SPHERICAL MAXIMAL FUNCTIONS ON TWO STEP NILPOTENT LIE GROUPS

JAEHYEON RYU ANDREAS SEEGER

Abstract. Consider $\mathbb{R}^d \times \mathbb{R}^m$ with the group structure of a two-step nilpotent Lie group and natural parabolic dilations. The maximal function originally introduced by Nevo and Thangavelu in the setting of the Heisenberg group deals with noncommutative convolutions associated to measures on spheres or generalized spheres in $\mathbb{R}^d$. We drop the nondegeneracy assumptions in the known results on Métivier groups and prove the sharp L^p boundedness result for all two step nilpotent Lie groups with $d \geq 3$.

1. Introduction

We consider the problem of bounding maximal operators for averages over spheres with higher codimension on a two-step nilpotent Lie group G which was introduced for the special case of the Heisenberg group by Nevo and Thangavelu [16]. The setup is as follows: The Lie algebra splits as a direct sum in two subspaces referred to as the horizontal and the vertical part, $\mathfrak{g} = \mathfrak{w}_{\mathrm{hor}} \oplus \mathfrak{w}_{\mathrm{vert}}$, where $\dim \mathfrak{w}_{\mathrm{hor}} = d$, $\dim \mathfrak{w}_{\mathrm{vert}} = m$, and $\mathfrak{w}_{\mathrm{vert}} \subseteq \mathfrak{z}(\mathfrak{g})$, with $\mathfrak{z}(\mathfrak{g})$ the center of the Lie algebra. We use the natural parabolic dilation structure on $\mathfrak{w}_{\mathrm{hor}} \oplus \mathfrak{w}_{\mathrm{vert}}$, and define for $\underline{X} \in \mathfrak{w}_{\mathrm{hor}}$, $\overline{X} \in \mathfrak{w}_{\mathrm{vert}}$, $\delta_t(\underline{X}, \overline{X}) = (t\underline{X}, t^2\overline{X})$. Using exponential coordinates on the group we identify G with $\mathfrak{w}_{\mathrm{hor}} \oplus \mathfrak{w}_{\mathrm{vert}} \equiv \mathbb{R}^d \oplus \mathbb{R}^m$. With $x = (\underline{x}, \overline{x}) \in \mathbb{R}^d \times \mathbb{R}^m$ the group law then becomes

$$(\underline{x}, \overline{x}) \cdot (\underline{y}, \overline{y}) = (\underline{x} + \underline{y}, \overline{x} + \overline{y} + \underline{x}^{\intercal} \vec{J} \underline{y}) \tag{1.1}$$

where $\underline{x}^{\intercal} \vec{J} \underline{y} := \sum_{i=1}^m \overline{e}_i \underline{x}^{\intercal} J_i \underline{y}$, $\{\overline{e}_i\}_{i=1}^m$ is the standard basis of unit vectors in $\mathbb{R}^m$ and $J_1, \dots, J_m$ are $d \times d$ skew-symmetric matrices. The above dilations on $\mathfrak{g}$ induce automorphisms $\delta_t : (\underline{x}, \overline{x}) \mapsto (t\underline{x}, t^2\overline{x})$ on the group. We will study averaging operators which will be convolution operators; the noncommutative convolution for functions $f * K(x) = \int f(y) K(y^{-1} \cdot x) dy$ is then given in the form

$$\begin{aligned} f * K(\underline{x}, \overline{x}) &= \int f(\underline{y}, \overline{y}) K(\underline{x} - \underline{y}, \overline{x} - \overline{y} + \underline{x}^{\intercal} \vec{J} \underline{y})\, dy \\ &= \int f(\underline{x} - \underline{w}, \overline{x} - \overline{w} - \underline{x}^{\intercal} \vec{J} \underline{w})\, K(\underline{w}, \overline{w}) dw. \end{aligned} \tag{1.2}$$

