim A(\sigma_1^{-2}\delta)\|f\|_p + \sigma_1^{-C} \max_{\mathrm{dist}(\mathfrak{q}^1_1, \mathfrak{q}^1_2) \gtrsim \sigma_1} \|T_\delta f_{\mathfrak{q}^1_1} T_\delta f_{\mathfrak{q}^1_2}\|_{\frac{p}{2}}^{\frac12}. \tag{43}$$

We now proceed to decompose the bilinear expression appearing in the left-hand side.

† Here we take the same convention for $\{\mathfrak{Q}^i\}$ as we do for $\{\mathfrak{q}^i\}$.

In the following section we explain how one can achieve trilinear decomposition out of (43) before we inductively obtain the full k-linear decomposition which we need for the proof of Theorem 1.1. Once one gets familiar with it, extension to a higher degree of multilinearity becomes more or less obvious.

2E3. *σ_2-scale decomposition.* Suppose that we are given two cubes $\mathfrak{q}_1^1$ and $\mathfrak{q}_2^1$ of first scale such that $\mathrm{dist}(\mathfrak{q}_1^1, \mathfrak{q}_2^1) \gtrsim \sigma_1$. For $i = 1, 2$, we denote by $\{\mathfrak{q}_i^2\}$ the collection of dyadic cubes $\mathfrak{q}_i^2$ of side length σ_2 contained in $\mathfrak{q}_i^1$ such that

$$\mathfrak{q}_i^1 = \bigcup_{\mathfrak{q}_i^2} \mathfrak{q}_i^2, \quad i = 1, 2. \tag{44}$$

We also denote by $\{\mathfrak{q}^2\}$ the set $\{\mathfrak{q}_1^2\} \cup \{\mathfrak{q}_2^2\}$. Then it follows that

$$T_\delta f_{\mathfrak{q}_i^1} = \sum_{\mathfrak{q}_i^2} T_\delta f_{\mathfrak{q}_i^2}, \quad i = 1, 2. \tag{45}$$

We may also assume that $\mathfrak{q}_1^2$, $\mathfrak{q}_2^2$ are contained in the $C\sigma_2$-neighborhood of $\Gamma(\psi)$ because $T_\delta f_{\mathfrak{q}_1^2}$, $T_\delta f_{\mathfrak{q}_2^2}$ are zero otherwise.

Decomposition from this stage is no longer as simple as in the σ_1-scale case. We need to use spatial localization in order to compare the values of the decomposed pieces. This makes it possible to bound large parts of the operator with transversal products.

Let us fix a cube $\mathfrak{Q}^2$ and x_0 be the center of $\mathfrak{Q}^2$. Let $\mathfrak{q}_{1*}^2 \in \{\mathfrak{q}_1^2\}$, $\mathfrak{q}_{2*}^2 \in \{\mathfrak{q}_2^2\}$ be the cubes such that

$$[T_\delta f_{\mathfrak{q}_{1*}^2}](x_0) = \max_{\mathfrak{q}_1^2}[T_\delta f_{\mathfrak{q}_1^2}](x_0), \quad [T_\delta f_{\mathfrak{q}_{2*}^2}](x_0) = \max_{\mathfrak{q}_2^2}[T_\delta f_{\mathfrak{q}_2^2}](x_0).$$

Let us define $\Lambda_i^2 \subset \{\mathfrak{q}_i^2\}$, $i = 1, 2$, by

$$\Lambda_i^2 = \big\{\mathfrak{q}_i^2 : [T_\delta f_{\mathfrak{q}_i^2}](x_0) \ge \sigma_2^{2d} \max([T_\delta f_{\mathfrak{q}_{1*}^2}](x_0), [T_\delta f_{\mathfrak{q}_{2*}^2}](x_0))\big\}.$$

Using (45), we split the summation to get

$$T_\delta f_{\mathfrak{q}_1^1} T_\delta f_{\mathfrak{q}_2^1} = \sum_{(\mathfrak{q}_1^2, \mathfrak{q}_2^2) \in \Lambda_1 \times \Lambda_2} T_\delta f_{\mathfrak{q}_1^2} T_\delta f_{\mathfrak{q}_2^2} + \sum_{(\mathfrak{q}_1^2, \mathfrak{q}_2^2) \notin \Lambda_1 \times \Lambda_2} T_\delta f_{\mathfrak{q}_1^2} T_\delta f_{\mathfrak{q}_2^2}. \tag{46}$$

Since there are at most $O(\sigma_2^{-2(d-1)})$ pairs $(\mathfrak{q}_1^2, \mathfrak{q}_2^2)$, the second sum in the right-hand side is bounded by

$$\sum_{(\mathfrak{q}_1^2, \mathfrak{q}_2^2) \notin \Lambda_1 \times \Lambda_2} |T_\delta f_{\mathfrak{q}_1^2}(x)||T_\delta f_{\mathfrak{q}_2^2}(x)| \le \sigma_2^d \max_{\mathfrak{q}^2}([T_\delta f_{\mathfrak{q}^2}](x_0))^2. \tag{47}$$

For a cube $\mathfrak{q}$ we denote by $\boldsymbol{c}(\mathfrak{q})$ the center of $\mathfrak{q}$. Let $\Pi = \Pi(\mathfrak{q}_{1*}^2, \mathfrak{q}_{2*}^2)$ be the 2-plane which is spanned by $\boldsymbol{n}_1 = \boldsymbol{n}(\boldsymbol{c}(\mathfrak{q}_{1*}^2))$, $\boldsymbol{n}_2 = \boldsymbol{n}(\boldsymbol{c}(\mathfrak{q}_{2*}^2))$, and define

$$\mathfrak{N} = \mathfrak{N}(\mathfrak{Q}^2, \mathfrak{q}_1^1, \mathfrak{q}_2^1) = \{\mathfrak{q}^2 \in \Lambda_1^2 \cup \Lambda_2^2 : \mathrm{dist}(\boldsymbol{n}(\mathfrak{q}^2), \Pi) \le C\sigma_2\}. \tag{48}$$

Clearly, $\mathrm{Vol}(\boldsymbol{n}_1, \boldsymbol{n}_2) \gtrsim \sigma_1$ and $\mathrm{dist}(\boldsymbol{n}(\mathfrak{q}^2), \Pi) \gtrsim \sigma_2$ if $\mathfrak{q}^2 \notin \mathfrak{N}$. Since $\sigma_1 \gg \sigma_2$, if $\mathfrak{q}^2 \notin \mathfrak{N}$, then $\mathrm{Vol}(\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{n}(\xi)) \gtrsim \sigma_1\sigma_2$ for $\xi \in \mathfrak{q}^2$. Also, $\boldsymbol{n}(\mathfrak{q}_{i*}^2) \subset \boldsymbol{n}_i + O(\sigma_2)$, $i = 1, 2$. So, it follows that

$$\mathrm{Vol}(\boldsymbol{n}(\xi_1), \boldsymbol{n}(\xi_2), \boldsymbol{n}(\xi_3)) \gtrsim \sigma_1\sigma_2 \tag{49}$$

if $\xi_1 \in \mathfrak{q}^2_{1*}$, $\xi_2 \in \mathfrak{q}^2_{2*}$, and $\xi_3 \in \mathfrak{q}^2 \not\subset \mathfrak{N}$. That is, $\mathfrak{q}^2_{1*}$, $\mathfrak{q}^2_{2*}$, $\mathfrak{q}^2$ are transversal. Hence, we split $\sum_{(\mathfrak{q}^2_1,\mathfrak{q}^2_2)\in\Lambda_1\times\Lambda_2} T_\delta f_{\mathfrak{q}^2_1} T_\delta f_{\mathfrak{q}^2_2}$ into

$$\sum_{\substack{(\mathfrak{q}^2_1,\mathfrak{q}^2_2)\in\Lambda_1\times\Lambda_2\\ \mathfrak{q}^2_1,\mathfrak{q}^2_2\in\mathfrak{N}}} T_\delta f_{\mathfrak{q}^2_1}(x) T_\delta f_{\mathfrak{q}^2_2}(x) + \sum_{\substack{(\mathfrak{q}^2_1,\mathfrak{q}^2_2)\in\Lambda_1\times\Lambda_2\\ \mathfrak{q}^2_1 \text{ or } \mathfrak{q}^2_2\notin\mathfrak{N}}} T_\delta f_{\mathfrak{q}^2_1}(x) T_\delta f_{\mathfrak{q}^2_2}(x). \tag{50}$$

Each term appearing in the second sum can be bounded by a product of three operators which satisfy the transversality condition. Indeed, suppose that $(\mathfrak{q}^2_1,\mathfrak{q}^2_2)\in\Lambda_1\times\Lambda_2$ and $\mathfrak{q}^2_2\not\subset\mathfrak{N}$. The case $\mathfrak{q}^2_1\not\subset\mathfrak{N}$ can be handled similarly by symmetry. Since $[T_\delta f_{\mathfrak{q}^2_2}](x_0)\ge\sigma_2^{2d}[T_\delta f_{\mathfrak{q}^2_{1*}}](x_0)$, we have

$$\begin{aligned}[T_\delta f_{\mathfrak{q}^2_1}](x_0)[T_\delta f_{\mathfrak{q}^2_2}](x_0) &\le \big([T_\delta f_{\mathfrak{q}^2_{1*}}](x_0)[T_\delta f_{\mathfrak{q}^2_2}](x_0)\big)^{\frac{2}{3}}\big([T_\delta f_{\mathfrak{q}^2_1}](x_0)[T_\delta f_{\mathfrak{q}^2_2}](x_0)\big)^{\frac{1}{3}}\\ &\le \sigma_2^{-\frac{2d}{3}}\big([T_\delta f_{\mathfrak{q}^2_{1*}}](x_0)[T_\delta f_{\mathfrak{q}^2_{2*}}](x_0)[T_\delta f_{\mathfrak{q}^2_2}](x_0)\big)^{\frac{2}{3}}.\end{aligned}$$

Hence, from this and (49) it follows that

$$\Bigg|\sum_{\substack{(\mathfrak{q}^2_1,\mathfrak{q}^2_2)\in\Lambda_1\times\Lambda_2\\ \mathfrak{q}^2_1 \text{ or } \mathfrak{q}^2_2\notin\mathfrak{N}}} T_\delta f_{\mathfrak{q}^2_1}(x) T_\delta f_{\mathfrak{q}^2_2}(x)\Bigg| \le \sigma_2^{-C}\sum_{\mathfrak{q}^2_1,\mathfrak{q}^2_2,\mathfrak{q}^2_3 \text{ trans}}\Bigg(\prod_{i=1}^3[T_\delta f_{\mathfrak{q}^2_i}](x_0)\Bigg)^{\frac{2}{3}}. \tag{51}$$

We combine (46), (47), (50) and (51) to get, for $x\in\mathfrak{Q}^2$,

$$\begin{aligned}&|T_\delta f_{\mathfrak{q}^1_1}(x) T_\delta f_{\mathfrak{q}^1_2}(x)|\\ &\quad\le \sigma_2^d\big(\max_{\mathfrak{q}^2}[T_\delta f_{\mathfrak{q}^2}](x_0)\big)^2 + \Bigg|\sum_{\substack{(\mathfrak{q}^2_1,\mathfrak{q}^2_2)\in\Lambda_1\times\Lambda_2\\ \mathfrak{q}^2_1,\mathfrak{q}^2_2\in\mathfrak{N}}} T_\delta f_{\mathfrak{q}^2_1}(x) T_\delta f_{\mathfrak{q}^2_2}(x)\Bigg| + \sigma_2^{-C}\sum_{\mathfrak{q}^2_1,\mathfrak{q}^2_2,\mathfrak{q}^2_3 \text{ trans}}\Bigg(\prod_{i=1}^3[T_\delta f_{\mathfrak{q}^2_i}](x_0)\Bigg)^{\frac{2}{3}}.\end{aligned}$$

Using Lemma 2.13 again, we have, for $x\in\mathfrak{Q}^2$,

$$\begin{aligned}|T_\delta f_{\mathfrak{q}^1_1}(x) T_\delta f_{\mathfrak{q}^1_2}(x)| &\le \sigma_2^d\big(\max_{\mathfrak{q}^2}|[T_\delta f_{\mathfrak{q}^2}]|(x)\big)^2\\ &\quad + \Bigg|\sum_{\substack{(\mathfrak{q}^2_1,\mathfrak{q}^2_2)\in\Lambda_1\times\Lambda_2\\ \mathfrak{q}^2_1,\mathfrak{q}^2_2\in\mathfrak{N}}} T_\delta f_{\mathfrak{q}^2_1}(x) T_\delta f_{\mathfrak{q}^2_2}(x)\Bigg| + \sigma_2^{-C}\sum_{\mathfrak{q}^2_1,\mathfrak{q}^2_2,\mathfrak{q}^2_3 \text{ trans}}\Bigg(\prod_{i=1}^3|[T_\delta f_{\mathfrak{q}^2_i}]|(x)\Bigg)^{\frac{2}{3}}.\end{aligned} \tag{52}$$

Taking $L^{\frac{p}{2}}$ on both sides of the inequality over each $\mathfrak{Q}^2$, summing along $\mathfrak{Q}^2$, and using Proposition 2.5 and Lemma 2.6, we get

$$\begin{aligned}\|T_\delta f_{\mathfrak{q}^1_1} T_\delta f_{\mathfrak{q}^1_2}\|_{\frac{p}{2}} &\lesssim (A(\sigma_2^{-2}\delta))^2\|f\|_p^2 + \Bigg(\sum_{\mathfrak{Q}^2}\Bigg\|\sum_{\mathfrak{q}^2_1,\mathfrak{q}^2_2\subset[\mathfrak{N}](\mathfrak{Q}^2,\mathfrak{q}^1_1,\mathfrak{q}^1_2)} T_\delta f_{\mathfrak{q}^2_1} T_\delta f_{\mathfrak{q}^2_2}\Bigg\|_{L^{p/2}(\mathfrak{Q}^2)}^{\frac{p}{2}}\Bigg)^{\frac{2}{p}}\\ &\quad + \sigma_2^{-C}\sup_{\tau_1,\tau_2,\tau_3}\max_{\mathfrak{q}^2_1,\mathfrak{q}^2_2,\mathfrak{q}^2_3 \text{ trans}}\|T_\delta(\tau_1 f_{\mathfrak{q}^2_1})T_\delta(\tau_2 f_{\mathfrak{q}^2_2})T_\delta(\tau_3 f_{\mathfrak{q}^2_3})\|_{\frac{p}{3}}^{\frac{2}{3}},\end{aligned} \tag{53}$$

where $[\mathfrak{N}](\mathfrak{Q}^2, \mathfrak{q}_1^1, \mathfrak{q}_2^1)$ is a subset of $\mathfrak{N}(\mathfrak{Q}^2, \mathfrak{q}_1^1, \mathfrak{q}_2^1)$. Here, for simplicity we now denote $\tau_{l_i} f$ by $\tau_i f$ just to indicate translation by a vector. The precise value of l_i is not significant in the overall argument. To show (53), for the first term in the right-hand side of (52) we may repeat the same argument as in (42). In fact, by (33) and the rapid decay of A_l[†] combined with Hölder's inequality to summation along l, l', and using Proposition 2.5 and Lemma 2.6 we have

$$\Big\|\max_{\mathfrak{q}^2} |[T_\delta f_{\mathfrak{q}^2}]|\Big\|_p \lesssim \sup_{\tau_2} \Big\|\max_{\mathfrak{q}^2} |T_\delta(\tau_2 f_{\mathfrak{q}^2})|\Big\|_p \lesssim A(\sigma_1^{-2}\delta)\|f\|_p.$$

For the third term of the right-hand side of (52), thanks to (33) and the rapid decay of A_l, it is enough to note that there are as many as $O(\sigma_2^{-C})$ transversal $\mathfrak{q}_1^2, \mathfrak{q}_2^2, \mathfrak{q}_3^2$.

We combine (53) with (43) to get

$$\begin{aligned} \|T_\delta f\|_p &\lesssim A(\sigma_1^{-2}\delta)\|f\|_p + \sigma_1^{-C} A(\sigma_2^{-2}\delta)\|f\|_p \\ &+ \sigma_1^{-C} \sup_{\tau_1,\tau_2} \max_{\mathfrak{q}_1^1,\mathfrak{q}_2^1 \text{ trans}} \Bigg(\sum_{\mathfrak{Q}^2} \Bigg\| \sum_{\substack{\mathfrak{q}_1^2 \subset \mathfrak{q}_1^1,\ \mathfrak{q}_2^2 \subset \mathfrak{q}_2^1 \\ \mathfrak{q}_1^2, \mathfrak{q}_2^2 \subset [\mathfrak{N}](\mathfrak{Q}^2, \mathfrak{q}_1^1, \mathfrak{q}_2^1)}} T_\delta(\tau_1 f_{\mathfrak{q}_1^2}) T_\delta(\tau_2 f_{\mathfrak{q}_2^2}) \Bigg\|_{L^{p/2}(\mathfrak{Q}^2)}^{\frac{p}{2}} \Bigg)^{\frac{1}{p}} \\ &+ \sigma_1^{-C}\sigma_2^{-C} \sup_{\tau_1,\tau_2,\tau_3} \max_{\mathfrak{q}_1^2,\mathfrak{q}_2^2,\mathfrak{q}_3^2 \text{ trans}} \|T_\delta(\tau_1 f_{\mathfrak{q}_1^2}) T_\delta(\tau_2 f_{\mathfrak{q}_2^2}) T_\delta(\tau_3 f_{\mathfrak{q}_3^2})\|_{\frac{p}{3}}^{\frac{1}{3}}. \end{aligned} \tag{54}$$

Here $[\mathfrak{N}](\mathfrak{Q}^2, \mathfrak{q}_1^1, \mathfrak{q}_2^1)$ also depends on τ_1, τ_2. We keep decomposing the trilinear transversal part in order to achieve a higher level of multilinearity.

2E4. *From k-transversal to $(k+1)$-transversal.* Now we proceed inductively. Suppose that we are given dyadic cubes $\mathfrak{q}_1^{k-1}, \mathfrak{q}_2^{k-1}, \dots, \mathfrak{q}_k^{k-1}$ of $(k-1)$-th scale which are transversal:

$$\mathrm{Vol}(\boldsymbol{n}(\xi_1), \boldsymbol{n}(\xi_2), \dots, \boldsymbol{n}(\xi_k)) \ge c\sigma_1\sigma_2\cdots\sigma_{k-1} \tag{55}$$

whenever $\xi_i \in \mathfrak{q}_i^{k-1}$, $i = 1, \dots, k$. As before, we denote by $\{\mathfrak{q}_i^k\}$ the collection of dyadic cubes of side length $2\sigma_k$ contained in $\mathfrak{q}_i^{k-1}$ such that

$$\bigcup_{\mathfrak{q}_i^k} \mathfrak{q}_i^k = \mathfrak{q}_i^{k-1}, \quad i = 1, \dots, k, \tag{56}$$

and we also denote by $\{\mathfrak{q}^k\}$ the set $\bigcup_{i=1}^k \{\mathfrak{q}_i^k\}$. Hence,

$$\prod_{i=1}^k T_\delta f_{\mathfrak{q}_i^{k-1}} = \prod_{i=1}^k \Bigg(\sum_{\mathfrak{q}_i^k} T_\delta f_{\mathfrak{q}_i^k} \Bigg) = \sum_{\mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k} \prod_{i=1}^k (T_\delta f_{\mathfrak{q}_i^k}). \tag{57}$$

Fix a dyadic cube $\mathfrak{Q}^k$ of side length $2M_i$ and let x_0 be the center of $\mathfrak{Q}^k$. For $i = 1, \dots, k$, let $\mathfrak{q}_{i*}^k \in \{\mathfrak{q}_i^k\}$ be such that

$$[T_\delta f_{\mathfrak{q}_{i*}^k}](x_0) = \max_{\mathfrak{q}_i^k} [T_\delta f_{\mathfrak{q}_i^k}](x_0)$$

[†]Note that the sequence is independent of $\mathfrak{Q}^2$.

and we set, for $i = 1, \dots, k$,

$$\Lambda_i^k = \{\mathfrak{q}_i^k : [T_\delta f_{\mathfrak{q}_i^k}](x_0) \ge (\sigma_k)^{kd} \max_{i=1,\dots,k} [T_\delta f_{\mathfrak{q}_{i*}^k}](x_0)\}.$$

Then, it follows that

$$\sum_{(\mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k)\notin\prod_{i=1}^k \Lambda_i^k} \prod_{i=1}^k [T_\delta f_{\mathfrak{q}_i^k}](x_0) \le \max [T_\delta f_{\mathfrak{q}^k}](x_0). \tag{58}$$

Let $\boldsymbol{n}_1, \dots, \boldsymbol{n}_k$ denote the normal vectors $\boldsymbol{n}(\boldsymbol{c}(\mathfrak{q}_{1*}^k)), \dots, \boldsymbol{n}(\boldsymbol{c}(\mathfrak{q}_{k*}^k))$, respectively, and let

$$\Pi^k = \Pi^k(\mathfrak{Q}^k, \mathfrak{q}_1^{k-1}, \mathfrak{q}_2^{k-1}, \dots, \mathfrak{q}_k^{k-1})$$

be the k-plane spanned by $\boldsymbol{n}_1, \dots, \boldsymbol{n}_k$. Now, for a sufficiently large constant $C > 0$, we define

$$\mathfrak{N} = \mathfrak{N}(\mathfrak{Q}^k, \mathfrak{q}_1^{k-1}, \mathfrak{q}_2^{k-1}, \dots, \mathfrak{q}_k^{k-1}) = \{\mathfrak{q}^k : \operatorname{dist}(\boldsymbol{n}(\mathfrak{q}^k), \Pi^k) \le C\sigma_k\}. \tag{59}$$

By (55) it follows that if $\mathfrak{q}_i^k \notin \mathfrak{N}$, (39) holds whenever $\xi_1 \in \mathfrak{q}_{1*}^k, \dots, \xi_k \in \mathfrak{q}_{k*}^k$ and $\xi_{k+1} \in \mathfrak{q}_i^k$. Hence, $\mathfrak{q}_{1*}^k, \dots, \mathfrak{q}_{k*}^k, \mathfrak{q}_i^k$ are transversal.

We write

$$\sum_{(\mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k)\in\prod_{i=1}^k \Lambda_i^k} \prod_{i=1}^k T_\delta f_{\mathfrak{q}_i^k} = \sum_{\substack{(\mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k)\in\prod_{i=1}^k \Lambda_i^k \\ \mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k\in\mathfrak{N}}} \prod_{i=1}^k T_\delta f_{\mathfrak{q}_i^k} + \sum_{\substack{(\mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k)\in\prod_{i=1}^k \Lambda_i^k \\ \mathfrak{q}_i^k\notin\mathfrak{N} \text{ for some } i}} \prod_{i=1}^k T_\delta f_{\mathfrak{q}_i^k}. \tag{60}$$

Consider a k-tuple $(\mathfrak{q}_1^k, \dots, \mathfrak{q}_k^k)$ which appears in the second sum. There is a $\mathfrak{q}_i^k \notin \mathfrak{N}$. By the same manipulation as before, we get

$$\prod_{i=1}^k [T_\delta f_{\mathfrak{q}_i^k}](x_0) \le \sigma_k^{-\frac{dk^2}{k+1}} \prod_{i=1}^k ([T_\delta f_{\mathfrak{q}_{i*}^k}](x_0))^{\frac{k}{k+1}} ([T_\delta f_{\mathfrak{q}_i^k}](x_0))^{\frac{k}{k+1}}.$$

Since $\mathfrak{q}_{1*}^k, \dots, \mathfrak{q}_{k*}^k, \mathfrak{q}_i^k$ are transversal, by Lemma 2.13 we have, for $x \in \mathfrak{Q}^k$,

$$\Bigg|\sum_{\substack{(\mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k)\in\prod_{i=1}^k \Lambda_i^k \\ \mathfrak{q}_i^k\notin\mathfrak{N} \text{ for some } i}} \prod_{i=1}^k T_\delta f_{\mathfrak{q}_i^k}(x)\Bigg| \lesssim \sigma_k^{-C} \sum_{\mathfrak{q}_1^k,\dots,\mathfrak{q}_{k+1}^k \text{ trans}} \prod_{i=1}^{k+1} ([T_\delta f_{\mathfrak{q}_i^k}](x_0))^{\frac{k}{k+1}}. \tag{61}$$

Combining (58) and (61) with (57) and (60), and applying Lemma 2.13 yield, for $x \in \mathfrak{Q}^k$,

$$\Bigg|\prod_{i=1}^k T_\delta f_{\mathfrak{q}_i^{k-1}}(x)\Bigg|$$
$$\lesssim \big(\max_{\mathfrak{q}^k} |[T_\delta f_{\mathfrak{q}^k}]|(x)\big)^k + \sigma_k^{-C} \sum_{\mathfrak{q}_1^k,\dots,\mathfrak{q}_{k+1}^k \text{ trans}} \prod_{i=1}^{k+1} \big(|[T_\delta f_{\mathfrak{q}_i^k}]|(x)\big)^{\frac{k}{k+1}} + \Bigg|\sum_{\substack{\mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k\in \\ [\mathfrak{N}](\mathfrak{Q}^k,\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1})}} \prod_{i=1}^k T_\delta f_{\mathfrak{q}_i^k}(x)\Bigg|,$$

where $[\mathfrak{N}](\mathfrak{Q}^k, \mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1})$ is a subset of $\mathfrak{N}(\mathfrak{Q}^k, \mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1})$. After taking the (p/k)-th power of both sides of the inequality, we integrate on $\mathbb{R}^d$, and use Proposition 2.5 and Lemma 2.6 along with (33) to get

$$\Big\| \prod_{i=1}^{k} T_\delta f_{\mathfrak{q}_i^{k-1}}(x) \Big\|_{L^{p/k}}^{\frac{1}{k}} \lesssim A(\sigma_k^{-2}\delta)\|f\|_p + \sigma_k^{-C} \sup_{\tau_1,\dots,\tau_{k+1}} \max_{\mathfrak{q}_1^k,\dots,\mathfrak{q}_{k+1}^k \text{ trans}} \Big\| \prod_{i=1}^{k+1} T_\delta(\tau_i f_{\mathfrak{q}_i^k}) \Big\|_{L^{p/(k+1)}}^{\frac{1}{k+1}}$$
$$+ \Big(\sum_{\mathfrak{Q}^k} \Big\| \sum_{\substack{\mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k \in \\ [\mathfrak{N}](\mathfrak{Q}^k, \mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1})}} \prod_{i=1}^{k} T_\delta f_{\mathfrak{q}_i^k} \Big\|_{L^{p/k}(\mathfrak{Q}^k)}^{\frac{p}{k}} \Big)^{\frac{1}{p}}. \tag{62}$$

2E5. *Multiscale decomposition.* For $k = 2, \dots, d-1$, let us set

$$\mathfrak{M}^k f = \sup_{\tau_1,\dots,\tau_k} \max_{\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1} \text{ trans}} \Big(\sum_{\mathfrak{Q}^k} \Big\| \sum_{\substack{\mathfrak{q}_i^k \subset \mathfrak{q}_i^{k-1} \\ \mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k \in [\mathfrak{N}](\mathfrak{Q}^k)}} \prod_{i=1}^{k} T_\delta(\tau_i f_{\mathfrak{q}_i^k}) \Big\|_{L^{p/k}(\mathfrak{Q}^k)}^{\frac{p}{k}} \Big)^{\frac{1}{p}}.$$

Here $[\mathfrak{N}](\mathfrak{Q}^k)$ depends on $\tau_1, \dots, \tau_k$, and $\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}$, but $\boldsymbol{n}(\mathfrak{q}^k)$, $\mathfrak{q}^k \in [\mathfrak{N}](\mathfrak{Q}^k)$, is contained in a k-plan. Starting from (54) we iteratively apply (62) to the transversal products to get

$$\|T_\delta f\|_p \lesssim \sum_{k=1}^{m} \sigma_{k-1}^{-C} A(\sigma_k^{-2}\delta)\|f\|_p$$
$$+ \sum_{k=2}^{m} \sigma_{k-1}^{-C} \mathfrak{M}^k f + \sigma_l^{-C} \sup_{\tau_1,\dots,\tau_{m+1}} \max_{\mathfrak{q}_1^m,\dots\mathfrak{q}_{m+1}^m \text{ trans}} \Big\| \prod_{i=1}^{m+1} T_\delta \tau_i f_{\mathfrak{q}_i^m} \Big\|_{L^{p/(m+1)}}^{\frac{1}{m+1}}. \tag{63}$$

2F. ***Proof of Proposition 2.3.*** For given $\beta > 0$, we need to show that $\mathcal{A}^\beta(s) \le C$ for $0 < s \le 1$ if $p \ge p_\circ(d)$. Let $\varepsilon > 0$ be small enough such that $(100d)^{-1}\beta \ge \varepsilon$, and choose a small $\varepsilon_\circ > 0$ and $N = N(\varepsilon)$ large enough such that Proposition 2.10 and Corollary 2.12 hold uniformly for $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$.

Let $0 < s < \delta \le 1$, and let $\sigma_1, \dots, \sigma_m$ be dyadic numbers satisfying (36). Since $A(\delta) \le C$ for $\delta \gtrsim 1$ and $s \le \sigma_k^{-2}\delta$, we see

$$A(\sigma_k^{-2}\delta) \le A(\sigma_k^{-2}\delta)\chi_{(0,10^{-2}]}(\sigma_k^{-2}\delta) + C \le (\sigma_k^{-2}\delta)^{-\frac{d-1}{2}+\frac{d}{p}-\beta}\mathcal{A}^\beta(s) + C. \tag{64}$$

By Proposition 2.10 and Lemma 2.6 we have, for $p \ge 2(m+1)/m$,

$$\sup_{\tau_1,\dots,\tau_{m+1}} \max_{\mathfrak{q}_1^m,\dots,\mathfrak{q}_{m+1}^m \text{ trans}} \Big\| \prod_{i=1}^{m+1} T_\delta \tau_i f_{\mathfrak{q}_i^m} \Big\|_{L^{p/(m+1)}}^{\frac{1}{m+1}} \lesssim (\sigma_1 \cdots \sigma_m)^{-C_\varepsilon} \delta^{-\varepsilon} \delta^{\frac{d}{p}-\frac{d-1}{2}} \|f\|_p, \tag{65}$$

which uniformly holds for $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$.

We have two types of estimates for $\mathfrak{M}^k f$. Since $\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}$ are already transversal,

$$\Big| \sum_{\substack{\mathfrak{q}_i^k \subset \mathfrak{q}_i^{k-1} \\ \mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k \subset [\mathfrak{N}](\mathfrak{Q}^k)}} \prod_{i=1}^{k} T_\delta(\tau_i f_{\mathfrak{q}_i^k}) \Big| \le \sum_{\mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k \text{ trans}} \Big| \prod_{i=1}^{k} T_\delta(\tau_i f_{\mathfrak{q}_i^k}) \Big|.$$

Here we slightly abuse the definition "trans" and "$\mathfrak{q}_1^k, \dots, \mathfrak{q}_k^k$ trans" means (55) holds if $\xi_i \in \mathfrak{q}_i^k$, $i = 1, \dots, k$. Since there are as many as $O(\sigma_k^{-C})$ k-tuples $(\mathfrak{q}_1^k, \dots, \mathfrak{q}_k^k)$ and the above inequality holds regardless of $\mathfrak{Q}^k$, we get

$$\mathfrak{M}^k f \lesssim \sigma_k^{-C} \sup_{\tau_1,\dots,\tau_k} \max_{\mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k \text{ trans}} \Big\| \prod_{i=1}^k T_\delta(\tau_i f_{\mathfrak{q}_i^k}) \Big\|_{\frac{p}{k}}^{\frac{1}{k}}.$$

Since $\mathfrak{q}_1^k, \dots, \mathfrak{q}_k^k$ are transversal, by Proposition 2.10 (also see Remark 2.8) and Lemma 2.6, we get, for $p \ge 2k/(k-1)$,

$$\Big\| \prod_{i=1}^k T_\delta(\tau_i f_{\mathfrak{q}_i^k}) \Big\|_{\frac{p}{k}}^{\frac{1}{k}} \lesssim (\sigma_1 \cdots \sigma_{k-1})^{-C_\varepsilon} \delta^{\frac{d}{p} - \frac{d-1}{2} - \varepsilon} \prod_{i=1}^k \|\tau_i f_{\mathfrak{q}_i^k}\|_p^{\frac{1}{k}} \lesssim \sigma_k^{-C_\varepsilon} \delta^{\frac{d}{p} - \frac{d-1}{2} - \varepsilon} \|f\|_p.$$

Hence, for $p \ge 2k/(k-1)$, we have the uniform estimate for $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$

$$\mathfrak{M}^k f \lesssim \sigma_k^{-C} \delta^{\frac{d}{p} - \frac{d-1}{2} - \varepsilon} \|f\|_p. \tag{66}$$

On the other hand, fixing $\tau_1, \dots, \tau_k$, $\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}$ trans, and $\mathfrak{Q}^k$, we consider the integrals appearing in the definition of $\mathfrak{M}^k f$. Let us write $\mathfrak{Q}^k = \mathfrak{q}(z, 1/\sigma_k)$. Using Corollary 2.12, for $2 \le p \le 2k/(k-1)$, we have

$$\Big\| \sum_{\substack{\mathfrak{q}_i^k \subset \mathfrak{q}_i^{k-1} \\ \mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k \in [\mathfrak{N}](\mathfrak{Q}^k)}} \prod_{i=1}^k T_\delta(\tau_i f_{\mathfrak{q}_i^k}) \Big\|_{L^{p/k}(\mathfrak{Q}^k)} \lesssim \sigma_{k-1}^{-C_\varepsilon} \sigma_k^{-\varepsilon} \prod_{i=1}^k \Big\| \Big(\sum_{\mathfrak{q}_i^k \in [\mathfrak{N}](\mathfrak{Q}^k)} |T_\delta(\tau_i f_{\mathfrak{q}_i^k})|^2 \Big)^{\frac{1}{2}} \rho_{B(z, \frac{C}{\sigma_k})} \Big\|_p. \tag{67}$$

Since $[\mathfrak{N}](\mathfrak{Q}^k) \subset \mathfrak{N}(\mathfrak{Q}^k, \mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1})$, it is clear that if $\mathfrak{q}_i^k \in [\mathfrak{N}](\mathfrak{Q}^k)$, then $\mathfrak{q}_i^k \subset \mathrm{N}^{-1}(\Pi) + O(\sigma_k)$ for a k-plane Π. Since $\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}$ are transversal and $\mathfrak{q}_i^k \subset \mathfrak{q}_i^{k-1}$, $i = 1, \dots, k$, we know $\sum_{\mathfrak{q}_1^k \in \mathfrak{N}(\mathfrak{Q}^k)} T_\delta(\tau_1 f_{\mathfrak{q}_1^k})$, $\dots$, $\sum_{\mathfrak{q}_k^k \in \mathfrak{N}(\mathfrak{Q}^k)} T_\delta(\tau_k f_{\mathfrak{q}_k^k})$ satisfy the assumptions of Corollary 2.12 (Proposition 2.11) with $\delta = \sigma_k$ and $\sigma = \sigma_1 \cdots \sigma_{k-1}$. Hence, Corollary 2.12 gives (67).

Recalling that the $\mathfrak{q}_i^k$ are contained in a $C\sigma_k$-neighborhood of $\Gamma(\psi)$, we see that $\#\mathfrak{N}(\mathfrak{Q}^k)$ is $\lesssim \sigma_k^{1-k}$. So, by Hölder's inequality we have

$$\Big\| \sum_{\substack{\mathfrak{q}_i^k \subset \mathfrak{q}_i^{k-1} \\ \mathfrak{q}_1^k,\dots,\mathfrak{q}_k^k \subset \mathfrak{N}(\mathfrak{Q}^k)}} \prod_{i=1}^k T_\delta(\tau_i f_{\mathfrak{q}_i^k}) \Big\|_{L^{p/k}(\mathfrak{Q}^k)}^{\frac{1}{k}} \lesssim \sigma_{k-1}^{-C} \sigma_k^{-\varepsilon - p(k-1)(\frac{1}{2} - \frac{1}{p})} \max_{1 \le i \le k} \Big\| \Big(\sum_{\mathfrak{q}^k} |T_\delta(\tau_i f_{\mathfrak{q}^k})|^p \Big)^{\frac{1}{p}} \rho_{B(z, \frac{C}{\sigma_k})} \Big\|_p.$$

Here we bound $\sigma_1, \dots, \sigma_{k-1}$ with σ_{k-1} using (36) and replace C_ε with a larger constant C, since ε is fixed. By using the rapid decay of ρ we sum the estimates along $\mathfrak{Q}^k$ to get

$$\mathfrak{M}^k f \lesssim \sigma_{k-1}^{-C} \sigma_k^{-\varepsilon - (k-1)(\frac{1}{2} - \frac{1}{p})} \sup_h \Big\| \Big(\sum_{\mathfrak{q}^k} |T_\delta(\tau_h f_{\mathfrak{q}^k})|^p \Big)^{\frac{1}{p}} \Big\|_p. \tag{68}$$

By Proposition 2.5, Lemma 2.6, and (64) we get, for $2 \le p \le 2k/(k-1)$,

$$\mathfrak{M}^k f \lesssim \big(\sigma_{k-1}^{-C} \sigma_k^{\beta + \frac{2d-k-1}{2} - \frac{2d-k+1}{p}} \delta^{-\frac{d-1}{2} + \frac{d}{p} - \beta} \mathcal{A}^\beta(s) + \sigma_k^{-C} \big) \|f\|_p.$$

Here we also use $(100d)^{-1}\beta \ge \varepsilon$. So, if $p \ge 2(2d-k+1)/(2d-k-1)$,

$$\mathfrak{M}^k f \lesssim (\sigma_{k-1}^{-C}\sigma_k^{\alpha}\delta^{-\frac{d-1}{2}+\frac{d}{p}-\beta}\mathcal{A}^{\beta}(s)+\sigma_k^{-C})\|f\|_p$$

for some $\alpha>0$. Combining this with (66), we have for some $\alpha>0$

$$\mathfrak{M}^k f \lesssim (\sigma_k^{-C}\delta^{-\frac{d-1}{2}+\frac{d}{p}-\varepsilon}+\sigma_{k-1}^{-C}\sigma_k^{\alpha}\delta^{-\frac{d-1}{2}+\frac{d}{p}-\beta}\mathcal{A}^{\beta}(s)+\sigma_k^{-C})\|f\|_p$$

provided that

$$p \ge \min\Big(\frac{2(2d-k+1)}{2d-k-1},\frac{2k}{k-1}\Big).$$

Since $(100d)^{-1}\beta \ge \varepsilon$ and $p_\circ > 2d/(d-1)$, from (64) we note that $A(\sigma_k^{-2}\delta) \lesssim \sigma_k^{\alpha}\delta^{-\frac{d-1}{2}+\frac{d}{p}-\beta}\mathcal{A}^{\beta}(s)$. Thus, by (63), the above inequality, (64), and (65) we obtain

$$\|T_\delta f\|_p \lesssim \sum_{k=1}^{m}(\sigma_{k-1}^{-C}\sigma_k^{\alpha}\mathcal{A}^{\beta}(s)+\sigma_k^{-C})\delta^{-\frac{d-1}{2}+\frac{d}{p}-\beta}\|f\|_p+\sigma_m^{-C}\delta^{-\frac{d-1}{2}+\frac{d}{p}-\beta}\|f\|_p \tag{69}$$

for some $\alpha>0$ provided that

$$p \ge \min\Big(\frac{2(2d-k+1)}{2d-k-1},\frac{2k}{k-1}\Big),\quad k=2,\dots,m \ \text{ and } \ p \ge \frac{2(m+1)}{m}. \tag{70}$$

Since the estimates (65)–(68) hold uniformly for $\psi \in \mathfrak{G}(\varepsilon_\circ, N)$, so does (69). Taking supremum along ψ and f, we have

$$A(\delta) \le \Big(\sum_{k=1}^{m} C\sigma_{k-1}^{-C}\sigma_k^{\alpha}\mathcal{A}^{\beta}(s)+C\sigma_m^{-C}\Big)\delta^{-\frac{d-1}{2}+\frac{d}{p}-\beta}.$$

By multiplying $\delta^{\frac{d-1}{2}-\frac{d}{p}-\beta}$ to both sides, $\delta^{\frac{d-1}{2}-\frac{d}{p}+\beta}A(\delta) \le \sum_{k=1}^{m} C\sigma_{k-1}^{-C}\sigma_k^{\alpha}\mathcal{A}^{\beta}(s)+C\sigma_m^{-C}$. This is valid as long as $s<\delta\le 1$. Hence, taking supremum for $s<\delta\le 1$ yields

$$\mathcal{A}^{\beta}(s) \le \sum_{k=1}^{m} C\sigma_{k-1}^{-C}\sigma_k^{\alpha}\mathcal{A}^{\beta}(s)+C\sigma_m^{-C}$$

if (70) is satisfied. Therefore, choosing $\sigma_1 \ll \cdots \ll \sigma_m$, successively, we can make $\sum_{k=1}^{m} C\sigma_{k-1}^{-C}\sigma_k^{\alpha} \le \frac{1}{2}$. This gives the desired $\mathcal{A}^{\beta}(s) \le C\sigma_m^{-C}$ provided that (70) holds.

Finally, we only need to check that the minimum of

$$\mathcal{P}(m)=\max\Big(\frac{2(m+1)}{m},\max_{k=2,\dots,m}\min\Big(\frac{2(2d-k+1)}{2d-k-1},\frac{2k}{k-1}\Big)\Big),\quad 2\le m\le d-1$$

is $p_\circ(d)$ as can be done by routine computation. □

Remark 2.15. The minimum of $\mathcal{P}$ is achieved when m is near $2d/3$. So, it doesn't seem that the argument makes use of the full strength of the multilinear restriction estimates.

3. Square function estimates

In this section we prove Theorem 1.2. We firstly obtain multi(sub)linear square-function estimates which are vector-valued extensions of multilinear restriction estimates. Then, we modify the argument in Section 2F to obtain the sharp square-function estimate from these multilinear estimates. Although the basic strategy here is similar to the one in the previous section, due to the additional integration in t we need to handle a family of surfaces. This argument in this section is very much in parallel with that of the previous section.

3A. ***One-parameter family of elliptic functions.*** As before, for $0 < \varepsilon_\circ \ll \frac{1}{2}$ and an integer $N \geq 100d$, we denote by $\overline{\mathfrak{G}}(\varepsilon_\circ, N)$ the class of smooth functions defined on $I^{d-1} \times I$ which satisfy

$$\|\psi - \psi_\circ - t\|_{C^N(I^{d-1}\times I)} \leq \varepsilon_\circ. \tag{71}$$

This clearly implies that, for all $(x,t) \in I^{d-1} \times I$,

$$\partial_t \psi(x,t) \in [1-\varepsilon_\circ, 1+\varepsilon_\circ]. \tag{72}$$

For $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$ and $z_0 = (\zeta_0, t_0) \in \frac{1}{2} I^d$, define

$$\psi^{\varepsilon}_{z_0}(\zeta,t) = \varepsilon^{-2}\biggl(\psi\biggl(\zeta_0 + \varepsilon \mathcal{H}^{\psi}_{z_0}\zeta, t_0 + \frac{\varepsilon^2 t}{\partial_t \psi(z_0)}\biggr) - \psi(z_0) - \varepsilon \nabla_\zeta \psi(z_0)\mathcal{H}^{\psi}_{z_0}\zeta\biggr),$$

where $\mathcal{H}^{\psi}_{z_0} = (\sqrt{H(\psi(\cdot, t_0))(\zeta_0)})^{-1}$. Then we have the following.

Lemma 3.1. *Let $z_0 \in \frac{1}{2} I^d$ and $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$. There is a $\kappa = \kappa(\varepsilon_\circ, N) > 0$, independent of ψ, ζ_0, t_0, such that $\psi^{\varepsilon}_{z_0}$ is contained in $\overline{\mathfrak{G}}(\varepsilon_\circ, N)$ if $0 < \varepsilon \leq \kappa$.*

Proof. It is sufficient to show that $|\partial^\alpha_\zeta \partial^\beta_t (\psi^\varepsilon_{z_0}(\zeta,t) - \psi_\circ(\zeta) - t)| \leq C\varepsilon$, with C independent of $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$, if $|\alpha| + \beta \leq N$ and $(\zeta, t) \in I^d$.