Let Ω be a bounded open convex domain in $\mathfrak{w}_{\mathrm{hor}} \equiv \mathbb{R}^d \times \{0\}$ containing the origin and assume throughout the paper that the boundary $\Sigma \equiv \partial\Omega$ is smooth with *nonvanishing Gaussian curvature.* In most previous papers one takes for Ω the unit ball in $\mathbb{R}^d \times \{0\}$. Let μ be the normalized surface measure on Σ. For $t > 0$ the dilate μ_t is defined by $\langle f, \mu_t\rangle = \int f(tx, 0)d\mu$. For Schwartz functions f on $\mathbb{R}^{d+m}$ the averages over dilated spheres are then given by the convolutions

$$Af(x,t) = f * \mu_t(x) = \int_\Sigma f\big(\underline{x} - t\omega, \overline{x} - t\,\underline{x}^\intercal \vec{J}\omega\big)d\mu(\omega). \tag{1.3}$$

The analogue of the Nevo–Thangavelu maximal operator is defined (a priori for Schwartz functions) by

$$\mathfrak{M}f(x) := \sup_{t>0} \big|f * \mu_t(x)\big|. \tag{1.4}$$

The objective is to establish an $L^p(\mathbb{R}^{d+m}) \to L^p(\mathbb{R}^{d+m})$ bound for $\mathfrak{M}$, in an optimal range of p. Taking $\Sigma = S^{d-1}$ a partial boundedness result for $p > \frac{d-1}{d-2}$ was first obtained by Nevo and Thangavelu on the Heisenberg groups $\mathbb{H}^n$, for $2n \equiv d \geq 4$; here $m = 1$ and $J = J_1$ is an invertible symplectic matrix. The optimal result on L^p boundedness on the Heisenberg group $\mathbb{H}^n$, for $n \geq 2$, namely that $\mathfrak{M}$ is bounded on $L^p(\mathbb{R}^{d+1})$ for $p > \frac{d}{d-1}$ was obtained by Müller and the second author [14] and independently by Narayanan and Thangavelu [15]. The paper [14] also establishes this result in the more general setting of Métivier groups, that is, under the nondegeneracy condition that for all $\theta \in \mathbb{R}^m \setminus \{0\}$ the matrices $\sum_{i=1}^m \theta_i J_i$ are invertible. Regarding the case $n = 1$ it is not currently known whether $\mathfrak{M}$ is bounded on $L^p(\mathbb{H}^1)$ for any $p < \infty$ (see however results restricted to Heisenberg radial functions in [5] and [12]).

The purpose of this paper is to examine the behavior of the maximal function on general two-step nilpotent Lie groups with $d \geq 3$, i.e. when the nondegeneracy condition on Métivier groups fails. A trivial special case occurs when all the matrices J_i are zero; in this case one immediately obtains the same L^p boundedness result for $p > \frac{d}{d-1}$, $d \geq 3$ by applying Stein's result [21] (or Bourgain's result [4] when $d = 2$) in the horizontal hyper-planes and then integrating in $\mathfrak{w}_{\mathrm{vert}}$. The two extreme cases of Euclidean and Métivier groups suggest that L^p boundedness for $p > \frac{d}{d-1}$ should hold independently of the choice of the matrices J_i. However the intermediate cases are harder, and neither the slicing argument nor the arguments in [14, 15, 11, 18] for the Heisenberg and Métivier cases seem to apply; this was posed as a problem in [14]. In particular there seems to be no regularity theorem on Fourier integral operators which covers the averages in this general case. The special case $m = 1$ was recently considered by Liu and Yan [13] who obtain L^p boundedness of $\mathfrak{M}$ in the partial range $p > \frac{d-1}{d-2}$ and $d \geq 4$. Here we prove the optimal result in the range $p > \frac{d}{d-1}$, for all two-step nilpotent Lie groups with $d \geq 3$.