Let $0 < \varepsilon \leq \frac{1}{4}$. If $(\zeta,t) \in I^d$ and $|\alpha| + 2\beta > 2$, trivially $|\partial^\alpha_\zeta \partial^\beta_t (\psi^\varepsilon_{z_0}(\zeta,t) - \psi_\circ(\zeta,t) - t)| \leq C\varepsilon$ because $z_0 = (\zeta_0, t_0) \in \frac{1}{2} I^d$. Thus, it is sufficient to consider the cases $\beta = 1,\ |\alpha| = 0$ and $\beta = 0,\ 0 \leq |\alpha| \leq 2$. The first case is easy to handle. Indeed, from Taylor's theorem and (72)

$$\partial_t(\psi^\varepsilon_{z_0}(\zeta,t) - \psi_\circ - t) = (\partial_t \psi(z_0))^{-1}\biggl(\partial_t \psi\biggl(\zeta_0 + \varepsilon \mathcal{H}^{\psi}_{z_0}\zeta, t_0 + \frac{\varepsilon^2 t}{\partial_t \psi(z_0)}\biggr) - \partial_t \psi(z_0)\biggr) = O(\varepsilon).$$

To handle the second case, we consider Taylor's expansion of ψ in t with integral remainder:

$$\psi(\zeta,t) = \psi(\zeta,t_0) + \partial_t \psi(\zeta,t_0)(t-t_0) + R_1(\zeta,t),$$

where

$$R_1(\zeta,t) = (t-t_0)^2 \int_0^1 (1-s)\partial_t^2 \psi(\zeta, (t-t_0)s + t_0)\, ds.$$

The change of variables $t \to t_0 + \varepsilon^2(\partial_t \psi(z_0))^{-1} t,\ \zeta \to \zeta_0 + \varepsilon \mathcal{H}^{\psi}_{z_0}\zeta$ gives

$$\psi\biggl(\zeta_0 + \varepsilon \mathcal{H}^{\psi}_{z_0}\zeta, t_0 + \frac{\varepsilon^2 t}{\partial_t \psi(z_0)}\biggr) = \varepsilon^2 \psi(\cdot, t_0)^\varepsilon_{\zeta_0}(\zeta) + \psi(z_0) + \varepsilon \nabla_\zeta \psi(z_0)\mathcal{H}^{\psi}_{z_0}\zeta + \frac{\varepsilon^2 \partial_t \psi(\zeta_0 + \varepsilon \mathcal{H}^{\psi}_{z_0}\zeta, t_0)}{\partial_t \psi(z_0)} t + \widetilde{R}(\zeta,t),$$

where $\psi(\,\cdot\,,t_0)^{\varepsilon}_{\zeta_0}$ is defined by (8) and $\widetilde{R}(\zeta,t)=R_1(\zeta_0+\varepsilon\mathcal{H}^{\psi}_{z_0}\zeta,t_0+\varepsilon^2(\partial_t\psi(z_0))^{-1}t)$. Hence, it follows that

$$\psi^{\varepsilon}_{z_0}-\psi_\circ-t=\psi(\,\cdot\,,t_0)^{\varepsilon}_{\zeta_0}(\zeta)-\psi_\circ+\frac{\partial_t\psi(\zeta_0+\varepsilon\mathcal{H}^{\psi}_{z_0}\zeta,t_0)-\partial_t\psi(z_0)}{\partial_t\psi(z_0)}t+\varepsilon^{-2}\widetilde{R}(\zeta,t).$$

Since $\psi(\,\cdot\,,t_0)-t_0\in\mathfrak{G}(\varepsilon_\circ,N)$ and $(\psi(\,\cdot\,,t_0)-t_0)^{\varepsilon}_{\zeta_0}=\psi(\,\cdot\,,t_0)^{\varepsilon}_{\zeta_0}$, we have $|\partial^{\alpha}_{\zeta}(\psi(\,\cdot\,,t_0)^{\varepsilon}_{\zeta_0}-\psi_\circ)|\le C\varepsilon$ on I^d for $|\alpha|=0,1,2$ (similarly to the proof of Lemma 2.1). By (72) and the mean value theorem we also have $(\partial_t\psi(z_0))^{-1}\,\partial^{\alpha}_{\zeta}(\partial_t\psi(\zeta_0+\varepsilon\mathcal{H}^{\psi}_{z_0}\zeta,t_0)-\partial_t\psi(z_0))t=O(\varepsilon)$ in $C^N(I^{d-1})$ for $|\alpha|=0,1,2$. Note that

$$\varepsilon^{-2}\widetilde{R}(\zeta,t)=\frac{\varepsilon^2t^2}{(\partial_t\psi(z_0))^2}\int_0^1(1-s)\partial_t^2\psi\big(\zeta_0+\varepsilon\mathcal{H}^{\psi}_{z_0}\zeta,\varepsilon^2(\partial_t\psi(z_0))^{-1}ts+t_0\big)\,ds.$$

Thus, again by (72) it is easy to see that $\partial^{\alpha}_{\zeta}(\varepsilon^{-2}\widetilde{R})=O(\varepsilon^{2+|\alpha|})$ for any α. Therefore, combining the all together we have $|\partial^{\alpha}_{\zeta}(\psi^{\varepsilon}_{z_0}(\,\cdot\,,t)-\psi_\circ-t)|\le C\varepsilon$ on I^{d-1} for $|\alpha|=0,1,2$. □

3B. *Square function with localized frequency.* Abusing the conventional notation we denote by $m(D)f$ the multiplier operator given by $\widehat{m(D)f}(\xi)=m(\xi)\hat{f}(\xi)$, and we also write $D=(D',D_d)$ where D', D_d correspond to the frequency variables ζ, τ, respectively.

In order to show (4), by the Littlewood–Paley decomposition, scaling, and further finite decompositions, it is sufficient to show

$$\Big\|\Big(\int_{1-\varepsilon^2}^{1+\varepsilon^2}\Big|\frac{\partial}{\partial t}\mathcal{R}^{\alpha}_t f(x)\Big|^2\,dt\Big)^{\frac12}\Big\|_p\le C\|f\|_p$$

for some small $\varepsilon>0$. And by decomposing $\hat{f}$, which may now be assumed to be supported in $S^{d-1}+O(\varepsilon^2)$, and rotation we may assume $\hat{f}$ is supported in $B(-e_d,c\varepsilon^2)$ with some $c>0$. Hence, by discarding the harmless smooth multiplier, the matter reduces to showing

$$\big\|\,\|(D_d+\sqrt{t^2-|D'|^2})_+^{\alpha-1}f\|_{L^2_t(1-\varepsilon^2,1+\varepsilon^2)}\big\|_p\le C\|f\|_p.$$

By changing variables in the frequency domain, $D_d\to D_d+1$, $(D',D_d)\to(\varepsilon D',\varepsilon^2D_d)$ and $t\to\varepsilon^2t+1$, this is equivalent to

$$\big\|\,\|(D_d-\psi_{br}(D',t)))_+^{\alpha-1}\chi_\circ(D)f\|_{L^2_t(I)}\big\|_p\le C\|f\|_p,\tag{73}$$

where $\psi_{br}(\zeta,t)=\varepsilon^{-2}(1-\sqrt{1+2\varepsilon^2t+\varepsilon^4t^2-\varepsilon^2|\zeta|^2})$ and $\chi_\circ$ is a smooth function supported in a small neighborhood of the origin. Clearly, ψ_{br} satisfies (71) with $\varepsilon_\circ=C\varepsilon^2$ for some $C>0$. Consequently, we are led to consider general $\psi\in\overline{\mathfrak{G}}(\varepsilon_\circ,N)$ rather than the specific ψ_{br}.

Let us define the class $\mathcal{E}(N)$ of smooth functions by setting

$$\mathcal{E}(N)=\big\{\eta\in C^\infty(I^d\times I):\|\eta\|_{C^N(I^d\times I)}\le1,\ \tfrac12\le\eta\le1\big\}.$$

Let $\psi\in\overline{\mathfrak{G}}(\varepsilon_\circ,N)$ and $\eta\in\mathcal{E}(N)$. For $0<\delta$ and f with $\hat{f}$ supported in $\frac12I^d$, we define $S_\delta=S_\delta(\psi,\eta)$ by

$$S_\delta f(x)=\Big\|\phi\Big(\frac{\eta(D,t)(D_d-\psi(D',t))}{\delta}\Big)f\Big\|_{L^2_t(I)}.\tag{74}$$

Compared to ψ, the role of η is less significant but this enables us to handle more general square functions (in particular, see Remark 3.3). By dyadic decomposition away from the singularity, the matter of showing (73) is reduced to obtaining the sharp bound

$$\|S_\delta f\|_p \le C\delta^{\frac{d}{p}-\frac{d-2}{2}-\varepsilon}\|f\|_p, \quad \varepsilon > 0, \tag{75}$$

when $\hat{f}$ is supported in a small neighborhood of the origin. This is currently verified for $p \ge 2(d+2)/d$ [Lee et al. 2012] by making use of the bilinear restriction estimate for the elliptic surfaces. The following is our main result concerning the estimate (75).

Proposition 3.2. *Let $p_s = p_s(d)$ be given by (5) and* supp $\hat{f} \subset \frac{1}{2}I^d$. *If $p \ge \min(p_s(d), 2(d+2)/d)$ and $\varepsilon_\circ$ is sufficiently small, for $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that (75) holds uniformly for $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$, $\eta \in \mathcal{E}(N)$.*

Proof of Theorem 1.2. By choosing a small $\varepsilon > 0$ in the above, we can make ψ_{br} be in $\overline{\mathfrak{G}}(\varepsilon_\circ, N)$ for any ε_0 and N. Hence, Proposition 3.2 gives (75) for any $\varepsilon > 0$ if $p \ge \min(p_s(d), 2(d+2)/d)$. Hence, dyadic decomposition of the multiplier operator in (73) and using (75) followed by summation along dyadic pieces gives (73) for $\alpha > d/2 - d/p$. This proves Theorem 1.2. □

Remark 3.3. As has been shown before, for the proof of Theorem 1.2 it suffices to consider an operator which is defined without η, but by allowing η in (74) we can handle the square-function estimates for the operator $f \to \phi((1-|D|/t)/\delta)f$, which is closely related to smoothing estimates for the solutions to the Schrödinger and wave equations; for example, see [Lee et al. 2012]. In fact, Proposition 3.2 implies, for $\varepsilon > 0$,

$$\left\| \left(\int_{\frac{1}{2}}^{2} \left| \phi\left(\frac{1-|D|/t}{\delta} \right) f \right|^2 dt \right)^{\frac{1}{2}} \right\|_p \le \delta^{\frac{d}{2}-\frac{d}{p}-\varepsilon} C\|f\|_p \tag{76}$$

if $p \ge p_s(d)$. Indeed, by finite decompositions, rotation and scaling, as before, it is sufficient to consider the time average over the interval $I_\varepsilon = (1-\varepsilon^2, 1+\varepsilon^2)$ and we may assume that $\hat{f}$ is supported in $B(-e_d, c\varepsilon^2)$. Writing

$$1 - |\xi|/t = t^{-2}(t+|\xi|)^{-1}(\tau - \sqrt{t^2-|\zeta|^2})(\tau + \sqrt{t^2-|\zeta|^2})$$

for $\xi \in B(-e_d, c\varepsilon^2)$, the same change of variables $D_d \to D_d + 1$, $(D', D_d) \to (\varepsilon D', \varepsilon^2 D_d)$ and $t \to \varepsilon^2 t + 1$ transforms $\phi((1-|\xi|/t)/\delta)$ to

$$\phi\left(\frac{\eta(\xi,t)(\tau - \psi_{br})}{\varepsilon^{-2}\delta/2} \right)$$

with a smooth η which satisfies $\eta \in (1-c\varepsilon/2, 1+c\varepsilon/2)$. Hence, we now apply Proposition 3.2 with sufficiently small ε to get (76).

As before, in order to control the L^p norm of S_δ we define $B(\delta) = B_p(\delta)$ by

$$B(\delta) \equiv \sup\{\|S_\delta(\psi,\eta)f\|_{L^p} : \psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N),\ \eta \in \mathcal{E}(N),\ \|f\|_p \le 1,\ \text{supp } \hat{f} \subset \tfrac{1}{2}I^d\}.$$

As before, using Lemma 2.9 it is easy to see that $B(\delta) \le C$ if $\delta \ge 1$, and $B(\delta) \le C\delta^{-c}$ for some $c > 0$ otherwise (for example, see the paragraph below Proposition 3.6). We also define for $\beta > 0$ and $\delta \in (0, 1)$,

$$\mathcal{B}^\beta(\delta) = \mathcal{B}_p^\beta(\delta) \equiv \sup_{\delta < s \le 1} s^{\frac{d-2}{2} - \frac{d}{p} + \beta} B_p(s).$$

Thus, Theorem 1.2 follows if we show $\mathcal{B}^\beta(\delta) \le C$ for any $\beta > 0$. As observed in the previous section, the bound for $S_\delta f$ improves if the Fourier transform of f is contained in a set of smaller diameter. The following plays a crucial role in the induction argument (see Section 3F).

Proposition 3.4. *Let* $0 < \delta \ll 1$, $\psi \in \overline{\mathfrak{C}}(\varepsilon_\circ, N)$, *and* $\eta \in \mathcal{E}(N)$. *Suppose that* $\hat{f}$ *is supported in* $\mathfrak{q}(a, \varepsilon)$, $10\sqrt{\delta} \le \varepsilon \le \frac{1}{2}$, *and* $a \in \frac{1}{2} I^d$. *Then, if* $\varepsilon_\circ > 0$ *is small enough, there is a* $\kappa = \kappa(\varepsilon_\circ, N)$ *such that*

$$\|S_\delta(\psi, \eta) f\|_p \le C\varepsilon^{\frac{1}{p} + \frac{1}{2}} B_p(\varepsilon^{-2}\delta)\|f\|_p \tag{77}$$

holds with C *independent of* ψ, *and* ε, *whenever* $\varepsilon \le \kappa$.

Proof. By breaking the support of $\hat{f}$ into a finite number of dyadic cubes, we may assume that $\hat{f}$ is supported in $\mathfrak{q}(a, \nu\varepsilon)$ for a small constant $\nu > 0$ satisfying $\nu^2 d^2 \in [2^{-5}, 2^{-4})$. This only increases the bound by a constant multiple. Since $\hat{f}$ is supported in $\mathfrak{q}(a, \nu\varepsilon)$ and $a = (a', a_d) \in \frac{1}{2} I^d$, from (72) and the fact that $\frac{1}{2} \le \eta \le 1$, it is clear that $\phi\big(\eta(D, t)(D_d - \psi(D', t))/\delta\big) f \not\equiv 0$ for t contained in an interval $[\alpha, \beta]$ of length $\lesssim \nu\varepsilon$ because $\phi\big(\eta(\xi, t)(\tau - \psi(\zeta, t))/\delta\big)$ is supported in an $O(\delta)$-neighborhood of $\tau = \psi(\zeta, t)$.

Let $\alpha = t_0 < t_1 < \cdots < t_l = \beta$, $l \le O(\varepsilon^{-1})$, be such that $t_{k+1} - t_k \le \nu^2\varepsilon^2$. Since $\delta \le 10^{-2}\varepsilon^2$, by (71) and (72) it follows that if $t \in [t_k, t_{k+1}]$, then $\phi\big(\eta(\xi, t)(\tau - \psi(\zeta, t))/\delta\big) \hat{f}(\xi)$ is supported in the parallelepiped

$$\mathcal{P}_k = \big\{(\zeta, \tau) : \max_{i=1,\dots,d-1} |\zeta_i - a_i'| < \nu\varepsilon,\ |\tau - \psi(a', t_k) - \nabla_\zeta \psi(a', t_k)(\zeta - a')| \le 2d^2\nu^2\varepsilon^2\big\}.$$

This follows from Taylor's theorem since $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$. By (72) it is easy to see that $\{\mathcal{P}_k\}_{k=1}^l$ are overlapping boundedly. In fact, $\phi\big(\eta(\xi, t)(\tau - \psi(\zeta, t))/\delta\big) \hat{f}(\xi)$, $t \in [t_k,\ t_{k+1}]$, is supported in

$$\widetilde{\mathcal{P}}_k = \{\xi \in \mathfrak{q}(a, c\varepsilon) : |\tau - \psi(\zeta, t_k))| \le C\varepsilon^2\}, \quad k = 0, \dots, l-1,$$

with $C \ge 3d^2\nu^2\varepsilon^2$ and the $\{\widetilde{\mathcal{P}}_k\}$ are boundedly overlapping because of (72), and by Taylor's expansion it is easy to see that $\mathcal{P}_k \subset \widetilde{\mathcal{P}}_k$ because the second remainder is uniformly $O(\varepsilon^2)$ for $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$.

Let φ be a smooth function supported in $2I^d$ and $\varphi = 1$ on I^d. Let $L_{\mathcal{P}_k}$ be the affine map which bijectively maps $\mathcal{P}_k$ to I^d, and set $\varphi_{\mathcal{P}_k} = \varphi(L_{\mathcal{P}_k} \cdot)$ so that $\varphi_{\mathcal{P}_k}$ vanishes outside of $2\mathcal{P}_k$ and equals 1 on $\mathcal{P}_k$. Here $2\mathcal{P}_k$ denotes the parallelepiped which is given by dilating $\mathcal{P}_k$ twice from the center of $\mathcal{P}_k$. Then we have

$$(S_\delta f(x))^2 = \sum_k \int_{I_k} \bigg| \phi\bigg(\frac{\eta(D, t)(D_d - \psi(D', t))}{\delta} \bigg) \varphi_{\mathcal{P}_k}(D) f(x) \bigg|^2 dt.$$

Since $p \ge 2$, by Hölder's inequality it follows that

$$S_\delta f(x) \le C\varepsilon^{\frac{1}{p} - \frac{1}{2}} \bigg(\sum_k \bigg\| \phi\bigg(\frac{\eta(D, t)(D_d - \psi(D', t))}{\delta} \bigg) \varphi_{\mathcal{P}_k}(D) f(x) \bigg\|_{L_t^2(I_k)}^p \bigg)^{\frac{1}{p}}.$$

Hence it is sufficient to show that

$$\left\| \left\| \phi\Big(\frac{\eta(D,t)(D_d-\psi(D',t))}{\delta}\Big)\varphi_{\mathcal{P}_k}(D)f\right\|_{L^2_t(I_k)}\right\|_p \le C\varepsilon B_p(\varepsilon^{-2}\delta)\|\varphi_{\mathcal{P}_k}(D)f\|_p \tag{78}$$

because $\big(\sum_k \|\varphi_{\mathcal{P}_k}(D)f\|_p^p\big)^{\frac1p} \le C\|f\|_p$ for $2\le p\le\infty$. This follows by interpolation between the estimates for $p=2$ and $p=\infty$. The first is an easy consequence of Plancherel's theorem because the $\{2\mathcal{P}_k\}$ are boundedly overlapping and the latter is clear since $\mathcal{F}^{-1}(\phi_{\mathcal{P}_k})\in L^1$ uniformly.

Now we make the change of variables

$$t\to\varepsilon^2(\partial_t\psi(a',t_k))^{-1}t+t_k,\quad \xi\to L(\xi)=(L'(\xi),L_d(\xi)),$$

where

$$L'(\xi)=\varepsilon\mathcal{H}^{\psi}_{(a',t_k)}\zeta+a',\quad L_d(\xi)=\varepsilon^2\tau+\psi(a',t_k)+\varepsilon\nabla_\zeta\psi(a',t_k)\mathcal{H}^{\psi}_{(a',t_k)}\zeta,$$

and

$$\varepsilon^2x_d\to x_d,\quad \varepsilon\mathcal{H}^{\psi}_{(a',t_k)}(x'+x_d\nabla_\zeta\psi(a',t_k))\to x'.$$

Then, (78) follows if we show

$$\left\| \left\| \phi\Big(\frac{\eta(L(D),t)(D_d-\psi^{\varepsilon}_{a',t_k}(D',t))}{\varepsilon^{-2}\delta}\Big)f\right\|_{L^r_t(0,2\nu^2)}\right\|_p \le CB_p(\varepsilon^{-2}\delta)\|f\|_p$$

when the support $\hat f$ is contained in $L^{-1}(2\mathcal{P}_k)$. Clearly, $\eta(L(\xi),t)\in\mathcal{E}(N)$ and $L^{-1}(2\mathcal{P}_k)$ is contained in the set $\{(\zeta,\tau):|\zeta|\le 4\nu,|\tau|\le 8d^2\nu^2\}\subset\frac12 I^d$. From Lemma 3.1 there exists $\kappa>0$ such that $\psi^{\varepsilon}_{a',t_k}\in\overline{\mathfrak{G}}(\varepsilon_\circ,N)$ if $0<\varepsilon\le\kappa$. Hence, using the definition of $B_p(\delta)$ we get the desired inequality for $\varepsilon\le\kappa$. □

3C. ***Multi(sub)linear square-function estimates.*** Let $\psi\in\overline{\mathfrak{G}}(\varepsilon_\circ,N)$ and set

$$\Gamma^t=\Gamma^t(\psi):=\big\{(\zeta,\psi(\zeta,t)):\zeta\in\tfrac12 I^d\big\}. \tag{79}$$

As before we denote by $\Gamma^t(\delta)$ the δ-neighborhood $\Gamma^t+O(\delta)$. Clearly, from (72) it follows that, for $\delta>0$,

$$\Gamma^t(\delta)\cap\Gamma^s(\delta)=\varnothing\quad\text{if } |t-s|\ge C\delta \tag{80}$$

for some $C>0$. We also denote by N^t the (upward) normal map from the surface Γ^t to $\mathbb{S}^{d-1}$.

Definition 3.5 (normal vector field $\boldsymbol{n}=\boldsymbol{n}(\psi)$). The map $(\zeta,t)\to(\zeta,\psi(\zeta,t))$ is clearly one-to-one and we may assume that the image of this map contains I^d by extending $\psi(\zeta,t)$ to a larger set $I^{d-1}\times CI$, while (71) is satisfied. Hence, for each $\xi=(\zeta,\tau)\in I^d$ there is a unique t such that $\xi=(\zeta,\psi(\zeta,t))$. Then we define $\boldsymbol{n}(\xi)$ to be the normal vector to Γ^t at ξ, which forms a vector field on I^d.

A natural attempt for the multilinear generalization of S_δ is to consider $\prod_{i=1}^k S_\delta f_i$ under a transversality condition between supp f_i. But, the induction-on-scale argument does not work well with this naive generalization and it doesn't seem easy to obtain the sharp multilinear square-function estimates directly. We get around this difficulty by considering a vector-valued extension in which we discard the exact structure of the operator S_δ. As is clearly seen in its proof, the estimate in Proposition 3.6 is not limited to the surfaces given by $\psi\in\overline{\mathfrak{G}}(\varepsilon_\circ,N)$ but it holds for a more general class of surfaces as long as the transversality is satisfied.

Proposition 3.6. *Let $2 \le k \le d$ be an integer and $0 < \sigma \ll 1$, and let Γ^t be given by $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$, and the functions G_i, $1 \le i \le k$, be defined on $\mathbb{R}^d \times I$. Suppose that, for each $t \in I$, $G_1(\cdot, t), \dots, G_k(\cdot, t)$ satisfy that, for $0 < \delta \ll \sigma$,*

$$\operatorname{supp} \widehat{G}_i(\cdot, t) \subset \Gamma^t(\delta), \quad t \in I, \tag{81}$$

and suppose that

$$\mathrm{Vol}(\boldsymbol{n}(\xi_1), \boldsymbol{n}(\xi_2), \dots, \boldsymbol{n}(\xi_k)) \gtrsim \sigma, \tag{82}$$

whenever $\xi_i \in \operatorname{supp} \widehat{G}_i(\cdot, t) + O(\delta)$ for some $t \in I$. Then, if $p \ge 2k/(k-1)$ and $\varepsilon_\circ > 0$ is small enough, for $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that

$$\Big\| \prod_{i=1}^k \|G_i\|_{L_t^2(I)} \Big\|_{L^{p/k}(B(x,\delta^{-1}))} \le C\sigma^{-C_\varepsilon} \delta^{-\varepsilon} \prod_{i=1}^k (\delta^{\frac12} \|G_i\|_{L^2_{x,t}}) \tag{83}$$

holds with C, C_ε independent of ψ.

Without being concerned about the optimal α for a while, we first observe that, for $p \ge 2$, there is an α such that

$$\big\| \|G_i\|_{L_t^2(I)} \big\|_{L^p(\mathbb{R}^d)} \le C\delta^{-\alpha} \|G_i\|_{L^2_{x,t}} \tag{84}$$

holds uniformly if $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$ and N is large enough ($N \ge 100d$). (It is enough to keep $\|\psi\|_{C^N(I^d)}$ uniformly bounded.) To see this, let φ be a smooth function supported in $2I$ with $\varphi = 1$ on I, and we set $K_\delta^t = \mathcal{F}^{-1}\big(\varphi\big((\tau - \psi(\zeta, t))/C\delta\big)\tilde\chi(\xi)\big)$. Then, by Lemma 2.9 $|K_\delta^t(x)| \le C\delta\mathfrak{K}_M(x)$ for a large M with C depending only on $\|\psi\|_{C^N(I^d)}$. Since $\operatorname{supp} \mathcal{F}(G_i(\cdot, t)) \subset \Gamma^t(\delta)$, we have $G_i(\cdot, t) = K_\delta^t * G_i(\cdot, t)$. So, $|G_i(x, t)| \le C\delta\mathfrak{K}_M * |G_i(\cdot, t)|$, $t \in I$, and by Minkowski's inequality we get

$$\|G_i(x,t)\|_{L_t^2(I)} \le C\delta\mathfrak{K}_M * (\|G_i(\cdot, t)\|_{L_t^2(I)})(x). \tag{85}$$

Young's convolution inequality gives (84), namely with $\alpha = d - 1$, if taking sufficiently large M.

Proof of Proposition 3.6. Since

$$\mathcal{F}(G_i(\cdot, t)) = \varphi\Big(\frac{\tau - \psi(\zeta, t)}{C\delta}\Big)\tilde\chi(\xi)\mathcal{F}(G_i(\cdot, t)),$$

by Schwarz's inequality and Plancherel's theorem, $|G_i(x, t)| \lesssim \delta^{\frac12}\|G_i(\cdot, t)\|_2$. So, this gives (83) for $p = \infty$. Thus, by interpolation it is sufficient to show (83) with $p = 2k/(k-1)$.

Let us set $R = \delta^{-1}$ and we may set $x = 0$. Following the same argument as in the proof of Proposition 2.11 we start with the assumption that, for $0 < \delta \ll \sigma$,

$$\Big\| \prod_{i=1}^k \|G_i\|_{L_t^2(I)} \Big\|_{L^{2/(k-1)}(B(0,R))} \lesssim R^\alpha R^{-\frac{k}{2}} \prod_{i=1}^k \|G_i\|_{L^2_{x,t}} \tag{86}$$

holds uniformly for $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$ whenever (81) and (82) are satisfied. By (84) and Hölder's inequality, this is true for a large $\alpha > 0$. Hence, it is sufficient to show (86) implies that for $\varepsilon > 0$ there is an $N = N(\varepsilon)$

such that, for some $\kappa > 0$,

$$\Big\| \prod_{i=1}^{k} \|G_i\|_{L_t^2(I)} \Big\|_{L^{2/(k-1)}(B(0,R))} \lesssim C_\varepsilon \sigma^{-\kappa} R^{\frac{\alpha}{2}+c\varepsilon} R^{-\frac{k}{2}} \prod_{i=1}^{k} \|G_i\|_{L^2_{x,t}} \tag{87}$$

holds uniformly for $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$. Then, iterating this implication from (86) to (87) gives the desired inequality; see the paragraph below (20).

Since $\hat{\rho}_{B(z,\sqrt{R})}$ is supported in a ball of radius $\sim R^{-\frac{1}{2}}$, the Fourier transform of $\rho_{B(z,\sqrt{R})} G_i(\cdot,t)$ is contained in $\Gamma^t + O(R^{-\frac{1}{2}})$ for each t and (82) holds with $\delta = R^{-\frac{1}{2}}$ since $\delta \ll \sigma$. Hence, by the assumption (86), it follows that

$$\Big\| \prod_{i=1}^{k} \|\rho_{B(z,\sqrt{R})} G_i\|_{L_t^2(I)} \Big\|_{L^{2/(k-1)}(B(z,\sqrt{R}))} \le C R^{\frac{\alpha}{2}} R^{-\frac{k}{4}} \prod_{i=1}^{k} \|\rho_{B(z,\sqrt{R})} G_i\|_{L^2_{x,t}}. \tag{88}$$

We now decompose $G_i(\cdot,t)$ into $\{G_{i,\boldsymbol{q}}(\cdot,t)\}$, which is defined by

$$\mathcal{F}(G_{i,\boldsymbol{q}}(\cdot,t)) = \chi_{\boldsymbol{q}} \mathcal{F}(G_i(\cdot,t)). \tag{89}$$

Here $\{\boldsymbol{q}\}$ are the dyadic cubes of side length l, $R^{-\frac{1}{2}} < l \le 2R^{-\frac{1}{2}}$, which we already used in the proof of Proposition 2.11. We write

$$G_i(x,t) = \sum_{\boldsymbol{q}} G_{i,\boldsymbol{q}}(x,t).$$

In what follows we assume $G_{i,\boldsymbol{q}} \ne 0$. By (81) it follows that, for each t, the cubes $\{\boldsymbol{q}\}$ appearing in the sum are contained in $\Gamma^t(R^{-\frac{1}{2}})$ because $G_{i,\boldsymbol{q}}(\cdot,t) = 0$, otherwise. We also note from (72) that there is an interval $I_{i,\boldsymbol{q}}$ of length $CR^{-\frac{1}{2}}$ such that $G_{i,\boldsymbol{q}}(\cdot,t) = 0$ if $t \notin I_{i,\boldsymbol{q}}$. Hence we may multiply the characteristic function of $\chi_{I_{i,\boldsymbol{q}}}$ so that

$$G_{i,\boldsymbol{q}} = G_{i,\boldsymbol{q}}(\cdot,t)\chi_{I_{i,\boldsymbol{q}}}(t). \tag{90}$$

Since the Fourier supports of $\{\rho_{B(z,\sqrt{R})} G_{i,\boldsymbol{q}}(\cdot,t)\}$ are boundedly overlapping, by Plancherel's theorem it follows that

$$\prod_{i=1}^{k} \|\rho_{B(z,\sqrt{R})} G_i\|_{L^2_{x,t}} \le C \prod_{i=1}^{k} \Big\| \Big(\sum_{\boldsymbol{q}} |\rho_{B(z,\sqrt{R})} G_{i,\boldsymbol{q}}|^2 \Big)^{\frac{1}{2}} \Big\|_{L^2_{x,t}}. \tag{91}$$

Combining this with (88) we have

$$\Big\| \prod_{i=1}^{k} \|\rho_{B(z,\sqrt{R})} G_i\|_{L_t^2(I)} \Big\|_{L^{2/(k-1)}} \le C R^{\frac{\alpha}{2}} R^{-\frac{k}{4}} \prod_{i=1}^{k} \Big\| \Big(\sum_{\boldsymbol{q}} |\rho_{B(z,\sqrt{R})} G_{i,\boldsymbol{q}}|^2 \Big)^{\frac{1}{2}} \Big\|_{L^2_{x,t}}.$$

Since $\rho_{B(z,\sqrt{R})}$ is rapidly decaying outside of $B(z,\sqrt{R})$, we have for any large $M > 0$

$$\Big\| \prod_{i=1}^{k} \|\rho_{B(z,\sqrt{R})} G_i\|_{L_t^2(I)} \Big\|_{L^{2/(k-1)}} \lesssim R^{\frac{\alpha}{2}-\frac{k}{4}} \prod_{i=1}^{k} \Big\| \chi_{B(z,R^{1/2+\varepsilon})} \Big(\sum_{\boldsymbol{q}} |G_{i,\boldsymbol{q}}|^2 \Big)^{\frac{1}{2}} \Big\|_{L^2_{x,t}} + R^{-M} \prod_{i=1}^{k} \|G_i\|_{L^2_{x,t}}. \tag{92}$$

We now partition the interval $I_{i,\boldsymbol{q}}$ further into intervals $I^l_{i,\boldsymbol{q}} = [t_l, t_{l+1}]$, $l = 1, \dots, \ell_0$, of length $\sim R^{-1}$. Then the Fourier support of $G_{i,\boldsymbol{q}}(\cdot, t)$, $t \in I^l_{i,\boldsymbol{q}} = [t_l, t_{l+1}]$, is contained in an $O(R^{-1})$ neighborhood of Γ^{t_l}. Let $(\zeta_{\boldsymbol{q}}, \tau_{\boldsymbol{q}})$ be the center of $\boldsymbol{q}$ and we define a set $\boldsymbol{r}^l_{i,\boldsymbol{q}}$ by

$$\boldsymbol{r}^l_{i,\boldsymbol{q}} = \big\{(\zeta,\tau) : |\zeta - \zeta_{\boldsymbol{q}}| \le C\delta^{\frac12},\ |\tau - \psi(\zeta_{\boldsymbol{q}}, t_l) - \nabla_\zeta \psi(\zeta_{\boldsymbol{q}}, t_l)\cdot(\zeta - \zeta_{\boldsymbol{q}})| \le C\delta\big\} \tag{93}$$

with a constant $C > 0$ large enough. It follows that the Fourier transform of $G_{i,\boldsymbol{q}}(\cdot, t)$, $t \in I^l_{i,\boldsymbol{q}}$, is supported in $\boldsymbol{r}^l_{i,\boldsymbol{q}}$. This is easy to see from the second-order Taylor approximation because $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$.

Also define $\mathfrak{m}^l_{i,\boldsymbol{q}}$ by

$$\mathfrak{m}^l_{i,\boldsymbol{q}} = \rho\bigg(\frac{\zeta - \zeta_{\boldsymbol{q}}}{C\sqrt{\delta}}, \frac{\tau - \psi(\zeta_q, t_l) - \nabla_\zeta\psi(\zeta_q, t_l)\cdot(\zeta - \zeta_{\boldsymbol{q}})}{C\delta}\bigg) \tag{94}$$

with a suitable $C > 0$ such that $\mathfrak{m}^l_{i,\boldsymbol{q}}$ is comparable to 1 on $\boldsymbol{r}^l_{i,\boldsymbol{q}}$. Now, we set

$$\mathcal{F}(G^l_{i,\boldsymbol{q}}(\cdot, t)) = (\mathfrak{m}^l_{i,\boldsymbol{q}})^{-1}\mathcal{F}(G_{i,\boldsymbol{q}}(\cdot, t))\chi_{I^l_{i,\boldsymbol{q}}}(t). \tag{95}$$

Denoting by $\boldsymbol{n}^l_{i,\boldsymbol{q}}$ the normal vector $\boldsymbol{n}(\zeta_{\boldsymbol{q}}, \psi(\zeta_{\boldsymbol{q}}, t_l))$, we also set with a large $C > 0$

$$\boldsymbol{T}^l_{i,\boldsymbol{q}} = \big\{x : |x\cdot \boldsymbol{n}^l_{i,\boldsymbol{q}}| \le C,\ |x - (x\cdot\boldsymbol{n}^l_{i,\boldsymbol{q}})\boldsymbol{n}^l_{i,\boldsymbol{q}}| \le CR^{-\frac12}\big\}.$$

Let us set $K^l_{i,\boldsymbol{q}} = \mathcal{F}^{-1}(\mathfrak{m}^l_{i,\boldsymbol{q}})$ so that $G_{i,\boldsymbol{q}}(\cdot, t) = G^l_{i,\boldsymbol{q}}(\cdot, t) * K^l_{i,\boldsymbol{q}}$ if $t \in I^l_{i,\boldsymbol{q}}$. Since $\hat\rho$ is supported in $\mathfrak{q}(0,1)$,

$$|K^l_{i,\boldsymbol{q}}| \lesssim R^{-\frac{d+1}{2}}\chi_{R\boldsymbol{T}^l_{i,\boldsymbol{q}}}.$$

By (90) it follows that

$$\sum_{\boldsymbol{q}} \|G_{i,\boldsymbol{q}}\|^2_{L^2_t(I)} = \sum_{\boldsymbol{q}} \|G_{i,\boldsymbol{q}}\|^2_{L^2_t(I_{i,\boldsymbol{q}})} = \sum_{\boldsymbol{q},l} \|G_{i,\boldsymbol{q}}\|^2_{L^2_t(I^l_{i,\boldsymbol{q}})}.$$

Thus, by (95) we have

$$\begin{aligned}\sum_{\boldsymbol{q}} \|G_{i,\boldsymbol{q}}\|^2_{L^2_t(I)} &= \sum_{\boldsymbol{q},l} \|G^l_{i,\boldsymbol{q}}(\cdot, t) * K^l_{i,\boldsymbol{q}}\|^2_{L^2(I^l_{i,\boldsymbol{q}})}\\ &\lesssim \sum_{\boldsymbol{q},l} \|G^l_{i,\boldsymbol{q}}(\cdot, t)\|^2_{L^2(I^l_{i,\boldsymbol{q}})} * |K^l_{i,\boldsymbol{q}}|\\ &\lesssim \sum_{\boldsymbol{q},l} \|G^l_{i,\boldsymbol{q}}(\cdot, t)\|^2_{L^2(I^l_{i,\boldsymbol{q}})} * (R^{-\frac{d+1}{2}}\chi_{R\boldsymbol{T}^l_{i,\boldsymbol{q}}}).\end{aligned} \tag{96}$$

We denote by $\widetilde{\boldsymbol{T}}^l_{i,\boldsymbol{q}}$ the tube $R^{1+\varepsilon}\boldsymbol{T}^l_{i,\boldsymbol{q}}$, which is an $R^{1+\varepsilon}$-times dilation of $\boldsymbol{T}^l_{i,\boldsymbol{q}}$ from its center. So, from (96) we have, for $x, y \in B(z, R^{\frac12+\varepsilon})$,

$$\sum_{\boldsymbol{q}} \|G_{i,\boldsymbol{q}}(y,\cdot)\|^2_{L^2_t(I)} \lesssim R^{c\varepsilon} \sum_{\boldsymbol{q},l} \|G^l_{i,\boldsymbol{q}}(\cdot, t)\|^2_{L^2(I^l_{i,\boldsymbol{q}})} * \bigg(\frac{\chi_{\widetilde{\boldsymbol{T}}^l_{i,\boldsymbol{q}}}}{|\widetilde{\boldsymbol{T}}^l_{i,\boldsymbol{q}}|}\bigg)(x).$$

Once we have this equality we can repeat the argument from (23) to (26) which is in the proof of Proposition 2.11 and also using (92), we have

$$\Big\|\prod_{i=1}^k \|G_i\|_{L_t^2(I)}\Big\|_{L^{2/(k-1)}(B(0,R))} \lesssim R^{c\varepsilon+\frac{\alpha}{2}+\frac{d-k}{4}}\Big\|\prod_{i=1}^k\Big(\sum_{\boldsymbol{q},l}\|G^l_{i,\boldsymbol{q}}(\cdot,t)\|^2_{L^2(I^l_{i,\boldsymbol{q}})}*\Big(\frac{\chi_{\widetilde{\boldsymbol{T}}^l_{i,\boldsymbol{q}}}}{|\widetilde{\boldsymbol{T}}^l_{i,\boldsymbol{q}}|}\Big)\Big)^{\frac12}\Big\|_{L^{2/(k-1)}(B(0,2R))}+\mathcal{E},$$

where $\mathcal{E}=R^{-M}\prod_{i=1}^k\|G_i\|_{L^2_{x,t}}$ for any large $M>0$. Hence, for (87) it suffices to show that

$$\Big\|\prod_{i=1}^k\Big(\sum_{\boldsymbol{q},l}\|G^l_{i,\boldsymbol{q}}(\cdot,t)\|^2_{L^2(I^l_{i,\boldsymbol{q}})}*\Big(\frac{\chi_{\widetilde{\boldsymbol{T}}^l_{i,\boldsymbol{q}}}}{|\widetilde{\boldsymbol{T}}^l_{i,\boldsymbol{q}}|}\Big)\Big)^{\frac12}\Big\|_{L^{2/(k-1)}(B(0,2R))}\lesssim\sigma^{-\kappa}R^{c\varepsilon}R^{-\frac{d+k}{4}}\prod_{i=1}^k\|G_i\|_{L^2_{x,t}}.$$

Since by (95) $\big\|\|G^l_{i,\boldsymbol{q}}\|_{L^2_t(I^l_{i,\boldsymbol{q}})}\big\|_2\sim\big\|\|G_{i,\boldsymbol{q}}\|_{L^2_t(I^l_{i,\boldsymbol{q}})}\big\|_2$, making use of the disjointness of $I^l_{i,\boldsymbol{q}}$ and the supports of $\mathcal{F}(G_{i,\boldsymbol{q}}(\cdot,t))$, and by Plancherel's theorem,

$$\sum_{\boldsymbol{q},l}\big\|\|G^l_{i,\boldsymbol{q}}\|_{L^2_t(I^l_{i,\boldsymbol{q}})}\big\|_2^2\sim\sum_{\boldsymbol{q}}\|G_{i,\boldsymbol{q}}\|^2_{L^2_t(I)}=\big\|\|G_i\|_{L^2_t(I)}\big\|_2^2.$$

Hence, the above inequality follows from

$$\Big\|\prod_{i=1}^k\sum_{\boldsymbol{q},l}f^l_{i,\boldsymbol{q}}*\Big(\frac{\chi_{\widetilde{\boldsymbol{T}}^l_{i,\boldsymbol{q}}}}{|\widetilde{\boldsymbol{T}}^l_{i,\boldsymbol{q}}|}\Big)\Big\|_{L^{1/(k-1)}(B(0,2R))}\le C\sigma^{-\kappa}R^{c\varepsilon}R^{-\frac{d+k}{2}}\prod_{i=1}^k\sum_{\boldsymbol{q},l}\|f^l_{i,\boldsymbol{q}}\|_1.$$

Let $\mathcal{I}_i=\{(\boldsymbol{q},l):G^l_{i,\boldsymbol{q}}\neq0\}$, $I_i\subset\mathcal{I}_i$ and $\mathcal{T}^l_{i,\boldsymbol{q}}$ be a finite subset of $\mathbb{R}^d$. By scaling and pigeonholing, losing $(\log R)^C$ in its bound, this reduces to

$$\Big\|\prod_{i=1}^k\sum_{(\boldsymbol{q},l)\in\mathcal{I}_i}\sum_{\tau\in\mathcal{T}^l_{i,\boldsymbol{q}}}\chi_{\boldsymbol{T}^l_{i,\boldsymbol{q}}+\tau}\Big\|_{L^{1/(k-1)}(B(0,2))}\le C\sigma^{-\kappa}R^{c\varepsilon}R^{\frac{d-k}{2}}\prod_{i=1}^k\sum_{(\boldsymbol{q},l)\in\mathcal{I}_i}\sum_{\tau\in\mathcal{T}^l_{i,\boldsymbol{q}}}|\boldsymbol{T}^l_{i,\boldsymbol{q}}+\tau|. \tag{97}$$

Here we note that if $G_{i,\boldsymbol{q}}\neq0$, then $\boldsymbol{q}\in\operatorname{supp}\mathcal{F}(G_i(\cdot,t))+O(\sqrt{\delta})$ for some t. So, by (82) we have $\mathrm{Vol}(\boldsymbol{n}_1,\dots,\boldsymbol{n}_k)\gtrsim\sigma$ whenever $\boldsymbol{n}_i\in\{\boldsymbol{n}^l_{i,\boldsymbol{q}}:G^l_{i,\boldsymbol{q}}\neq0\}$, $i=1,\dots,k$. Therefore, the estimate follows from the multilinear Kakeya estimate which is stated below in Theorem 3.7. □

Theorem 3.7 [Bennett et al. 2006; Guth 2010; Carbery and Valdimarsson 2013]. *Let $2\le k\le d$, $1\ll R$ and $\mathfrak{T}_i$, $i=1,2,\dots,k$, be collections of tubes of width $R^{-\frac12}$ (possibly with infinite length), with major axes parallel to the vectors in $\Theta_i\subset\mathbb{S}^{d-1}$. Suppose $\mathrm{Vol}(\theta_1,\theta_2,\dots,\theta_k)\ge\sigma$ holds whenever $\theta_i\in\Theta_i$, $i=1,\dots,k$. Then there is a constant C such that, for any subset $\mathcal{T}_i\subset\mathfrak{T}_i$, $i=1,\dots,k$,*

$$\Big\|\prod_{i=1}^k\Big(\sum_{T_i\in\mathcal{T}_i}\chi_{T_i}\Big)\Big\|_{L^{1/(k-1)}(B(0,1))}\le CR^{\frac{d-k}{2}}\sigma^{-1}\prod_{i=1}^k\Big(\sum_{T_i\in\mathcal{T}_i}|T_i|\Big).$$

This is a rescaled version of the estimate due to Guth [2010] (the case $d=k$) and Carbery and Valdimarsson [2013]; also see [Bennett et al. 2006]. However, we don't need the endpoint estimate for our purpose and the estimate in [Bennett et al. 2006] is actually enough because we allow a $\delta^{-\varepsilon}$ loss in our estimate.