Theorem 1.1. *Let $d \geq 3$, let G be a general two step nilpotent Lie group of dimension $d+m$, with group law* (1.1). *Let $\frac{d}{d-1} < p < \infty$. Then for $f \in L^p(G)$ and almost every $x \in G$ the functions $t \to Af(x,t)$ are continuous and the maximal operator $\mathfrak{M}$ extends to a bounded operator on $L^p(G)$.*

By a standard argument Theorem 1.1 implies

Corollary 1.2. *Let $f \in L^p(G)$, $p > \frac{d}{d-1}$. Then $\lim_{t\to 0} Af(x,t) = f(x)$ almost everywhere.*

To show in Theorem 1.1 the continuity in t, for a.e. $x \in G$, we shall prove a stronger inequality involving the standard Besov space $B^{1/p}_{p,1}(\mathbb{R})$ which is embedded in the space of bounded continuous functions. Namely, for $u \in C_c^\infty((\frac14, 4))$ and $\lambda > 0$ the functions $s \to u(s)A_{\lambda s}f(x)$ are in $B^{1/p}_{p,1}$; this implies that $\sup_{t>0} |A_t f(x)| = \sup_{t\in\mathbb{Q}} |A_t f(x)|$ almost everywhere and establishes $\mathfrak{M}f$ as a well defined measurable function, for every $f \in L^p$. In fact one gets the following global result which implies Theorem 1.1.

Theorem 1.3. *Let $d \geq 3$, $\frac{d}{d-1} < p < \infty$. Then for $u \in C_c^\infty((\frac14, 4))$*

$$\Big(\int_G \big[\sup_{n\in\mathbb{Z}} \|u(\cdot)Af(x,2^n\cdot)\|_{B^{1/p}_{p,1}(\mathbb{R})}\big]^p dx\Big)^{1/p} \lesssim \|f\|_p. \tag{1.5}$$

Remark 1.4. The smoothness parameter for the Besov space in the s-variable can be increased in Theorem 1.3. In fact we can replace $B^{1/p}_{p,1}$ with $B^{\beta+1/p}_{p,1}$ where $\beta < d-1+\frac{d}{p}$ if $\frac{d}{d-1} < p \leq 2$ and $\beta < \frac{d-2}{p}$ if $2 \leq p < \infty$, see also a relevant discussion in §4. Such improvements will not yield additional insights on the maximal operator.

It is also of interest to consider a local variant for which we get a restricted weak type inequality at the endpoint $p = \frac{d}{d-1}$:

Theorem 1.5. *Let $d \geq 3$, $p \geq \frac{d}{d-1}$ and let $I \subset (0,\infty)$ be a compact interval. Then A maps $L^{p,1}(G)$ to $L^{p,\infty}(G; L^\infty(I))$.*

The optimality of the p-range in the above maximal function theorems is shown by modifying an example of Stein [21], see also the discussion in [13].

Outline of the paper and methodology. In §2 we reduce matters to the case where the matrices $J_1, \dots, J_m$ are linearly independent. In §3 we set up standard dyadic frequency decompositions of the underlying spherical measures and formulate the main Proposition 3.1 to be proved for the boundedness of the local maximal operator (with dilation parameters in a compact interval). The arguments to extend to the global maximal operator (and the slightly stronger Theorem 1.3) are modifications from those in [14]; this is taken up in §4. The main L^2 estimates are discussed in §5; here we first recast Proposition 3.1 in a convenient form using Fourier integral operators, and then reduce matters to the problem of getting uniform estimates for a family of oscillatory integral operators acting on functions in $\mathbb{R}^d$. The main L^2 estimates for this family are stated in Proposition 5.2. The crucial and most novel part of the paper is §6 where we give the proof of this proposition via two decompositions of the operator into more elementary building blocks which are combined via almost orthogonality arguments.

Notation. For nonnegative quantities a, b write $a \lesssim b$ to indicate $a \le Cb$ for some constant C. We write $a \approx b$ to indicate $a \lesssim b$ and $b \lesssim a$.

Acknowledgements. The second author thanks Shaoming Guo for re-raising the question, Detlef Müller for useful remarks on Carnot groups, and Brian Street for conversations on general pointwise convergence issues. We are indebted to the anonymous referee for a thorough and careful reading of the manuscript and for very helpful suggestions to improve it. The first author was supported in part by KIAS Individual Grant MG087001 during a research stay at UW Madison. The second author was supported in part by NSF Grant DMS-2054220.