Corollary 3.8. *Let $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$, $\eta \in \mathcal{E}(N)$, and $0 < \delta \ll \sigma$. Suppose that (82) holds whenever $\xi_i \in \operatorname{supp} \hat{f}_i + O(\delta)$, $i = 1, 2, \dots, k$. Then, if $p \geq 2k/(k-1)$ and $\varepsilon_\circ$ is small enough, for $\varepsilon > 0$, there is an $N = N(\varepsilon)$ such that the following estimate holds with C, C_ε, independent of ψ and η:*

$$\Big\| \prod_{i=1}^{k} S_\delta(\psi, \eta) f_i \Big\|_{L^{p/k}(B(x,\delta^{-1}))} \leq C \sigma^{-C_\varepsilon} \delta^{-\varepsilon} \prod_{i=1}^{k} (\delta \| f_i \|_2).$$

To show this we need only to replace G_i with $\phi\big(\eta(D,t)(D_d - \psi(D',t))/\delta\big) f_i$ and apply Proposition 3.6. The assumptions in Proposition 3.6 are satisfied with $G_1, \dots, G_k$. Thus, the estimate is straightforward because $\|\phi\big(\eta(D,t)(D_d - \psi(D',t))/\delta\big) f_i\|_{L^2_{x,t}} \lesssim \delta^{\frac{1}{2}} \|f\|_2$, which follows by Plancherel's theorem and taking t-integration first.

The following result is a consequence of Corollary 3.8 and localization argument in the proof of Proposition 2.10.

Proposition 3.9. *Let $0 < \delta \ll \sigma \ll \tilde{\sigma} \ll 1$ and $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$, $\eta \in \mathcal{E}(N)$ and let $Q_1, \dots, Q_k \subset \frac{1}{2} I^d$ be dyadic cubes of side length $\tilde{\sigma}$. Suppose that (82) is satisfied whenever $\xi_i \in Q_i$, $i = 1, \dots, k$, and suppose that $\operatorname{supp} \hat{f}_i \subset Q_i$, $i = 1, \dots, k$. Then, if $p \geq 2k/(k-1)$ and $\varepsilon_\circ$ is small enough, for $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that*

$$\Big\| \prod_{i=1}^{k} S_\delta(\psi, \eta) f_i \Big\|_{\frac{p}{k}} \leq C \sigma^{-C_\varepsilon} \delta^{-\varepsilon} \prod_{i=1}^{k} (\delta^{\frac{d}{p} - \frac{d-2}{2}} \| f_i \|_p) \tag{98}$$

holds with C, C_ε, independent of ψ and η.

Proof. The proof is similar to that of Proposition 2.10. So, we shall be brief. Let φ, $\widetilde{Q}_i$, $\tilde{\chi}_i$, $\{\mathcal{B}\}$, and $\{\widetilde{\mathcal{B}}\}$ be the same as in the proof of Proposition 2.10. We set

$$K_i^t = \mathcal{F}^{-1}\bigg(\phi\bigg(\frac{\eta(\xi,t)(\tau - \psi(\zeta,t))}{\delta}\bigg)\tilde{\chi}_i(\xi)\bigg).$$

Then $S_\delta(\psi, \eta) f_i = \|K_i^t * f_i\|_{L^2_t(I)}$. The (p/k)-th power of the left-hand side of (98) is bounded by

$$\sum_{\mathcal{B}} \int_{\mathcal{B}} \prod_{i=1}^{k} \|K_i^t * f_i\|_{L^2_t(I)}^{\frac{p}{k}} \, dx \lesssim I + II,$$

where

$$I = \sum_{\mathcal{B}} \int_{\mathcal{B}} \prod_{i=1}^{k} \|K_i^t * (\chi_{\widetilde{\mathcal{B}}} f_i)\|_{L^2_t(I)}^{\frac{p}{k}} \, dx, \quad II = \sum_{\mathcal{B}} \bigg(\sum_{\substack{g_i = \chi_{\widetilde{\mathcal{B}}^c} f_i \\ \text{for some } i}} \int_{\mathcal{B}} \prod_{i=1}^{k} \|K_i^t * g_i\|_{L^2_t(I)}^{\frac{p}{k}} \, dx \bigg).$$

As before, the second sum is taken over all choices with $g_i = \chi_{\widetilde{\mathcal{B}}} f_i$ or $\chi_{\widetilde{\mathcal{B}}^c} f_i$, and $g_i = \chi_{\widetilde{\mathcal{B}}^c} f_i$ for some i. By choosing $c > 0$ small enough, we see that $\tilde{\chi}_1(D)(\chi_{\widetilde{\mathcal{B}}} f_1), \dots, \tilde{\chi}_k(D)(\chi_{\widetilde{\mathcal{B}}} f_k)$ satisfy the assumption of Corollary 3.8. Since $K_i^t * (\chi_{\widetilde{\mathcal{B}}} f_i)) = \phi\big(\eta(D,t)(D_d - \psi(D',t))/\delta\big)\tilde{\chi}_i(D)(\chi_{\widetilde{\mathcal{B}}} f_i)$, by Corollary 3.8

and Hölder's inequality

$$I \lesssim \sigma^{-C_\varepsilon}\Big(\frac{1}{\delta}\Big)^{\varepsilon} \sum_{\mathcal{B}} \prod_{i=1}^{k} \delta^{\frac{p}{k}} \|\chi_{\widetilde{\mathcal{B}}} f_i\|_2^{\frac{p}{k}} \lesssim \sigma^{-C_\varepsilon}\Big(\frac{1}{\delta}\Big)^{c\varepsilon}\Big(\prod_{i=1}^{k} \delta^{\frac{d}{p}-\frac{d-2}{2}} \|f_i\|_p\Big)^{\frac{p}{k}}.$$

To handle II we note from Lemma 2.9 that $|K_i^t(x)| \le C\delta\mathfrak{K}_M(x)$ with C depending only on $\|\psi\|_{C^N(I^{d-1})}$, $\|\eta\|_{C^N(I^d)}$. Thus,

$$\|K_i^t * (\chi_{\widetilde{\mathcal{B}}^c} f_i)(x)\|_{L_t^2} \le C\delta\delta^{\varepsilon(M-d-1)}\mathfrak{K}_{d+1} * |f_i|(x)$$

if $x \in B$, and $\|K_i * f_i(x)\|_{L_t^2(I)} \le C\delta\mathfrak{K}_{d+1} * |f_i|(x)$. The rest of proof is the same as before. We omit the details. □

3D. *Multilinear square-function estimate with confined direction sets.* From the point of view of Proposition 2.11 we may expect a better estimate thanks to the smallness of supports of the Fourier transforms of the input functions when they are confined in a small neighborhood of a k-dimensional submanifold. The following is a vector-valued generalization of Proposition 2.11.

Proposition 3.10. *Let k, $2 \le k \le d$, be an integer, $0 < \sigma \ll 1$ be fixed, and $\Pi \subset \mathbb{R}^d$ be a k-plane containing the origin. Let $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$ and Γ^t be defined by (79). For $0 < \delta \ll \sigma$, suppose that the functions $G_1, \dots, G_k$ defined on $\mathbb{R}^d \times I$ satisfy (81) for $t \in I$ and (82) whenever $\xi_i \in \operatorname{supp} \mathcal{F}(G_i(\cdot, t)) + O(\delta)$, $i = 1, 2, \dots, k$, for some $t \in I$. Additionally we assume that, for all $t \in I$,*

$$\boldsymbol{n}(\operatorname{supp} \widehat{G}_1(\cdot, t)), \dots, \boldsymbol{n}(\operatorname{supp} \widehat{G}_k(\cdot, t)) \subset \mathbb{S}^{d-1} \cap (\Pi + O(\delta)). \tag{99}$$

Then, if $2 \le p \le 2k/(k-1)$ and ε_0 is sufficiently small, for $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that

$$\Big\| \prod_{i=1}^{k} \|G_i\|_{L_t^2(I)} \Big\|_{L^{p/k}(B(x,\delta^{-1}))} \lesssim \sigma^{-C_\varepsilon} \delta^{dk(\frac{1}{2}-\frac{1}{p})-\varepsilon} \prod_{i=1}^{k} \|G_i\|_{L_{x,t}^2} \tag{100}$$

holds uniformly for $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$.

The following is an easy consequence of (100).

Corollary 3.11. *Let $\{\mathfrak{q}\}$, $\mathfrak{q} \subset \frac{1}{2}I^d$, be the collection of dyadic cubes of side length l, $\delta < l \le 2\delta$. Define $G_{i,\mathfrak{q}}$ by $\mathcal{F}(G_{i,\mathfrak{q}}(\cdot, t)) = \chi_{\mathfrak{q}}\mathcal{F}(G_i(\cdot, t))$ and set $R = \frac{1}{\delta}$. Suppose that the same assumptions as in Proposition 3.10 are satisfied. Then, if $2 \le p \le 2k/(k-1)$ and $\varepsilon_\circ$ is small enough, for $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that*

$$\Big\| \prod_{i=1}^{k} \|G_i\|_{L_t^2(I)} \Big\|_{L^{p/k}(B(x,R))} \lesssim \sigma^{-C_\varepsilon} \delta^{-\varepsilon} \prod_{i=1}^{k} \Big\| \Big(\sum_{\mathfrak{q}} \|G_{i,\mathfrak{q}}\|_{L_t^2(I)}^2 \Big)^{\frac{1}{2}} \rho_{B(x,R)} \Big\|_p \tag{101}$$

holds uniformly for $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$.

Proof. Observe that

$$\Big\| \prod_{i=1}^{k} \|G_i\|_{L_t^2(I)} \Big\|_{L^{p/k}(B(x,R))} \le \Big\| \prod_{i=1}^{k} \Big\| \rho\Big(\frac{\cdot - x}{R}\Big) G_i \Big\|_{L_t^2(I)} \Big\|_{L^{p/k}}.$$

Then, the functions $\rho((\cdot - x)/R)G_i$, $i = 1, \dots, k$, satisfy the assumption in Proposition 3.10 because $\operatorname{supp} \mathcal{F}\big(\rho((\cdot - x)/R)G_i(\cdot, t)\big) = \operatorname{supp} \widehat{G}(\cdot, t) + O(R^{-1})$. So, from Proposition 3.10 we get

$$\Big\| \prod_{i=1}^k \|G_i\|_{L^2_t(I)} \Big\|_{L^{p/k}(B(x,R))} \lesssim \sigma^{-C_\varepsilon} R^\varepsilon \prod_{i=1}^k R^{-d(\frac{1}{2}-\frac{1}{p})} \Big\| \Big\| \rho\Big(\frac{\cdot - x}{R}\Big) G_i \Big\|_{L^2} \Big\|_{L^2_t(I)}.$$

Since $G_i = \sum_{\mathfrak{q}} G_{i,\mathfrak{q}}$ and the supports of $\big\{\mathcal{F}\big(\rho((\cdot - x)/R)G_{i,\mathfrak{q}}(\cdot, t)\big)\big\}_{\mathfrak{q}}$ are boundedly overlapping, by Plancherel's theorem it follows that

$$\Big\| \Big\| \rho\Big(\frac{\cdot - x}{R}\Big) G_i \Big\|_{L^2_x} \Big\|_{L^2_t(I)} \lesssim \Big\| \Big(\sum_{\mathfrak{q}} \Big\| \rho\Big(\frac{\cdot - x}{R}\Big) G_{i,\mathfrak{q}} \Big\|_2^2 \Big)^{\frac{1}{2}} \Big\|_{L^2_t(I)}.$$

Combining this with the above inequality, we get

$$\Big\| \prod_{i=1}^k \|G_i\|_{L^2_t(I)} \Big\|_{L^{p/k}(B(x,R))} \lesssim \sigma^{-C_\varepsilon} R^\varepsilon \prod_{i=1}^k R^{-d(\frac{1}{2}-\frac{1}{p})} \Big\| \Big|\rho\Big(\frac{\cdot - x}{R}\Big)\Big| \Big(\sum_{\mathfrak{q}} \|G_{i,\mathfrak{q}}\|^2_{L^2_t(I)} \Big)^{\frac{1}{2}} \Big\|_2.$$

Now Hölder's inequality gives the desired estimate (101). □

As an application of Corollary 3.11 we obtain the following.

Corollary 3.12. *Let $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$, $\eta \in \mathcal{E}(N)$, $0 < \delta \ll \tilde{\sigma} \ll \sigma$, and $S_\delta = S_\delta(\psi, \eta)$ be defined by (74). Let Π be a k-plane which contains the origin. Suppose (82) holds whenever $\xi_i \in \operatorname{supp} \hat{f}_i + O(\tilde{\sigma})$, $i = 1, 2, \dots, k$, and*

$$\boldsymbol{n}(\operatorname{supp} \hat{f}_i) \subset \Pi + O(\tilde{\sigma}), \quad i = 1, 2, \dots, k. \tag{102}$$

Let $\{\mathfrak{q}\}$, $\mathfrak{q} \in \frac{1}{2} I^d$, be the collection of dyadic cubes of side length l, $\tilde{\sigma} < l \le 2\tilde{\sigma}$. Define $f_{i,\mathfrak{q}}$ by $\mathcal{F}(f_{i,\mathfrak{q}}) = \chi_{\mathfrak{q}} \mathcal{F}(f_i)$. Then, if $2k/(k-1) \le p \le 2$ and $\varepsilon_\circ$ is sufficiently small, for $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that

$$\Big\| \prod_{i=1}^k S_\delta f_i \Big\|_{L^{p/k}(B(x,1/\tilde{\sigma}))} \lesssim \sigma^{-C_\varepsilon} \tilde{\sigma}^{-\varepsilon} \prod_{i=1}^k \Big\| \Big(\sum_{\mathfrak{q}} |S_\delta f_{i,\mathfrak{q}}|^2 \Big)^{\frac{1}{2}} \rho_{B(x,1/\tilde{\sigma})} \Big\|_{L^p}$$

holds uniformly for ψ and η.

This follows from Corollary 3.11. Indeed, it suffices to check that

$$G_i = \rho(\tilde{\sigma}(\cdot - x)) \phi\big((D_d - \psi(D', t))/\sigma\big) f_i$$

satisfies the assumption of Corollary 3.11 with $\delta = \tilde{\sigma}$ as long as $\sigma \ll \tilde{\sigma}$. This is clear because

$$\widehat{G}_i(\cdot, t) = \tilde{\sigma}^{-d} \Big(e^{i\langle \cdot, x \rangle} \rho\Big(\frac{\cdot}{\tilde{\sigma}}\Big) \Big) * \Big(\phi\Big(\frac{\tau - \psi(\zeta, t)}{\sigma}\Big) \hat{f}_i \Big).$$

Proof of Proposition 3.10. The argument here is similar to the proof of Proposition 3.6. The estimate for $p = 2$ follows from Hölder's inequality and Plancherel's theorem. So, by interpolation it is sufficient to show (100) for $p = 2k/(k-1)$.

Let us set $R=\frac{1}{\delta}\gg 1$ and we may set $x=0$. As usual we start with the assumption that, for $0<\delta\ll\sigma$,

$$\Big\|\prod_{i=1}^{k}\|G_i\|_{L_t^2(I)}\Big\|_{L^{p/k}(B(0,R))}\le CR^{\alpha}R^{-\frac{d}{2}}\prod_{i=1}^{k}\|G_i\|_{L^2_{x,t}} \tag{103}$$

holds uniformly for $\psi\in\overline{\mathfrak{G}}(\varepsilon_\circ,N)$ whenever $G_1,\dots,G_k$ satisfy (81), (82) and (99). By (84) and Hölder's inequality, (103) is true for some large α. As before it is sufficient to show that (103) implies for any $\varepsilon>0$ there is an $N=N(\varepsilon)$ such that

$$\Big\|\prod_{i=1}^{k}\|G_i\|_{L_t^2(I)}\Big\|_{L^{p/k}(B(0,R))}\le C\sigma^{-\kappa}R^{\frac{\alpha}{2}+c\varepsilon}R^{-\frac{d}{2}}\prod_{i=1}^{k}\|G_i\|_{L^2_{x,t}}$$

holds uniformly for $\psi\in\overline{\mathfrak{G}}(\varepsilon_\circ,N)$. Then iteration of this implication gives the desired estimate (100).

Fix $z\in\mathbb{R}^d$ and consider $\rho_{B(z,\sqrt{R})}G_1(\cdot,t),\dots,\rho_{B(z,\sqrt{R})}G_k(\cdot,t)$. Then it is clear from (81) and (99) that $\operatorname{supp}\mathcal{F}(\rho_{B(z,\sqrt{R})}G_i(\cdot,t))\subset\Gamma^t+O(R^{-\frac{1}{2}})$ and

$$\boldsymbol{n}(\operatorname{supp}\mathcal{F}(\rho_{B(z,\sqrt{R})}G_i(\cdot,t)))\subset\Pi+O(R^{-\frac{1}{2}}).$$

Also, since $\delta\ll\sigma$, (82) holds if $\xi_i\in\operatorname{supp}\mathcal{F}(\rho_{B(z,\sqrt{R})}G_i(\cdot,t))$. Hence, by the assumption (103) we get

$$\Big\|\prod_{i=1}^{k}\|\rho_{B(z,\sqrt{R})}G_i\|_{L_t^2(I)}\Big\|_{L^{2/(k-1)}}\lesssim R^{\frac{\alpha}{2}}R^{-\frac{d}{4}}\prod_{i=1}^{k}\|\rho_{B(z,\sqrt{R})}G_i\|_{L^2_{x,t}}. \tag{104}$$

Now we proceed in the same way as in the proof of Proposition 3.6, and we keep using the same notations. As before, let $\{\boldsymbol{q}\}$ be the collection of dyadic cubes (hence essentially disjoint) of side length $\sim R^{-\frac{1}{2}}$ such that $I^d=\bigcup\boldsymbol{q}$. We decompose the function $G_i(\cdot,t)$ into $G_{i,\boldsymbol{q}}(\cdot,t)$, which is defined by (89), and get (91), which is clear. Then, combining (91) and (104), we have

$$\Big\|\prod_{i=1}^{k}\|\rho_{B(z,\sqrt{R})}G_i\|_{L_t^2(I)}\Big\|_{L^{2/(k-1)}}\le CR^{\frac{\alpha}{2}}R^{-\frac{d}{4}}\prod_{i=1}^{k}\Big\|\Big(\sum_{\boldsymbol{q}}|\rho_{B(z,\sqrt{R})}G_{i,\boldsymbol{q}}|^2\Big)^{\frac{1}{2}}\Big\|_{L^2_{x,t}}.$$

Then this gives

$$\Big\|\prod_{i=1}^{k}\|\rho_{B(z,\sqrt{R})}G_i\|_{L_t^2(I)}\Big\|_{L^{2/(k-1)}}\lesssim R^{\frac{\alpha}{2}-\frac{d}{4}}\prod_{i=1}^{k}\Big\|\chi_{B(z,R^{1/2+\varepsilon})}\Big(\sum_{\boldsymbol{q}}|G_{i,\boldsymbol{q}}|^2\Big)^{\frac{1}{2}}\Big\|_{L^2_{x,t}}+\mathcal{E}, \tag{105}$$

where $\mathcal{E}=R^{-M}\prod_{i=1}^{k}\|G_i\|_{L^2_{x,t}}$ for any large M.

We also denote by $(\mathrm{N}^t)^{-1}$ (defined from $\mathrm{N}^t(I^{d-1})$ to I^{d-1}) the inverse of $\mathrm{N}^t:\Gamma^t\to\mathbb{S}^{d-1}$ which is well defined because $\psi\in\overline{\mathfrak{G}}(\varepsilon_\circ,N)$. Since $\partial_t\psi\in(1-\varepsilon_\circ,1+\varepsilon_\circ)$, there is an interval $I_{i,\boldsymbol{q}}$ of length $CR^{-\frac{1}{2}}$ such that $G_{i,\boldsymbol{q}}(\cdot,t)=0$ if $t\notin I_{i,\boldsymbol{q}}$; see (90). As in the proof of Proposition 3.6 we partition $I_{i,\boldsymbol{q}}$ into intervals $I^l_{i,\boldsymbol{q}}=[t_l,t_{l+1}]$, $l=1,\dots,l_0$, of side length $\sim R^{-1}$. Since the Fourier transform of $G_i(\cdot,t)$ is supported in $\Gamma^{t_l}+O(\delta)$ if $t\in I^l_{i,\boldsymbol{q}}=[t_l,t_{l+1}]$ and the normal vectors are confined in $\Pi+O(\delta)$, it follows that

$$\operatorname{supp}\mathcal{F}(G_{i,\boldsymbol{q}}(\cdot,t))\subset\Gamma^{t_l}(\delta)\cap((\mathrm{N}^{t_l})^{-1}(\Pi)+O(\delta)),\quad t\in[t_l,t_{l+1}].$$

Fix t_l, and let us set

$$\xi_{i,\boldsymbol{q}}^{t_l} = (\zeta_{i,\boldsymbol{q}}^{t_l}, \tau_{i,\boldsymbol{q}}^{t_l}) \in ((\mathrm{N}^{t_l})^{-1}(\Pi) \cap \Gamma^{t_l}) \cap (\operatorname{supp} \mathcal{F}(G_{i,\boldsymbol{q}}(\cdot, t_l)) + O(\delta)).$$

(As before, we may assume that this set is nonempty, otherwise the associated function $G_{i,\boldsymbol{q}}^l$ is equal to 0. See below.) Let $v_1, \dots, v_{k-1}$ be an orthonormal basis for the tangent space $T_{\xi_{i,\boldsymbol{q}}^{t_l}}((\mathrm{N}^{t_l})^{-1}(\Pi))$ at $\xi_{i,\boldsymbol{q}}^{t_l}$, and $u_1, \dots, u_{d-k}$ be a set of orthonormal vectors such that $\{\mathrm{N}^{t_l}(\xi_{i,\boldsymbol{q}}^{t_l}), v_1, \dots, v_{k-1}, u_1, \dots, u_{d-k}\}$ forms an orthonormal basis for $\mathbb{R}^d$. Let us set

$$\begin{aligned}\boldsymbol{r}_{i,\boldsymbol{q}}^{t_l} &= \{\xi : |(\xi - \xi_{i,\boldsymbol{q}}^{t_l}) \cdot \mathrm{N}^{t_l}(\xi_{i,\boldsymbol{q}}^{t_l})| \le C\delta,\ |(\xi - \xi_{i,\boldsymbol{q}}^{t_l}) \cdot v_i| \le C\sqrt{\delta},\ i = 1, \dots, k-1, \\ &\qquad\qquad \text{and } |(\xi - \xi_{i,\boldsymbol{q}}^{t_l}) \cdot u_i| \le C\delta,\ i = 1, \dots, d-k\}, \\ \boldsymbol{P}_{i,\boldsymbol{q}}^{t_l} &= \{\xi : |\xi \cdot \mathrm{N}^{t_l}(\xi_{i,\boldsymbol{q}}^{t_l})| \le C,\ |\xi \cdot v_i| \le C\sqrt{\delta},\ i = 1, \dots, k-1, \text{ and } |\xi \cdot u_i| \le C,\ i = 1, \dots, d-k\}\end{aligned}$$

with a sufficiently large $C > 0$. Then $\mathcal{F}(G_{i,\boldsymbol{q}}(\cdot, t))$, $t \in [t_l, t_{l+1}]$ is supported in $\boldsymbol{r}_{i,\boldsymbol{q}}^{t_l}$.

The rest of proof is similar to that of Proposition 3.6, so we shall be brief. Let $\mathfrak{m}_{i,\boldsymbol{q}}^{t_l}$ be a smooth function naturally adapted to $\boldsymbol{r}_{i,\boldsymbol{q}}^{t_l}$ such that $\mathfrak{m}_{i,\boldsymbol{q}}^{t_l} \sim 1$ on $\boldsymbol{r}_{i,\boldsymbol{q}}^{t_l}$ and $\mathcal{F}^{-1}(\mathfrak{m}_{i,\boldsymbol{q}}^{t_l})$ is supported in $R\boldsymbol{P}_{i,\boldsymbol{q}}^{t_l}$. This can be done by using ρ and composing it with an appropriate affine map; for example, see (94). As before we define $G_{i,\boldsymbol{q}}^l(\cdot, t)$ by (95) and let $K_{i,\boldsymbol{q}}^{t_l} = \mathcal{F}^{-1}(\mathfrak{m}_{i,\boldsymbol{q}}^{t_l})$ so that $G_{i,\boldsymbol{q}}^l(\cdot, t) = G_{i,\boldsymbol{q}}^l(\cdot, t) * K_{i,\boldsymbol{q}}^{t_l}$ if $t \in I_{i,\boldsymbol{q}}^l$. Hence,

$$\sum_{\boldsymbol{q}} G_{i,\boldsymbol{q}} = \sum_{\boldsymbol{q},l} G_{i,\boldsymbol{q}}^l(\cdot, t) * K_{i,\boldsymbol{q}}^{t_l}, \quad |K_{i,\boldsymbol{q}}^{t_l}| \lesssim |R\boldsymbol{P}_{i,\boldsymbol{q}}^{t_l}|^{-1} \chi_{R\boldsymbol{P}_{i,\boldsymbol{q}}^{t_l}}.$$

Let us set $\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}^{t_l} = R^{1+\varepsilon} \boldsymbol{P}_{i,\boldsymbol{q}}^{t_l}$. Hence, from the same lines of inequalities as in (96) and repeating an argument similar to that in the proof of Proposition 3.6 we have, for $x \in B(y, R^{\frac{1}{2}+\varepsilon})$,

$$\prod_{i=1}^k \Big(\sum_{\boldsymbol{q}} \|G_{i,\boldsymbol{q}}\|_{L_t^2(I)}^2(x) \Big) \lesssim R^{c\varepsilon} \prod_{i=1}^k \sum_{\boldsymbol{q},i} \|G_{i,\boldsymbol{q}}^l(\cdot, t)\|_{L^2(I_{\boldsymbol{q}}^l)}^2 * \Big(\frac{\chi_{\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}^{t_l}}}{|\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}^{t_l}|} \Big)(y).$$

Now, we use the lines of argument from (23) to (26), and combine this with (105) to get

$$\begin{aligned}&\Big\| \prod_{i=1}^k \|G_i\|_{L_t^2(I)} \Big\|_{L^{2/(k-1)}(B(0,R))} \\ &\qquad \lesssim R^{c\varepsilon + \frac{\alpha}{2}} \Big\| \prod_{i=1}^k \Big(\sum_{\boldsymbol{q},l} \|G_{i,\boldsymbol{q}}^l(\cdot, t)\|_{L^2(I_{\boldsymbol{q}}^l)}^2 * \Big(\frac{\chi_{\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}^{t_l}}}{|\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}^{t_l}|} \Big) \Big)^{\frac{1}{2}} \Big\|_{L^{2/(k-1)}(B(0,2R))} + \mathcal{E}.\end{aligned}$$

Since $\sum_{\boldsymbol{q},l} \big\| \|\widetilde{G}_{i,\boldsymbol{q}}\|_{L_t^2(I_{\boldsymbol{q}}^l)} \big\|_2^2 \sim \sum_{\boldsymbol{q}} \big\| \|G_{i,\boldsymbol{q}}\|_{L_t^2(I_{\boldsymbol{q}})} \big\|_2^2 \sim \|G_i\|_{L_{x,t}^2}$, the proof is completed if we show

$$\Big\| \prod_{i=1}^k \Big(\sum_{\boldsymbol{q},l} f_{\boldsymbol{q},l} * \frac{\chi_{\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}^{t_l}}}{|\widetilde{\boldsymbol{P}}_{i,\boldsymbol{q}}^{t_l}|} \Big) \Big\|_{L^{2/(k-1)}(B(0,2R))} \le C R^{c\varepsilon} \sigma^{-1} R^{-d} \prod_{i=1}^k \Big(\sum_{\boldsymbol{q},l} \|f_{\boldsymbol{q},l}\|_1 \Big).$$

Finally, to show the above inequality we may repeat the argument in the last part in the proof of Proposition 2.11. In fact, we need only to show the associated Kakeya estimate; for example, see (28) and (97). Using the coordinates $(u, v) \times \Pi \times \Pi^\perp = \mathbb{R}^d$ as before, it is sufficient to show that the longer sides of $\boldsymbol{P}^{t_l}_{i,\boldsymbol{q}}$ are transverse to Π. More precisely, if $\varepsilon_\circ$ is sufficiently small and N is large enough, there exists a constant $c > 0$, independent of $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$, such that, for

$$w \in \big(T_{\xi^{t_l}_{i,\boldsymbol{q}}}(\mathrm{N}^{-1}(\Pi)) \oplus \operatorname{span}\{\mathrm{N}(\xi^{t_l}_{i,\boldsymbol{q}})\}\big)^{\perp}, \tag{106}$$

(29) holds. Since $\psi(\zeta, t) = \frac{1}{2}|\zeta|^2 + t + \mathcal{R}$ with $\|\mathcal{R}\|_{C^N(I^d\times I)} \le \varepsilon_\circ$, by the same perturbation argument it is sufficient to consider $\psi(\zeta, t) = \frac{1}{2}|\zeta|^2 + t$. For this case (29) clearly holds for w satisfying (106) because translation by t doesn't have any effect. The same argument works without modification. □

3E. *Multiscale decomposition for $S_\delta f$.* In this section we obtain a multiscale decomposition for the square function, which is to be combined with multilinear square-function estimates to prove Proposition 3.2. This will be carried out in a way similar to how we obtained the decomposition in Section 2, though we need to take care of the additional t-average.

Let $0 < \varepsilon_\circ \ll 1$, $1 \ll N$, $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$, $\eta \in \mathcal{E}(N)$, and S_δ be given by (74). Let N^t, $\boldsymbol{n}$ be given by Definition 3.5. Let $\kappa = \kappa(\varepsilon_\circ, N)$ be the number given in Proposition 3.4 so that (77) holds whenever $0 < \varepsilon \le \kappa$, $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$, and $\eta \in \mathcal{E}(N)$. As before, let $\sigma_1, \dots, \sigma_m$, and $M_1, \dots, M_m$ be dyadic numbers such that

$$\delta \ll \sigma_{d-1} \ll \cdots \ll \sigma_1 \ll \min(\kappa, 1), \qquad M_i = \frac{1}{\sigma_i}. \tag{107}$$

We assume that f is Fourier supported in $\frac{1}{2}I^d$. We keep using the same notation as in Section 2E. In particular, $\{\mathfrak{q}^i\}$, $\{\mathfrak{Q}^i\}$ are the collection of (closed) dyadic intervals of side length $2\sigma_i$, $2M_i$, respectively, so that (37) and (40) holds.

3E1. *Decomposition by normal vector sets.* Let $\{\theta^i\}$ be a discrete subset of $\mathbb{S}^{d-1}$ whose elements are separated by distance $\sim \sigma_i$. Let $\mathfrak{d}^i$ be disjoint subsets of $\{\mathfrak{q}^i\}$ which satisfies, for some θ^i,

$$\mathfrak{d}^i \subset \{\mathfrak{q}^i : \operatorname{dist}(\boldsymbol{n}(\mathfrak{q}^i), \theta^i) \le C\sigma_i\} \tag{108}$$

and

$$\bigcup_{\mathfrak{d}^i} \mathfrak{d}^i = \{\mathfrak{q}^i\}, \quad i = 1, \dots, m. \tag{109}$$

Obviously, such a partitioning of $\{\mathfrak{q}^i\}$ is possible. Disjointness between $\mathfrak{d}^i$ will be useful later for decomposing the square function. Then we also define an auxiliary operator by

$$\mathfrak{S}_{\mathfrak{d}^i} f = \Big(\sum_{\mathfrak{q}^i \in \mathfrak{d}^i} |S_\delta f_{\mathfrak{q}^i}|^2\Big)^{\frac{1}{2}}.$$

As before, $\mathfrak{d}^i$, $\mathfrak{d}^i_*$, $\mathfrak{d}^i_j$, and $\mathfrak{d}^i_{j*}$ denote the elements in $\{\mathfrak{d}^i\}$ for the rest of this section.

Definition 3.13. We define $\boldsymbol{n}(\mathfrak{d}^i)$ to be a vector† $\theta \in \{\theta^i\}$ such that $\operatorname{dist}(\boldsymbol{n}(\mathfrak{q}^i), \theta) \le C\sigma_i$ whenever $\mathfrak{q}^i \in \mathfrak{d}^i$. Particularly, we may set $\boldsymbol{n}(\mathfrak{d}^i) = \theta^i$ if (108) holds.

†Possibly, there is more than one θ. In that case we simply choose one of them. Ambiguity of the definition does not cause any problem in what follows.

Since the map N^t is injective for each t, the elements of $\mathfrak{d}^i$ are contained in an $O(\sigma_i)$ neighborhood of the curve $\{\xi : \boldsymbol{n}(\xi) = \theta^i\}$ with $\theta^i = \boldsymbol{n}(\mathfrak{d}^i)$. From (72) we observe that for any interval J of length σ_i there are as many as $O(1)$ $\mathfrak{q}^i \in \mathfrak{d}^i$ such that $\phi\big((D_d - \psi(D', t))/\delta\big) f_{\mathfrak{q}^i} \neq 0$ if $t \in J$. Hence, dividing I intervals of length $\sim \sigma_i$ and taking integration in t we see that

$$S_\delta\Big(\sum_{\mathfrak{q}^i \in \mathfrak{d}^i} f_{\mathfrak{q}^i}\Big) \lesssim \Big(\sum_{\mathfrak{q}^i \in \mathfrak{d}^i} |S_\delta f_{\mathfrak{q}^i}|^2\Big)^{\frac{1}{2}} = \mathfrak{S}_{\mathfrak{d}^i} f \tag{110}$$

with the implicit constant independent of $\mathfrak{d}^i$. Since $S_\delta f \le \sum_{\mathfrak{d}^i} S_\delta\big(\sum_{\mathfrak{q}^i \in \mathfrak{d}^i} f_{\mathfrak{q}^i}\big)$, $i = 1, \dots, m$, we also have

$$S_\delta f \lesssim \sum_{\mathfrak{d}^i}\Big(\sum_{\mathfrak{q}^i \in \mathfrak{d}^i} |S_\delta f_{\mathfrak{q}^i}|^2\Big)^{\frac{1}{2}} = \sum_{\mathfrak{d}^i} \mathfrak{S}_{\mathfrak{d}^i} f. \tag{111}$$

3E2. *σ_1-scale decomposition.* Decomposition at this stage is similar to that of T_δ in Section 2. So, we shall be brief. Fix $x \in \mathbb{R}^d$ and let $\mathfrak{d}^1_* \in \{\mathfrak{d}^1\}$ such that

$$\mathfrak{S}_{\mathfrak{d}^1_*} f(x) = \max_{\mathfrak{d}^1} \mathfrak{S}_{\mathfrak{d}^1} f(x).$$

Considering the cases $\sum_{\mathfrak{d}^1} \mathfrak{S}_{\mathfrak{d}^1} f(x) \le 100^d \mathfrak{S}_{\mathfrak{d}^1_*} f(x)$ and $\sum_{\mathfrak{d}^1} \mathfrak{S}_{\mathfrak{d}^1} f(x) > 100^d \mathfrak{S}_{\mathfrak{d}^1_*} f(x)$ separately, we have

$$\begin{aligned} S_\delta f(x) &\lesssim \sum_{\mathfrak{d}^1} \mathfrak{S}_{\mathfrak{d}^1} f(x) \lesssim \mathfrak{S}_{\mathfrak{d}^1_*} f(x) + \sigma_1^{1-d} \max_{\substack{\mathfrak{d}^1 \\ |\boldsymbol{n}(\mathfrak{d}^1_*) - \boldsymbol{n}(\mathfrak{d}^1)| \gtrsim \sigma_1}} (\mathfrak{S}_{\mathfrak{d}^1_*} f(x) \mathfrak{S}_{\mathfrak{d}^1} f(x))^{\frac{1}{2}} \\ &\lesssim \mathfrak{S}_{\mathfrak{d}^1_*} f(x) + \sigma_1^{1-d} \max_{\substack{\mathfrak{d}^1_1, \mathfrak{d}^1_2 \\ |\boldsymbol{n}(\mathfrak{d}^1_1) - \boldsymbol{n}(\mathfrak{d}^1_2)| \gtrsim \sigma_1}} (\mathfrak{S}_{\mathfrak{d}^1_1} f(x) \mathfrak{S}_{\mathfrak{d}^1_2} f(x))^{\frac{1}{2}}. \end{aligned}$$

Since $\#\mathfrak{d}^i \lesssim \sigma_1^{-1}$ and $\mathfrak{S}_{\mathfrak{d}^1_1} f \mathfrak{S}_{\mathfrak{d}^1_2} f = \big(\sum_{\mathfrak{q}^1_1 \in \mathfrak{d}^1_1, \mathfrak{q}^1_2 \in \mathfrak{d}^1_2} (S_\delta f_{\mathfrak{q}^1_1} S_\delta f_{\mathfrak{q}^1_2})^2\big)^{\frac{1}{2}}$,

$$S_\delta f(x) \lesssim \sigma_1^{\frac{1}{p} - \frac{1}{2}} \Big(\sum_{\mathfrak{q}^1 \in \mathfrak{d}^1_*} |S_\delta f_{\mathfrak{q}^1}|^p\Big)^{\frac{1}{p}} + \sigma_1^{-C} \Big(\sum_{\substack{\mathfrak{d}^1_1, \mathfrak{d}^1_2 \\ |\boldsymbol{n}(\mathfrak{d}^1_1) - \boldsymbol{n}(\mathfrak{d}^1_2)| \gtrsim \sigma_1}} (S_\delta f_{\mathfrak{q}^1_1} S_\delta f_{\mathfrak{q}^1_2})^{\frac{p}{2}}\Big)^{\frac{1}{p}}.$$

Taking the L^p norm on both side of the inequality yields

$$\|S_\delta f\|_p \lesssim \sigma_1^{\frac{1}{p} - \frac{1}{2}} \Big(\sum_{\mathfrak{q}^1} \|S_\delta f_{\mathfrak{q}^1}\|_p^p\Big)^{\frac{1}{p}} + \sigma_1^{-C} \Big(\sum_{\mathfrak{q}^1_1, \mathfrak{q}^1_2 \text{ trans}} \|S_\delta f_{\mathfrak{q}^1_1} S_\delta f_{\mathfrak{q}^1_2}\|_{\frac{p}{2}}^{\frac{p}{2}}\Big)^{\frac{1}{p}}.$$

Hence, using Proposition 3.4 and Lemma 2.6, we have

$$\|S_\delta f\|_p \lesssim \sigma_1^{\frac{2}{p}} B_p(\sigma_1^{-2}\delta) \|f\|_p + \sigma_1^{-C} \max_{\mathfrak{q}^1_1, \mathfrak{q}^1_2 \text{ trans}} \|S_\delta f_{\mathfrak{q}^1_1} S_\delta f_{\mathfrak{q}^1_2}\|_{\frac{p}{2}}^{\frac{1}{2}}. \tag{112}$$

We proceed to decompose those terms appearing in the bilinear expression.

3E3. *σ_k-scale decomposition, $k \ge 2$.* Fixing σ, for $l \in \sigma^{-1}\mathbb{Z}^d$, let A_l and τ_l be given by (32). The following is a slight modification of Lemma 2.13.

Lemma 3.14. *Let $\mathfrak{d}$ be a subset of $\{\mathfrak{q}^i\}$. Set $\mathfrak{S}_{\mathfrak{d}} f = \big(\sum_{\mathfrak{q}^i \in \mathfrak{d}} |S_\delta f_{\mathfrak{q}^i}|^2\big)^{\frac{1}{2}}$, and set*

$$[\mathfrak{S}_{\mathfrak{d}} f] = \sum_{l \in M_i \mathbb{Z}^d} A_l^{\frac{1}{2}} \mathfrak{S}_{\mathfrak{d}}(\tau_l f), \quad |[\mathfrak{S}_{\mathfrak{d}} f]| = \sum_{l, l' \in M_i \mathbb{Z}^d} (A_l A_{l'})^{\frac{1}{2}} \mathfrak{S}_{\mathfrak{d}}(\tau_{(l+l')} f).$$

If $x, x_0 \in \mathfrak{Q}^i$, the following inequality holds with the implicit constants independent of $\mathfrak{d}$:

$$\mathfrak{S}_{\mathfrak{d}} f(x) \lesssim [\mathfrak{S}_{\mathfrak{d}} f](x_0) \lesssim |[\mathfrak{S}_{\mathfrak{d}} f]|(x). \tag{113}$$

Proof. Note that $\mathfrak{q}^i$ is a cube of side length $2\sigma_i$. Since $x, x_0 \in \mathfrak{Q}^i$, using (35) and the Cauchy–Schwarz inequality, we get

$$\left|\phi\Big(\frac{D_d - \psi(D', t)}{\delta}\Big) f_{\mathfrak{q}^i}(x)\right|^2 \lesssim \sum_{l \in M_i \mathbb{Z}^d} A_l \left|\phi\Big(\frac{\eta(D, t)(D_d - \psi(D', t))}{\delta}\Big) \tau_l f_{\mathfrak{q}^i}(x_0)\right|^2.$$

Integrating in t we get

$$(S_\delta f_{\mathfrak{q}^i}(x))^2 \lesssim \sum_{l \in M_i \mathbb{Z}^d} A_l (S_\delta(\tau_l f_{\mathfrak{q}^i})(x_0))^2. \tag{114}$$

Summation in $\mathfrak{q}^i \in \mathfrak{d}$ gives

$$\Big(\sum_{\mathfrak{q}^i \in \mathfrak{d}} (S_\delta f_{\mathfrak{q}^i}(x))^2\Big)^{\frac{1}{2}} \lesssim \sum_{l \in M_i \mathbb{Z}^d} A_l^{\frac{1}{2}} \Big(\sum_{\mathfrak{q}^i \in \mathfrak{d}} (S_\delta(\tau_l f_{\mathfrak{q}^i})(x_0))^2\Big)^{\frac{1}{2}},$$

from which we get the first inequality of (113). By interchanging the roles of x and x_0 in (114) and summation in $\mathfrak{q}^i \in \mathfrak{d}$, it follows that

$$\sum_{\mathfrak{q}^i \in \mathfrak{d}} (S_\delta(\tau_l f_{\mathfrak{q}^i})(x_0))^2 \lesssim \sum_{l \in M_i \mathbb{Z}^d} A_{l'} \sum_{\mathfrak{q}^i \in \mathfrak{d}} (S_\delta(\tau_{(l+l')} f_{\mathfrak{q}^i})(x))^2.$$

Putting this in the right-hand side of the above inequality and repeating the same argument, we get the second inequality of (113). □

Now we have the bilinear decomposition (112) on which we build a higher degree of multilinear decomposition.