2. Preliminary reductions

In the following theorem (which will be proved in subsequent sections) we formulate the main results, with a hypothesis of linear independence on the skew-symmetric matrices entering in the group structure. We then show how the proof of Theorems 1.3 and 1.5 are reduced to this result.

Theorem 2.1. *Let $S_1, \dots, S_m$ be a linearly independent set of $d \times d$ skew symmetric matrices. Let $\mathcal{U}$ be a neighborhood of the origin of $\mathbb{R}^{d-1}$ and let $g : \mathcal{U} \to \mathbb{R}$ be a C^∞ function satisfying $g(0) = 1$, $g'(0) = 0$, $g''(0)$ positive-definite. There exists $\rho > 0$ such that the following holds for C_c^∞ functions β_0 supported in a ball $\mathcal{U}_\rho \subset \mathcal{U}$ of diameter ρ centered at the origin in $\mathbb{R}^{d-1}$: Let $\Gamma_g(\omega') = g(\omega')e_1 + \sum_{i=2}^d \omega_i e_i$, and define $\mathcal{A}f(x, t) \equiv \mathcal{A}_t f(x)$ by*

$$\mathcal{A}f(x,t) = \int f\big(\underline{x} - t\Gamma_g(\omega'), \overline{x} - t\sum_{i=1}^m \overline{e}_i \underline{x}^\intercal S_i \Gamma_g(\omega')\big)\beta_0(\omega')d\omega' . \tag{2.1}$$

Let $u \in C_c^\infty((\frac14, 4))$. Then for $p > \frac{d}{d-1}$ the inequality

$$\Big\| \sup_n \|u(s)\mathcal{A}f(\cdot, 2^n s)\|_{B^{1/p}_{p,1}(\mathbb{R},ds)} \Big\|_{L^p(\mathbb{R}^{d+m})} \le C\|f\|_{L^p(\mathbb{R}^{d+m})} \tag{2.2}$$

holds for all functions $f \in L^p(\mathbb{R}^{d+m})$. Here C depends on p but is independent of f, and independent of β_0 as β_0 ranges over a bounded subset of $C_c^\infty(\mathcal{U}_\rho)$. Moreover for a compact interval $I \subset (0, \infty)$

$$\big\|\operatorname{ess\,sup}_{t\in I}|\mathcal{A}_t f|\big\|_{L^{p,\infty}(\mathbb{R}^{d+m})} \le C_I \|f\|_{L^{p,1}(\mathbb{R}^{d+m})}, \quad p = \tfrac{d}{d-1}. \tag{2.3}$$

We shall now show in several steps how Theorems 1.1, 1.3 and 1.5 are implied by Theorem 2.1. In our first reduction we reduce to the situation that the manifold $\Sigma \subset \mathbb{R}^m$ that supports the measure μ is given as a graph with the property required in Theorem 2.1.

We fix a point $y_\circ \in \Sigma$ and consider the operator $f \mapsto f * (\chi^{y_\circ}\mu)_t$ where $\chi^{y_\circ}$ is a C_c^∞ function supported in a neighborhood of $y_\circ$. It suffices to prove the analogues of Theorem 1.1 and 1.3 for these convolutions; once this is achieved one can use a compactness and partition of unity argument to deduce Theorems 1.1, 1.3 in their original formulation. Let $e_{\circ,1} = y_\circ/|y_\circ|$ (recall that the origin lies in the domain surrounded by Σ and thus $y_\circ \neq 0$). Pick unit vectors $e_{\circ,i}$, $2 \le i \le d$ so that $e_{\circ,1}, \dots, e_{\circ,d}$ is an orthonormal basis of $\mathbb{R}^d$. As $e_{\circ,1}$ does not belong to the tangent space to $|y_\circ|^{-1}\Sigma$ at $e_{\circ,1}$ we may parametrize $|y_\circ|^{-1}\Sigma$ near $e_{\circ,1}$ by

$$\Gamma(\omega') = G(\omega')e_{\circ,1} + \sum_{i=2}^{d} \omega_i e_{\circ,i};$$