3E4. *From k-transversal to $(k+1)$-transversal, $2 \le k \le m$.* Let us be given cubes $\mathfrak{q}_1^{k-1}, \mathfrak{q}_2^{k-1}, \dots, \mathfrak{q}_k^{k-1}$ of side length σ_{k-1} which satisfy (55). Though we use the same notation as in the multiplier-estimate case, it should be noted that the normal vector field $\boldsymbol{n}$ is defined on $I^{d-1} \times CI$ (see Definition 3.5). As before, we denote by $\{\mathfrak{q}_i^k\}$ the collection of dyadic cubes of side length σ_k contained in $\mathfrak{q}_i^{k-1}$, see (56), which are partitioned into the subsets of $\{\mathfrak{d}_i^k\}$ so that

$$\bigcup_{\mathfrak{d}_i^k} \Big(\bigcup_{\mathfrak{q}_i^k \in \mathfrak{d}_i^k} \mathfrak{q}_i^k\Big) = \mathfrak{q}_i^{k-1}, \quad i = 1, \dots, k.$$

So, we can write

$$\prod_{i=1}^k S_\delta\Big(\sum_{\mathfrak{q}_i^k \subset \mathfrak{q}_i^{k-1}} f_{\mathfrak{q}_i^{k-1}}\Big) = \prod_{i=1}^k S_\delta\Big(\sum_{\mathfrak{d}_i^k} \sum_{\mathfrak{q}_i^k \in \mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big)$$

and recall the definition

$$\mathfrak{S}_{\mathfrak{d}_i^k} F_{\mathfrak{q}_i^{k-1}} := \Big(\sum_{\mathfrak{q}_i^k \in \mathfrak{d}_j^k} |S_\delta F_{\mathfrak{q}_i^k}|^2\Big)^{\frac{1}{2}}.$$

Fix $\mathfrak{Q}^k$ and let x_0 be the center of $\mathfrak{Q}^k$. Let $\mathfrak{d}_{i*}^k \in \{\mathfrak{d}_i^k\}$ be an angular partition such that

$$\mathfrak{S}_{\mathfrak{d}_{i*}^k} f_{\mathfrak{q}_i^{k-1}}(x_0) = \max_{\mathfrak{d}_i^k} \mathfrak{S}_{\mathfrak{d}_i^k} f_{\mathfrak{q}_i^{k-1}}(x_0).$$

Let us set

$$\bar{\Lambda}_i^k = \{\mathfrak{d}_i^k : [\mathfrak{S}_{\mathfrak{d}_i^k} f_{\mathfrak{q}_i^{k-1}}](x_0) > (\sigma_k)^{kd} \max_{1\le j\le k} [\mathfrak{S}_{\mathfrak{d}_{j*}^k} f_{\mathfrak{q}_j^{k-1}}](x_0)\}, \quad 1 \le i \le k. \tag{115}$$

We split the sum to get

$$\prod_{i=1}^k S_\delta\Big(\sum_{\mathfrak{d}_i^k}\sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big) \le \prod_{i=1}^k S_\delta\Big(\sum_{\mathfrak{d}_i^k\in\bar{\Lambda}_i^k}\sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big) + \sum_{(\mathfrak{d}_1^k,\dots,\mathfrak{d}_k^k)\notin\prod_{i=1}^k \bar{\Lambda}_i^k} \prod_{i=1}^k S_\delta\Big(\sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big). \tag{116}$$

Thus, if $x \in \mathfrak{Q}^k$, by (113) and (110) the second term in the right-hand side is bounded by

$$\begin{aligned}
\sum_{(\mathfrak{d}_1^k,\dots,\mathfrak{d}_k^k)\notin\prod_{i=1}^k \bar{\Lambda}_i^k} \prod_{i=1}^k S_\delta\Big(\sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big)(x) &\lesssim \sum_{(\mathfrak{d}_1^k,\dots,\mathfrak{d}_k^k)\notin\prod_{i=1}^k \bar{\Lambda}_i^k} \prod_{i=1}^k [\mathfrak{S}_{\mathfrak{d}_i^k} f_{\mathfrak{q}_i^{k-1}}](x_0)\\
&\lesssim \big(\max_{1\le j\le k} [\mathfrak{S}_{\mathfrak{d}_{j*}^k} f_{\mathfrak{q}_j^{k-1}}](x_0)\big)^k\\
&\lesssim \big(\max_{1\le j\le k} [\mathfrak{S}_{\mathfrak{d}_{j*}^k} f](x_0)\big)^k \lesssim \big(\max_{\mathfrak{d}^k} |[\mathfrak{S}_{\mathfrak{d}^k} f]|(x)\big)^k.
\end{aligned} \tag{117}$$

Here $\{\mathfrak{d}^k\} = \bigcup_{1\le i\le k}\{\mathfrak{d}_i^k\}$ and the third inequality follows from the definition of $\mathfrak{S}_{\mathfrak{d}_j^k} f$ because $\mathfrak{q}_i^k \subset \mathfrak{q}_i^{k-1}$. Since (117) holds for each $\mathfrak{Q}^k$, integrating over all $\mathfrak{Q}^k$, using Lemma 3.14, Proposition 3.4 and Lemma 2.6, we get

$$\begin{aligned}
\Big\|\sum_{(\mathfrak{d}_1^k,\dots,\mathfrak{d}_k^k)\notin\prod_{i=1}^k \bar{\Lambda}_i^k} \prod_{i=1}^k S_\delta\Big(\sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big)\Big\|_{\frac{p}{k}}^{\frac{1}{k}} &\lesssim \big\|\max_{\mathfrak{d}^k} |[\mathfrak{S}_{\mathfrak{d}^k} f]|\big\|_p \lesssim \sup_h \big\|\max_{\mathfrak{d}^k} \mathfrak{S}_{\mathfrak{d}^k}(\tau_h f)\big\|_p\\
&\lesssim \sup_h \Big(\sum_{\mathfrak{d}^k} \|\mathfrak{S}_{\mathfrak{d}^k}(\tau_h f)\|_p^p\Big)^{\frac{1}{p}}\\
&\lesssim \sup_h \sigma_k^{(\frac{1}{p}-\frac{1}{2})} \Big(\sum_{\mathfrak{q}_i^k} \|S_\delta \tau_h f_{\mathfrak{q}_i^k}\|_p^p\Big)^{\frac{1}{p}}\\
&\lesssim \sigma_k^{\frac{2}{p}} B_p(\sigma_k^{-2}\delta)\|f\|_p.
\end{aligned} \tag{118}$$

The second-to-last inequality follows from the definition of $\mathfrak{S}_{\mathfrak{d}^k} f$ and Hölder's inequality since there are as many as $O(\sigma_k^{-1})$ $\mathfrak{q}^k \subset \mathfrak{d}^k$.

We note that vectors $\boldsymbol{n}(\mathfrak{d}_{1*}^k), \dots, \boldsymbol{n}(\mathfrak{d}_{k*}^k)$ are linearly independent because $\mathfrak{q}_1^{k-1}, \mathfrak{q}_2^{k-1}, \dots, \mathfrak{q}_k^{k-1}$ are transversal. We also denote by $\Pi_*^k = \Pi_*^k(\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}, \mathfrak{Q}^k)$ the k-plane spanned by the vectors $\boldsymbol{n}(\mathfrak{d}_{1*}^k), \dots, \boldsymbol{n}(\mathfrak{d}_{k*}^k)$. Let us set

$$\bar{\mathfrak{N}} = \bar{\mathfrak{N}}(\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}, \mathfrak{Q}^k) = \{\mathfrak{d}^k : \operatorname{dist}(\boldsymbol{n}(\mathfrak{d}^k), \Pi_*^k) \le C\sigma_k\}.$$

We split the sum and use the triangle inequality so that

$$\prod_{i=1}^{k} S_\delta\Big(\sum_{\mathfrak{d}_i^k\in\bar\Lambda_i^k}\sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big) \le \prod_{i=1}^{k} S_\delta\Big(\sum_{\substack{\mathfrak{d}_i^k\in\bar\Lambda_i^k\\ \mathfrak{d}_i^k\in\overline{\mathfrak{N}}}}\sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big) + \sum_{\substack{\mathfrak{d}_i^k\in\bar\Lambda_i^k\\ \mathfrak{d}_i^k\notin\overline{\mathfrak{N}}\text{ for some } i}} \prod_{i=1}^{k} S_\delta\Big(\sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big). \tag{119}$$

For the k-tuples $(\mathfrak{d}_1^k,\dots,\mathfrak{d}_k^k)$ appearing in the second summation of the right-hand side, there is a $\mathfrak{d}_i^k$ for which $\boldsymbol{n}(\mathfrak{d}_i^k)$ is not contained in $\Pi_*^k+O(\sigma_k)$. In particular, suppose that $\boldsymbol{n}(\mathfrak{d}_1^k)\notin\Pi_*^k+O(\sigma_k)$. Then, by (113) and (115) we have

$$\prod_{i=1}^{k} S_\delta\Big(\sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big)(x) \lesssim \prod_{i=1}^{k}[\mathfrak{S}_{\mathfrak{d}_i^k} f_{\mathfrak{q}_i^{k-1}}](x_0) \le \sigma_k^{-C}\big([\mathfrak{S}_{\mathfrak{d}_1^k} f_{\mathfrak{q}_1^{k-1}}](x_0)\big)^{\frac{k}{k+1}}\prod_{i=1}^{k}\big([\mathfrak{S}_{\mathfrak{d}_{i*}^k} f_{\mathfrak{q}_i^{k-1}}](x_0)\big)^{\frac{k}{k+1}}.$$

Recall that $\mathrm{Vol}(\boldsymbol{n}(\xi_1),\boldsymbol{n}(\xi_2),\dots,\boldsymbol{n}(\xi_k))\gtrsim\sigma_1\cdots\sigma_{k-1}$ if $\xi_i\in\mathfrak{q}_i^{k-1}$, $i=1,\dots,k$. From the definition of $\overline{\mathfrak{N}}$ it follows that $\mathrm{dist}(\boldsymbol{n}(\mathfrak{q}^k),\Pi_*^k)\gtrsim\sigma_k$ if $\mathfrak{q}^k\in\mathfrak{d}^k$ and $\boldsymbol{n}(\mathfrak{d}^k)\notin\overline{\mathfrak{N}}$. Hence

$$\mathrm{Vol}\big(\boldsymbol{n}(\xi_1),\boldsymbol{n}(\xi_2),\dots,\boldsymbol{n}(\xi_k),\boldsymbol{n}(\xi_{k+1})\big)\gtrsim\sigma_1\cdots\sigma_k$$

if $\xi_i\in\mathfrak{q}_i^k$ and $\mathfrak{q}_i^k\in\mathfrak{d}_{i*}^k$, $i=1,\dots,k$, and $\xi_{k+1}\in\mathfrak{q}_{k+1}^k$ and $\mathfrak{q}_{k+1}^k\in\mathfrak{d}_1^k$. So these cubes are transversal. Since there are only $O(\sigma_k^{-C})$ σ_k-scale cubes, by (113) and Hölder's inequality

$$\begin{aligned}\prod_{i=1}^{k} S_\delta\Big(\sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big)(x) &\lesssim \sigma_k^{-C}\big(|[\mathfrak{S}_{\mathfrak{d}_1^k} f_{\mathfrak{q}_1^{k-1}}]|(x)\big)^{\frac{k}{k+1}}\prod_{i=1}^{k}\big(|[\mathfrak{S}_{\mathfrak{d}_{i*}^k} f_{\mathfrak{q}_i^{k-1}}]|(x)\big)^{\frac{k}{k+1}}\\ &\lesssim \sigma_k^{-C}\sum_{l_1,l_1',\dots,l_{k+1},l_{k+1}'\in M_k\mathbb{Z}^d}\prod_{i=1}^{k+1}\widetilde{A}_{l_i}\widetilde{A}_{l_i'}\Big(\sum_{\mathfrak{q}_1^k,\dots,\mathfrak{q}_{k+1}^k\text{ trans}}\Big(\prod_{i=1}^{k+1}S_\delta(\tau_{(l_i+l_i')}f_{\mathfrak{q}_i^k})(x)\Big)^{\frac{p}{k+1}}\Big)^{\frac{k}{p}}.\end{aligned}$$

Here $\widetilde{A}_{l_i},\widetilde{A}_{l_i'}$ are rapidly decaying sequences. The same is true for any $\mathfrak{d}_1^k,\dots,\mathfrak{d}_k^k$ satisfying $\mathfrak{d}_i^k\in\bar\Lambda_i^k$, $1\le i\le k$, and $\mathfrak{d}_i^k\notin\overline{\mathfrak{N}}$ for some i and this holds regardless of $\mathfrak{Q}^k$. So, we have, for any x,

$$\begin{aligned}\sum_{\substack{\mathfrak{d}_i^k\in\bar\Lambda_i^k\\ \mathfrak{d}_i^k\notin\overline{\mathfrak{N}}\text{ for some } i}}\prod_{i=1}^{k} S_\delta\Big(\sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big)(x) &\lesssim \sigma_k^{-C}\sum_{l_1,l_1',\dots,l_{k+1},l_{k+1}'}\prod_{i=1}^{k+1}\widetilde{A}_{l_i}\widetilde{A}_{l_i'}\Big(\sum_{\mathfrak{q}_1^k,\dots,\mathfrak{q}_{k+1}^k\text{ trans}}\Big(\prod_{i=1}^{k+1}S_\delta(\tau_{(l_i+l_i')}f_{\mathfrak{q}_i^k})(x)\Big)^{\frac{p}{k+1}}\Big)^{\frac{k}{p}}.\end{aligned} \tag{120}$$

Since $\widetilde{A}_{l_i},\widetilde{A}_{l_i'}$ are rapidly decaying, taking the $L^{\frac{p}{k}}$ norm and a simple manipulation give

$$\Big\|\sum_{\substack{\mathfrak{d}_i^k\in\bar\Lambda_i^k\\ \mathfrak{d}_i^k\notin\overline{\mathfrak{N}}\text{ for some } i}}\prod_{i=1}^{k} S_\delta\Big(\sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big)\Big\|_{\frac{p}{k}} \lesssim \sigma_k^{-C}\sup_{\tau_1,\dots,\tau_{k+1}}\max_{\mathfrak{q}_1^k,\dots,\mathfrak{q}_{k+1}^k\text{ trans}}\Big\|\prod_{i=1}^{k+1}S_\delta(\tau_i f_{\mathfrak{q}_i^k})\Big\|_{\frac{p}{k+1}}^{\frac{k}{k+1}}. \tag{121}$$

We now combine the inequalities (116), (117), (119), (120) to get

$$\prod_{i=1}^{k} S_\delta\Big(\sum_{\mathfrak{d}_i^k}\sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big) \lesssim \big(\max_{\mathfrak{d}^k}|[\mathfrak{S}_{\mathfrak{d}^k} f]|(x)\big)^k + \chi_{\mathfrak{Q}^k}\prod_{i=1}^{k} S_\delta\Big(\sum_{\substack{\mathfrak{d}_i^k\in\bar{\Lambda}_i^k\\ \mathfrak{d}_i^k\in\bar{\mathfrak{N}}}}\ \sum_{\mathfrak{q}_i^k\in\mathfrak{d}_i^k} f_{\mathfrak{q}_i^k}\Big) + \sigma_k^{-C}\sum_{l_1,l_1',\dots,l_{k+1},l_{k+1}'}\prod_{i=1}^{k+1}\widetilde{A}_{l_i}\widetilde{A}_{l_i'}\bigg(\sum_{\mathfrak{q}_1^k,\dots,\mathfrak{q}_{k+1}^k\ \text{trans}}\Big(\prod_{i=1}^{k+1} S_\delta(\tau_{(l_i+l_i')}f_{\mathfrak{q}_i^k})(x)\Big)^{\frac{p}{k+1}}\bigg)^{\frac{k}{p}}.$$

Here $\bar{\mathfrak{N}}$ depends on $\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1},\mathfrak{Q}^k$. By taking the $\frac{1}{k}$-th power, integrating on $\mathbb{R}^d$ and using (118) and (121) we get

$$\bigg\|\Big(\prod_{i=1}^{k} S_\delta\Big(\sum_{\mathfrak{q}_i^k\subset\mathfrak{q}_i^{k-1}} f_{\mathfrak{q}_i^k}\Big)\Big)^{\frac{1}{k}}\bigg\|_p \lesssim \sigma_k^{\frac{2}{p}} B_p(\sigma_k^{-2}\delta)\|f\|_p + \sigma_k^{-C}\sup_{\tau_1,\dots,\tau_{k+1}}\ \max_{\mathfrak{q}_1^k,\dots,\mathfrak{q}_{k+1}^k\ \text{trans}}\Big\|\prod_{i=1}^{k+1} S_\delta(\tau_i f_{\mathfrak{q}_i^k})\Big\|_{\frac{p}{k+1}}^{\frac{k}{k+1}} + \bigg(\sum_{\mathfrak{Q}^k}\Big\|\prod_{i=1}^{k} S_\delta\Big(\sum_{\substack{\mathfrak{d}_i^k\in\\ [\bar{\mathfrak{N}}](\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1},\mathfrak{Q}^k)}}\ \sum_{\substack{\mathfrak{q}_i^k\in\mathfrak{d}_i^k\\ \mathfrak{q}_i^k\subset\mathfrak{q}_i^{k-1}}} f_{\mathfrak{q}_i^k}\Big)\Big\|_{L^{p/k}(\mathfrak{Q}^k)}^{\frac{p}{k}}\bigg)^{\frac{1}{p}}, \tag{122}$$

where $[\bar{\mathfrak{N}}](\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1},\mathfrak{Q}^k)$ denotes a subset of $\bar{\mathfrak{N}}(\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1},\mathfrak{Q}^k)$ which depends on $\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1},\mathfrak{Q}^k$.

3E5. *Multiscale decomposition.* For $k=2,\dots,m$, let us set

$$\bar{\mathfrak{M}}^k f = \sup_{\tau_1,\dots,\tau_k}\ \max_{\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1}\ \text{trans}}\bigg(\sum_{\mathfrak{Q}^k}\Big\|\prod_{i=1}^{k} S_\delta\Big(\sum_{\substack{\mathfrak{d}_i^k\in\\ [\bar{\mathfrak{N}}](\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1},\mathfrak{Q}^k)}}\ \sum_{\substack{\mathfrak{q}_i^k\in\mathfrak{d}_i^k\\ \mathfrak{q}_i^k\subset\mathfrak{q}_i^{k-1}}} \tau_i f_{\mathfrak{q}_i^k}\Big)\Big\|_{L^{p/k}(\mathfrak{Q}^k)}^{\frac{p}{k}}\bigg)^{\frac{1}{p}}.$$

Here $[\bar{\mathfrak{N}}](\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1},\mathfrak{Q}^k)$ also depends on $\tau_1,\dots,\tau_k$ but this doesn't affect the overall bound. Starting from (112) we successively apply (122) to k-scale transversal products (given by $\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1}$ transversal). After decomposition up to the m-th scale we get

$$\|S_\delta f\|_p \lesssim \sum_{k=1}^{m}\sigma_{k-1}^{-C}\sigma_k^{\frac{2}{p}} B_p(\sigma_k^{-2}\delta)\|f\|_p + \sum_{k=2}^{m}\sigma_{k-1}^{-C}\bar{\mathfrak{M}}^k f + \sigma_m^{-C}\sup_{\tau_1,\dots,\tau_{m+1}}\ \max_{\mathfrak{q}_1^m,\dots,\mathfrak{q}_{m+1}^m\ \text{trans}}\Big\|\prod_{i=1}^{m+1} S_\delta\tau_i f_{\mathfrak{q}_i^m}\Big\|_{L^{p/(m+1)}}^{\frac{1}{m+1}}. \tag{123}$$

3F. ***Proof of Proposition 3.2.*** We may assume $d\ge 9$ since $p_s\ge 2(d+2)/d$ for $d<9$ and the sharp bound for $p\ge 2(d+2)/d$ is verified in [Lee et al. 2012]. So, we have $p_s(d)\ge 2(d-1)/(d-2)$. The proof is similar to that of Proposition 2.3. Let $\beta>0$ and we aim to show that $\mathcal{B}^\beta(s)\le C$ for $0<s\le 1$ if $p\ge p_s(d)$. We choose $\varepsilon>0$ such that $(100d)^{-1}\beta\ge\varepsilon$. Fix $\varepsilon_\circ>0$ and $N=N(\varepsilon)$ such that Corollaries 3.8, 3.11 and 3.12 hold uniformly for $\psi\in\bar{\mathfrak{G}}(\varepsilon_\circ,N)$.

Let $s < \delta \le 1$. Obviously, $(\sigma_k^{-2}\delta)^{\frac{d-2}{2}-\frac{d}{p}+\beta} B(\sigma_k^{-2}\delta) \le \mathcal{B}^\beta(s) + \sigma_k^{-C}$ because $s \le \sigma_k^{-2}\delta$ and $B(\delta) = B_p(\delta) \le C$ for $\delta \gtrsim 1$. Hence, it follows that

$$\sigma_k^{\frac{2}{p}} B(\sigma_k^{-2}\delta) \lesssim \sigma_k^{2(\frac{d-2}{2}-\frac{d-1}{p})+2\beta} \delta^{-\frac{d-2}{2}+\frac{d}{p}-\beta} (\mathcal{B}^\beta(s) + \sigma_k^{-C}). \tag{124}$$

We first consider the $(m+1)$-product in (123). By Corollary 3.8 we have, for $p \ge 2(m+1)/m$,

$$\sup_{\tau_1,\dots,\tau_{m+1}} \max_{\mathfrak{q}_1^m,\dots,\mathfrak{q}_{m+1}^m \text{ trans}} \Big\| \prod_{i=1}^{m+1} S_\delta \tau_i f_{\mathfrak{q}_i^m} \Big\|_{L^{p/(m+1)}}^{\frac{1}{m+1}} \le C_\varepsilon \sigma_m^{-C} \delta^{-\frac{d-2}{2}+\frac{d}{p}-\varepsilon} \|f\|_p. \tag{125}$$

For $\overline{\mathfrak{M}}^k$, as before we have two types of estimates. The first one follows from Corollary 3.8, while the second one is a consequence of the square-function estimates in Corollary 3.12. From the definition of $\overline{\mathfrak{M}}^k$, we note that $\mathfrak{q}_1^k, \mathfrak{q}_2^k, \dots, \mathfrak{q}_k^k$ are contained, respectively, in $\mathfrak{q}_1^{k-1}, \mathfrak{q}_2^{k-1}, \dots, \mathfrak{q}_k^{k-1}$, which are transversal. Hence, we have

$$\prod_{i=1}^k S_\delta \Big(\sum_{\mathfrak{d}_i^k \in [\overline{\mathfrak{N}}](\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1},\mathfrak{Q}^k)} \sum_{\substack{\mathfrak{q}_i^k \in \mathfrak{d}_i^k \\ \mathfrak{q}_i^k \subset \mathfrak{q}_i^{k-1}}} \tau_i f_{\mathfrak{q}_i^k} \Big)(x) \le \sum_{\mathfrak{q}_1^k,\mathfrak{q}_2^k,\dots,\mathfrak{q}_k^k \text{ trans}} \prod_{i=1}^k S_\delta(\tau_i f_{\mathfrak{q}_i^k})(x).$$

Here "$\mathfrak{q}_1^k, \mathfrak{q}_2^k, \dots, \mathfrak{q}_k^k$ trans" means $\mathrm{Vol}(\boldsymbol{n}(\xi_1), \dots, \boldsymbol{n}(\xi_k)) \ge \sigma_1 \cdots \sigma_{k-1}$ provided $\xi_i \in \mathfrak{q}_i^k$, $i = 1, \dots, k$. Since there are as many as $O(\sigma_{k-1}^{-C})$ k-tuples $(\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1})$ and the above holds regardless of $\mathfrak{Q}^k$, by Corollary 3.12 we have, for $p \ge 2k/(k-1)$,

$$\overline{\mathfrak{M}}^k f \lesssim \sigma_k^{-C} \sup_{\tau_1,\dots,\tau_k} \sum_{\mathfrak{q}_1^k,\mathfrak{q}_2^k,\dots,\mathfrak{q}_k^k \text{ trans}} \Big\| \prod_{i=1}^k S_\delta(\tau_i f_{\mathfrak{q}_i^k}) \Big\|_{L^{p/k}} \lesssim \sigma_k^{-C} \delta^{-\frac{d-2}{2}+\frac{d}{p}-\varepsilon} \|f\|_p. \tag{126}$$

Estimates for $\overline{\mathfrak{M}}^k$ via Corollary 3.11. By fixing $\tau_1, \dots, \tau_k$, and $(\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1})$ satisfying $\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}$ are transversal, we first handle the integral over $\mathfrak{Q}^k$ which is in the definition of $\overline{\mathfrak{M}}^k$. For $i = 1, \dots, k$, set

$$f_i = \sum_{\mathfrak{d}_i^k \in [\overline{\mathfrak{N}}](\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1},\mathfrak{Q}^k)} \Big(\sum_{\substack{\mathfrak{q}_i^k \in \mathfrak{d}_i^k \\ \mathfrak{q}_i^k \subset \mathfrak{q}_i^{k-1}}} \tau_i f_{\mathfrak{q}_i^k} \Big).$$

Since $\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}$ are transversal, (82) holds with $\sigma = \sigma_1 \cdots \sigma_{k-1}$ whenever $\xi_i \in \operatorname{supp} \hat{f}_i + O(\sigma_k)$, $i = 1, 2, \dots, k$. Also note that $\boldsymbol{n}(\mathfrak{d}_1^k), \dots, \boldsymbol{n}(\mathfrak{d}_k^k) \subset \Pi_*^k(\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}, \mathfrak{Q}^k)$. Hence, it follows that (102) holds with $\tilde\sigma = \sigma_k$. Let us set

$$\mathcal{Q}(\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}, \mathfrak{Q}^k) = \{\mathfrak{q}^k : \boldsymbol{n}(\mathfrak{q}^k) \in [\overline{\mathfrak{N}}](\mathfrak{q}_1^{k-1}, \dots, \mathfrak{q}_k^{k-1}, \mathfrak{Q}^k)\}.$$

Let write $\mathfrak{Q}^k = \mathfrak{q}(z, 1/\sigma_k)$. Then, by Corollary 3.12 we have, for $2 \le p \le 2k/(k-1)$,

$$\Big\| \Big(\prod_{i=1}^k S_\delta f_i \Big)^{\frac{1}{k}} \Big\|_{L^p(\mathfrak{Q}^k)}^p \lesssim \sigma_{k-1}^{-C_\varepsilon} \sigma_k^{-\varepsilon} \prod_{i=1}^k \Big\| \Big(\sum_{\substack{\mathfrak{q}_i^k \in \mathfrak{q}_{k-1}^i \\ \mathfrak{q}_i^k \in \mathcal{Q}(\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1},\mathfrak{Q}^k)}} |S_\delta \tau_i f_{\mathfrak{q}^k}|^2 \Big)^{\frac{1}{2}} \rho_{B(z,\frac{C}{\sigma_k})} \Big\|_{L^p}^{\frac{p}{k}}.$$

The dyadic cubes of side length σ_k in $\mathcal{Q}(\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1},\mathfrak{Q}^k)$ are contained in an $O(\sigma_k)$-neighborhood of $\boldsymbol{n}^{-1}(\Pi_*^k)$ which is a smooth k-dimensional surface. Thus,

$$\#\{\mathfrak{q}_i^k\subset\mathfrak{q}_{k-1}^i:\mathfrak{q}_i^k\in\mathcal{Q}(\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1},\mathfrak{Q}^k)\}\le C\sigma_k^{-k}.$$

Now, by Hölder's inequality we get

$$\Big\|\Big(\prod_{i=1}^k S_\delta f_i\Big)^{\frac1k}\Big\|_{L^p(\mathfrak{Q}^k)}^p\lesssim\sigma_{k-1}^{-C_\varepsilon}\sigma_k^{-\varepsilon-k(\frac p2-1)}\prod_{i=1}^k\Big\|\Big(\sum_{\mathfrak{q}_i^k\subset\mathfrak{q}_i^{k-1}}|S_\delta\tau_i f_{\mathfrak{q}_i^k}|^p\Big)^{\frac1p}\rho_{\mathfrak{Q}^k}\Big\|_{L^p}^{\frac pk}.$$

Summation along $\mathfrak{Q}^k$ using the rapid decay of the Schwartz function ρ gives

$$\Big\|\Big(\prod_{i=1}^k S_\delta f_i\Big)^{\frac1k}\Big\|_{L^p}\lesssim\sigma_{k-1}^{-C_\varepsilon}\sigma_k^{-\varepsilon-k(\frac12-\frac1p)}\prod_{i=1}^k\Big\|\Big(\sum|S_\delta\tau_i f_{\mathfrak{q}_i^k}|^p\Big)^{\frac1p}\Big\|_p^{\frac1k}.$$

Hence, using Proposition 3.4, Lemma 2.6, and (124), for $2\le p\le 2k/(k-1)$, we have

$$\begin{aligned}\Big\|\Big(\prod_{i=1}^k S_\delta f_i\Big)^{\frac1k}\Big\|_{L^p}&\lesssim\sigma_{k-1}^{-C}\sigma_k^{-\varepsilon-\frac{k-1}{2}+\frac{k+1}{p}}B(\sigma_k^{-2}\delta)\|f\|_p\\&\lesssim\sigma_{k-1}^{-C}\delta^{-\frac{d-2}{2}+\frac dp-\beta}\sigma_k^{\beta+\frac{2d-k-3}{2}-\frac{2d-k-1}{p}}(\sigma_k^{-C}+\mathcal{B}^\beta(s))\|f\|_p\\&\lesssim\sigma_{k-1}^{-C}\delta^{-\frac{d-2}{2}+\frac dp-\beta}(\sigma_k^{-C}+\sigma_k^\alpha\mathcal{B}^\beta(s))\|f\|_p\end{aligned}$$

with some $\alpha>0$ if $p\ge 2(2d-k-1)/(2d-k-3)$. Here we have used $(100d)^{-1}\beta\ge\varepsilon$. We note that the right-hand side of the above is independent of $\tau_1,\dots,\tau_k$ and there are only $O(\sigma_{k-1}^{-C})$ many k-tuples $(\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1})$ satisfying $\mathfrak{q}_1^{k-1},\dots,\mathfrak{q}_k^{k-1}$ are transversal. Thus, recalling the definition of $\overline{\mathfrak{M}^k f}$, we have for $2\le p\le 2k/(k-1)$

$$\overline{\mathfrak{M}^k f}\lesssim\sigma_{k-1}^{-C}\delta^{-\frac{d-2}{2}+\frac dp-\beta}\big(\sigma_k^{-C}+\sigma_k^\alpha\mathcal{B}^\beta(s)\big)\|f\|_p$$

with some $\alpha>0$ provided that $p\ge 2(2d-k-1)/(2d-k-3)$. Combining this and (126) we have, for some $\alpha>0$,

$$\overline{\mathfrak{M}}^k f\le C\delta^{-\frac{d-2}{2}+\frac dp-\beta}\big(\sigma_k^{-C}+\sigma_k^\alpha\mathcal{B}^\beta(s)\big)\|f\|_p. \tag{127}$$

provided that $p\ge\min\big(2(2d-k-1)/(2d-k-3),2k/(k-1)\big)$.

Closing induction. Let us set

$$p(m)=\max\Big(\max_{1\le k\le m}\min\Big(\frac{2(2d-k-1)}{2d-k-3},\frac{2k}{k-1}\Big),\frac{2(m+1)}{m}\Big).$$

Since $p\ge p_s>2(d-1)/(d-2)$ and $(100d)^{-1}\beta\ge\varepsilon$, we have

$$\sigma_k^{\frac2p}B(\sigma_k^{-2}\delta)\lesssim\sigma_k^\alpha\delta^{-\frac{d-2}{2}+\frac dp-\beta}(\mathcal{B}^\beta(s)+\sigma_k^{-C})$$

for some $\alpha>0$. Using (123), we combine the estimates (124), (125), and (127) to get

$$\|S_\delta f\|_p\le C\sum_{k=1}^m(\sigma_{k-1}^{-C}+\sigma_{k-1}^{-C}\sigma_k^\alpha\mathcal{B}^\beta(s))\delta^{-\frac{d-2}{2}+\frac dp-\beta}\|f\|_p+C\sigma_m^{-C}\delta^{-\frac{d-2}{2}+\frac dp-\beta}\|f\|_p$$

for some $\alpha > 0$ as long as $p \ge p(m)$. The rest of proof is similar to that in Section 2F, so we intend to be brief. By using the stability of the estimates along $\psi \in \overline{\mathfrak{G}}(\varepsilon_\circ, N)$, $\eta \in \mathcal{E}(N)$, multiplying by $\delta^{\frac{d-2}{2}-\frac{d}{p}+\beta}$ on both sides and taking the supremum along ψ, η and f, and taking the supremum along δ, $s < \delta \le 1$, we get

$$\mathcal{B}^\beta(s) \le C\Big(\sum_{k=1}^m \sigma_{k-1}^{-C}\sigma_k^\alpha\Big)\mathcal{B}^\beta(s) + C\sum_{k=1}^m \sigma_k^{-C}$$

for some $\alpha > 0$ provided that $p \ge p(m)$. Choosing $\sigma_1, \dots, \sigma_{m-1}$ such that $C\big(\sum_{k=1}^{m-1}\sigma_{k-1}^{-C}\sigma_k^\alpha\big) \le \frac{1}{2}$ gives $\mathcal{B}^\varepsilon(\delta) \le C\sigma_m^{-C}$ for $p \ge p(m)$. Therefore, to complete the proof we need only to check that the minimum of $p(m)$, $2 \le m \le d-1$, is p_s. This can be done by a simple computation.

Acknowledgements

S. Lee was partially supported by NRF (Republic of Korea) grant No. 2015R1A2A2A05000956. He would like thank Andreas Seeger for discussions on related problems.

Note in proof

The range of sharp L^p bound for the Bochner–Riesz means in Theorem 1.1 was recently improved by Guth, Hickman and Iliopoulou [Guth et al. 2017] for $d \ge 4$.

References

[Annoni 2017] M. Annoni, "Almost everywhere convergence for modified Bochner–Riesz means at the critical index for $p \ge 2$", *Pacific J. Math.* **286**:2 (2017), 257–275. MR Zbl

[Bejenaru 2016] I. Bejenaru, "Optimal multilinear restriction estimates for a class of surfaces with curvature", preprint, 2016. arXiv

[Bejenaru 2017] I. Bejenaru, "The optimal trilinear restriction estimate for a class of hypersurfaces with curvature", *Adv. Math.* **307** (2017), 1151–1183. MR Zbl

[Bennett 2014] J. Bennett, "Aspects of multilinear harmonic analysis related to transversality", pp. 1–28 in *Harmonic analysis and partial differential equations*, edited by P. Cifuentes et al., Contemp. Math. **612**, Amer. Math. Soc., Providence, RI, 2014. MR Zbl

[Bennett et al. 2006] J. Bennett, A. Carbery, and T. Tao, "On the multilinear restriction and Kakeya conjectures", *Acta Math.* **196**:2 (2006), 261–302. MR Zbl

[Bennett et al. 2015] J. Bennett, N. Bez, T. C. Flock, and S. Lee, "Stability of the Brascamp–Lieb constant and applications", preprint, 2015. arXiv

[Bourgain 1991a] J. Bourgain, "Besicovitch type maximal operators and applications to Fourier analysis", *Geom. Funct. Anal.* **1**:2 (1991), 147–187. MR Zbl

[Bourgain 1991b] J. Bourgain, "On the restriction and multiplier problems in $\mathbb{R}^3$", pp. 179–191 in *Geometric aspects of functional analysis (1989–90)*, edited by V. D. Milman and J. Lindenstrauss, Lecture Notes in Math. **1469**, Springer, 1991. MR Zbl

[Bourgain 1991c] J. Bourgain, "L^p-estimates for oscillatory integrals in several variables", *Geom. Funct. Anal.* **1**:4 (1991), 321–374. MR Zbl

[Bourgain 2013] J. Bourgain, "On the Schrödinger maximal function in higher dimension", *Tr. Mat. Inst. Steklova* **280** (2013), 53–66. In Russian; translated in *Proc. Steklov Inst. Math.* **280**:1 (2013), 46–60. 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 and Guth 2011] J. Bourgain and L. Guth, "Bounds on oscillatory integral operators based on multilinear estimates", *Geom. Funct. Anal.* **21**:6 (2011), 1239–1295. MR Zbl

[Carbery 1983] A. Carbery, "The boundedness of the maximal Bochner–Riesz operator on $L^4(\mathbb{R}^2)$", *Duke Math. J.* **50**:2 (1983), 409–416. MR Zbl

[Carbery 1985] A. Carbery, "Radial Fourier multipliers and associated maximal functions", pp. 49–56 in *Recent progress in Fourier analysis* (El Escorial, 1983), edited by I. Peral and J. L. Rubio de Francia, North-Holland Math. Stud. **111**, North-Holland, Amsterdam, 1985. MR Zbl

[Carbery and Valdimarsson 2013] A. Carbery and S. I. Valdimarsson, "The endpoint multilinear Kakeya theorem via the Borsuk–Ulam theorem", *J. Funct. Anal.* **264**:7 (2013), 1643–1663. MR Zbl

[Carbery et al. 1984] A. Carbery, G. Gasper, and W. Trebels, "Radial Fourier multipliers of $L^p(\mathbb{R}^2)$", *Proc. Nat. Acad. Sci. U.S.A.* **81**:10 (1984), 3254–3255. MR

[Carbery et al. 1988] A. Carbery, J. L. Rubio de Francia, and L. Vega, "Almost everywhere summability of Fourier integrals", *J. London Math. Soc.* (2) **38**:3 (1988), 513–524. MR Zbl

[Carleson and Sjölin 1972] L. Carleson and P. Sjölin, "Oscillatory integrals and a multiplier problem for the disc", *Studia Math.* **44** (1972), 287–299. MR Zbl

[Christ 1985] M. Christ, "On almost everywhere convergence of Bochner–Riesz means in higher dimensions", *Proc. Amer. Math. Soc.* **95**:1 (1985), 16–20. MR Zbl

[Christ 1987] M. Christ, "Weak type endpoint bounds for Bochner–Riesz multipliers", *Rev. Mat. Iberoamericana* **3**:1 (1987), 25–31. MR Zbl

[Christ 1988] M. Christ, "Weak type $(1, 1)$ bounds for rough operators", *Ann. of Math.* (2) **128**:1 (1988), 19–42. MR Zbl

[Du et al. 2017] X. Du, L. Guth, and X. Li, "A sharp Schrödinger maximal estimate in $\mathbb{R}^2$", *Ann. of Math.* (2) **186**:2 (2017), 607–640. MR Zbl

[Fefferman 1970] C. Fefferman, "Inequalities for strongly singular convolution operators", *Acta Math.* **124** (1970), 9–36. MR Zbl

[Fefferman 1971] C. Fefferman, "The multiplier problem for the ball", *Ann. of Math.* (2) **94** (1971), 330–336. MR Zbl

[Fefferman 1973] C. Fefferman, "A note on spherical summation multipliers", *Israel J. Math.* **15** (1973), 44–52. MR Zbl

[Rubio de Francia 1985] J. L. Rubio de Francia, "A Littlewood–Paley inequality for arbitrary intervals", *Rev. Mat. Iberoamericana* **1**:2 (1985), 1–14. MR Zbl

[Guth 2010] L. Guth, "The endpoint case of the Bennett–Carbery–Tao multilinear Kakeya conjecture", *Acta Math.* **205**:2 (2010), 263–286. MR Zbl

[Guth 2016a] L. Guth, "A restriction estimate using polynomial partitioning", *J. Amer. Math. Soc.* **29**:2 (2016), 371–413. MR Zbl

[Guth 2016b] L. Guth, "Restriction estimates using polynomial partitioning, II", preprint, 2016. arXiv

[Guth et al. 2017] L. Guth, J. Hickman, and M. Iliopoulou, "Sharp estimates for oscillatory integral operators via polynomial partitioning", preprint, 2017. arXiv

[Hörmander 1973] L. Hörmander, "Oscillatory integrals and multipliers on FL^p", *Ark. Mat.* **11** (1973), 1–11. MR Zbl

[Lee 2004] S. Lee, "Improved bounds for Bochner–Riesz and maximal Bochner–Riesz operators", *Duke Math. J.* **122**:1 (2004), 205–232. MR Zbl

[Lee 2006a] S. Lee, "Linear and bilinear estimates for oscillatory integral operators related to restriction to hypersurfaces", *J. Funct. Anal.* **241**:1 (2006), 56–98. MR Zbl

[Lee 2006b] S. Lee, "On pointwise convergence of the solutions to Schrödinger equations in $\mathbb{R}^2$", *Int. Math. Res. Not.* **2006** (2006), art. id. 32597. MR Zbl

[Lee and Seeger 2015] S. Lee and A. Seeger, "On radial Fourier multipliers and almost everywhere convergence", *J. Lond. Math. Soc.* (2) **91**:1 (2015), 105–126. MR Zbl

[Lee and Vargas 2008] S. Lee and A. Vargas, "Sharp null form estimates for the wave equation", *Amer. J. Math.* **130**:5 (2008), 1279–1326. MR Zbl

[Lee and Vargas 2010] S. Lee and A. Vargas, "Restriction estimates for some surfaces with vanishing curvatures", *J. Funct. Anal.* **258**:9 (2010), 2884–2909. MR Zbl

[Lee and Vargas 2012] S. Lee and A. Vargas, "On the cone multiplier in $\mathbb{R}^3$", *J. Funct. Anal.* **263**:4 (2012), 925–940. MR Zbl

[Lee et al. 2012] S. Lee, K. M. Rogers, and A. Seeger, "Improved bounds for Stein's square functions", *Proc. Lond. Math. Soc.* (3) **104**:6 (2012), 1198–1234. MR Zbl

[Lee et al. 2013] S. Lee, K. M. Rogers, and A. Seeger, "On space-time estimates for the Schrödinger operator", *J. Math. Pures Appl.* (9) **99**:1 (2013), 62–85. MR Zbl

[Lee et al. 2014] S. Lee, K. M. Rogers, and A. Seeger, "Square functions and maximal operators associated with radial Fourier multipliers", pp. 273–302 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

[Ou and Wang 2017] Y. Ou and H. Wang, "A cone restriction estimate using polynomial partitioning", preprint, 2017. arXiv

[Ramos 2016] J. Ramos, "The trilinear restriction estimate with sharp dependence on the transversality", preprint, 2016. arXiv

[Seeger 1986] A. Seeger, "On quasiradial Fourier multipliers and their maximal functions", *J. Reine Angew. Math.* **370** (1986), 61–73. MR Zbl

[Seeger 1996] A. Seeger, "Endpoint inequalities for Bochner–Riesz multipliers in the plane", *Pacific J. Math.* **174**:2 (1996), 543–553. MR Zbl

[Shayya 2017] B. Shayya, "Weighted restriction estimates using polynomial partitioning", *Proc. Lond. Math. Soc.* (3) **115**:3 (2017), 545–598. MR Zbl

[Stein 1958] E. M. Stein, "Localization and summability of multiple Fourier series", *Acta Math.* **100** (1958), 93–147. MR Zbl

[Stein 1986] E. M. Stein, "Oscillatory integrals in Fourier analysis", pp. 307–355 in *Beijing lectures in harmonic analysis* (Beijing, 1984), edited by E. M. Stein, Ann. of Math. Stud. **112**, Princeton Univ. Press, 1986. MR Zbl

[Stein 1993] E. M. Stein, *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*, Princeton Mathematical Series **43**, Princeton University Press, 1993. MR Zbl

[Sunouchi 1967] G. Sunouchi, "On the Littlewood–Paley function g^* of multiple Fourier integrals and Hankel multiplier transformations", *Tôhoku Math. J.* (2) **19** (1967), 496–511. MR Zbl

[Tao 1996] T. Tao, "Weak-type endpoint bounds for Riesz means", *Proc. Amer. Math. Soc.* **124**:9 (1996), 2797–2805. MR Zbl

[Tao 1998] T. Tao, "The weak-type endpoint Bochner–Riesz conjecture and related topics", *Indiana Univ. Math. J.* **47**:3 (1998), 1097–1124. MR Zbl

[Tao 1999] T. Tao, "The Bochner–Riesz conjecture implies the restriction conjecture", *Duke Math. J.* **96**:2 (1999), 363–375. MR Zbl

[Tao 2002] T. Tao, "On the maximal Bochner–Riesz conjecture in the plane for $p < 2$", *Trans. Amer. Math. Soc.* **354**:5 (2002), 1947–1959. MR Zbl

[Tao 2003] T. Tao, "A sharp bilinear restrictions estimate for paraboloids", *Geom. Funct. Anal.* **13**:6 (2003), 1359–1384. MR Zbl

[Tao and Vargas 2000a] T. Tao and A. Vargas, "A bilinear approach to cone multipliers, I: Restriction estimates", *Geom. Funct. Anal.* **10**:1 (2000), 185–215. MR Zbl

[Tao and Vargas 2000b] T. Tao and A. Vargas, "A bilinear approach to cone multipliers, II: Applications", *Geom. Funct. Anal.* **10**:1 (2000), 216–258. MR Zbl

[Tao et al. 1998] T. Tao, A. Vargas, and L. Vega, "A bilinear approach to the restriction and Kakeya conjectures", *J. Amer. Math. Soc.* **11**:4 (1998), 967–1000. MR Zbl

[Temur 2014] F. Temur, "A Fourier restriction estimate for surfaces of positive curvature in $\mathbb{R}^6$", *Rev. Mat. Iberoam.* **30**:3 (2014), 1015–1036. MR Zbl

[Tomas 1975] P. A. Tomas, "A restriction theorem for the Fourier transform", *Bull. Amer. Math. Soc.* **81** (1975), 477–478. MR Zbl

[Wisewell 2005] L. Wisewell, "Kakeya sets of curves", *Geom. Funct. Anal.* **15**:6 (2005), 1319–1362. MR Zbl

[Wolff 2001] T. Wolff, "A sharp bilinear cone restriction estimate", *Ann. of Math.* (2) **153**:3 (2001), 661–698. MR Zbl

[Zhang 2015] R. Zhang, "The endpoint perturbed Brascamp–Lieb inequality with examples", preprint, 2015. arXiv

Received 4 Oct 2017. Accepted 12 Jan 2018.

SANGHYUK LEE: shklee@snu.ac.kr
School of Mathematical Sciences, Seoul National University, Seoul, South Korea

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. 6 2018