here $(\omega')^\intercal = (\omega_2, \dots, \omega_d)$ and the function G satisfies

$$G(0) = 1,\ G'(0) = b \text{ and } G''(0) \text{ positive definite.}$$

We then have for $\nu = \chi^{y_\circ}\mu$

$$f * \nu_t(x) = \int f(\underline{x} - t|y_\circ|\Gamma(\omega'), \overline{x} - t|y_\circ|\underline{x}^\intercal \vec{J}\Gamma(\omega'))\chi_\circ(\omega')d\omega'$$

with $\chi_\circ(\omega') = \chi^{y_\circ}(|y_\circ|\Gamma(\omega'))|y_\circ|^d(1+|G'(\omega')|^2)^{1/2}$ and $\underline{x}^\intercal \vec{J}\Gamma(\omega')$ denotes the vector $\sum_{i=1}^m \overline{e}_i \underline{x}^\intercal J_i \Gamma(\omega')$.

Let P be the $(d-1)\times d$ matrix defined by $P = \begin{pmatrix} 0 & I_{d-1} \end{pmatrix}$, corresponding to the projection $(x_1, x_2, \dots, x_d)^\intercal \mapsto (x_2, \dots, x_d)^\intercal$. Let R denote the rotation satisfying $Re_{\circ,i} = e_i$ for $i = 1, \dots, d$. Then

$$R\Gamma(\omega') = \Gamma_G(\omega') := G(\omega')e_1 + \sum_{i=2}^{d} \omega_i e_i \equiv G(\omega')e_1 + P^\intercal \omega'.$$

Setting $\widetilde{J}_i = |y_\circ|^2 R J_i R^\intercal$ we define $\mathcal{A}^{[1]} f(x,t) \equiv \mathcal{A}_t^{[1]} f(x)$ by

$$\mathcal{A}_t^{[1]} f(x) = \int f(\underline{x} - t\Gamma_G(\omega'), \overline{x} - t\sum_{i=1}^{m} \overline{e}_i \underline{x}^\intercal \widetilde{J}_i \Gamma_G(\omega'))\chi_\circ(\omega')d\omega'. \tag{2.4}$$

Then we compute

$$f * \nu_t(x) = \mathcal{A}_t^{[1]} h(|y_\circ|^{-1}R\underline{x}, \overline{x}), \text{ with } h(\underline{x}, \overline{x}) = f(|y_\circ|R^\intercal \underline{x}, \overline{x}).$$

Since $R^{-1} = R^\intercal$ it suffices to prove (1.5) and the maximal bounds with $\mathcal{A}^{[1]}$ in place of A.

We now use another transformation to reduce to the situation in Theorem 2.1. To this end we set $g(\omega') = G(\omega') - b^\intercal\omega'$ so that $g'(0) = 0$ as in Theorem 2.1 (recall $b = G'(0)$).

We then have (splitting $\underline{x} = (x_1, x') \in \mathbb{R} \times \mathbb{R}^{d-1}$)

$$\mathcal{A}_t^{[1]} f(x) = \int f(x_1 - tg(\omega') - tb^\intercal\omega', x' - t\omega', \overline{x} - t\overline{v}(\underline{x}, \omega'))\chi_\circ(\omega')d\omega'$$

where $\overline{v} = (\overline{v}_1, \dots, \overline{v}_m)$ with

$$\overline{v}_i(\underline{x}, \omega') = \underline{x}^\intercal \widetilde{J}_i(g(\omega') + b^\intercal\omega')e_1 + \underline{x}^\intercal \widetilde{J}_i P^\intercal \omega'.$$

Now write

$$x_1 - tg(\omega') - tb^\intercal\omega' = x_1 - b^\intercal x' - tg(\omega') + b^\intercal(x' - t\omega') \tag{2.5}$$

and

$$\begin{aligned}\overline{v}_i(\underline{x}, \omega') =& (x')^\intercal P\widetilde{J}_i e_1 g(\omega') + (x')^\intercal P\widetilde{J}_i e_1(b^\intercal\omega') + (b^\intercal x')e_1^\intercal \widetilde{J}_i P^\intercal\omega' \\ &+ (x_1 - b^\intercal x')e_1^\intercal \widetilde{J}_i P^\intercal\omega' + (x')^\intercal P\widetilde{J}_i P^\intercal\omega'.\end{aligned}$$

Observe that, with $\Gamma_g(\omega') = g(\omega)e_1 + P^\intercal\omega'$,

$$\begin{aligned}(x')^\intercal P\widetilde{J}_i e_1(b^\intercal\omega') &= -\underline{x}^\intercal P^\intercal P\widetilde{J}_i^\intercal e_1 b^\intercal P\Gamma_g(\omega') \\ (b^\intercal x')e_1^\intercal \widetilde{J}_i P^\intercal\omega' &= \underline{x}^\intercal P^\intercal b e_1^\intercal \widetilde{J}_i P^\intercal P\Gamma_g(\omega'),\end{aligned}$$

also the analogous formulas remain true if on the right hand sides $\underline{x}$ is replaced with $(x_1 - b^\intercal x')e_1 + P^\intercal x'$. Furthermore

$$\begin{aligned}(x')^\intercal P\widetilde{J}_i e_1 g(\omega) + (x_1 - b^\intercal x')e_1^\intercal \widetilde{J}_i P^\intercal\omega' + (x')^\intercal P\widetilde{J}_i P^\intercal\omega' \\ = \big[(x_1 - b^\intercal x')e_1 + P^\intercal x'\big]^\intercal \widetilde{J}_i\Gamma_g(\omega').\end{aligned}$$

We combine the above observations, and setting

$$\mathcal{J}_i = \widetilde{J}_i + P^\intercal b e_1^\intercal \widetilde{J}_i P^\intercal P - P^\intercal P\widetilde{J}_i^\intercal e_1 b^\intercal P \tag{2.6}$$

we see that $\mathcal{J}_i$ are skew symmetric $d \times d$ matrices satisfying $\mathcal{J}_i e_1 = \widetilde{J}_i e_1$, and that

$$\overline{v}_i(\underline{x}, \omega') = \big[(x_1 - b^\intercal x')e_1 + P^\intercal x'\big]^\intercal \mathcal{J}_i\Gamma_g(\omega'). \tag{2.7}$$

Now if we define $\mathcal{A}_t^{[2]}$ by

$$\mathcal{A}_t^{[2]} f(x) = \int f(\underline{x} - t\Gamma_g(\omega'), \overline{x} - t\sum_{i=1}^m \overline{e}_i \underline{x}^\intercal \mathcal{J}_i\Gamma_g(\omega'))\chi_\circ(\omega')d\omega' \tag{2.8}$$

then it follows from (2.5) and (2.7) that

$$\begin{aligned}\mathcal{A}_t^{[1]} f(x_1, x', \overline{x}) &= \mathcal{A}_t^{[2]} f_b(x_1 - b^\intercal x', x', \overline{x}) \\ &\text{with } f_b(y_1, y', \overline{y}) = f(y_1 + b^\intercal y', y', \overline{y}).\end{aligned} \tag{2.9}$$

Hence the desired bounds for the families $(\mathcal{A}_t^{[1]})_{t>0}$ and $(\mathcal{A}_t^{[2]})_{t>0}$ are equivalent. For the case that the matrices $\mathcal{J}_1, \dots, \mathcal{J}_m$ are linearly independent (1.5) and the L^p boundedness of the maximal operator in theorem 1.1 can now be obtained from Theorem 2.1 (using $S_i = \mathcal{J}_i$ in that theorem).

In the other extreme case, when all $\mathcal{J}_i$ are the zero-matrices the L^p boundedness of the maximal operator operator (and the analogue of (1.5)) follows by an application of the spherical maximal theorems in the Euclidean case in $\mathbb{R}^d$ ([21]) and integration in the vertical variables. In this case we also have the result for $d = 2$ by using Bourgain's theorem [4] (although this is not needed in our proof). If $d \geq 3$ the restricted weak type inequality for $p = \frac{d}{d-1}$ can be deduced from [3] and a slicing argument. These slicing arguments also apply to the variants where a $B^{1/p}_{p,1}$-norm is used on the dilation parameter.

It remains to consider the case where the matrices $\mathcal{J}_i$ are not all zero but are linearly dependent. For this case we need a further reduction.

Lemma 2.2. *Assume that $\mathcal{J}_1, \dots, \mathcal{J}_m$ are not all zero. Then there exist linearly independent skew symmetric $d \times d$ matrices $S_1, \dots, S_n$, with $1 \leq n \leq m$, and an orthogonal matrix $V \in O(m)$ such that for $\mathcal{A}^{[2]}_t$ as in (2.8)*

$$\mathcal{A}^{[2]}_t f(x) = \mathcal{A}^{[3]}_t f_V(\underline{x}, V\overline{x}), \tag{2.10}$$

with $f_V(y) = f(\underline{y}, V^\intercal \overline{y})$ and

$$\mathcal{A}^{[3]}_t f(x) = \int f(\underline{x} - t\Gamma_g(\omega'), \overline{x} - t\sum_{i=1}^n \overline{e}_i \underline{x}^\intercal S_i \Gamma_g(\omega'))\chi_\circ(\omega')d\omega' . \tag{2.11}$$

Proof. Consider a basis $E_1, \dots E_{\frac{d(d-1)}{2}}$ in the space of $d \times d$ skew symmetric matrices. We can express the $\mathcal{J}_i$ in terms of the basis matrices, and obtain $\mathcal{J}_i = \sum_{j=1}^{\frac{d(d-1)}{2}} c_{ij} E_j$, $i = 1, \dots, m$ for suitable scalars c_{ij}. We denote by C the $m \times \frac{d(d-1)}{2}$ matrix whose (i, j) entry is given by c_{ij}. We apply the singular value decomposition of the transposed matrix $C^\intercal$. That is, we decompose $C^\intercal = UDV$ where U is an orthogonal $\frac{d(d-1)}{2} \times \frac{d(d-1)}{2}$ matrix, V is an orthogonal $m \times m$ matrix and D is a $\frac{d(d-1)}{2} \times m$ matrix such that

$$D_{ij} = \begin{cases} s_i & \text{if } 1 \leq i = j \leq n \\ 0 & \text{otherwise.} \end{cases}$$

Here $n \leq \min\{\frac{d(d-1)}{2}, m\}$ and $s_1 \geq \dots \geq s_n > 0$ are the singular values. For the coefficients of C we then get

$$c_{ij} = (C^\intercal)_{ji} = \sum_{k=1}^{\frac{d(d-1)}{2}} \sum_{\ell=1}^m U_{jk} D_{k\ell} V_{\ell i} = \sum_{k=1}^n U_{jk} s_k V_{ki}.$$

Defining

$$S_k = s_k \sum_{j=1}^{\frac{d(d-1)}{2}} U_{jk} E_j, \quad k = 1, \dots n,$$

it is clear by the invertibility of U that $S_1, \dots, S_n$ are linearly independent skew symmetric matrices and we obtain

$$\mathcal{J}_i = \sum_{j=1}^{\frac{d(d-1)}{2}} c_{ij} E_j = \sum_{j=1}^{\frac{d(d-1)}{2}} \sum_{k=1}^{n} U_{jk} s_k V_{ki} E_j = \sum_{k=1}^{n} V_{ki} S_k\,, \quad i = 1, \dots, m.$$

Hence (using $V^\intercal = V^{-1}$)

$$\overline{x} - t \sum_{i=1}^{m} \overline{e}_i \underline{x}^\intercal \mathcal{J}_i \Gamma_g(\omega') = V^\intercal \big[V\overline{x} - t \sum_{k=1}^{n} \overline{e}_k \underline{x}^\intercal S_k \Gamma_g(\omega') \big]$$

which gives (2.10). □

By Lemma 2.2 the desired bounds for $\mathcal{A}_t^{[2]}$ and $\mathcal{A}_t^{[3]}$ are equivalent. We show how Theorem 2.1 yields the analogue of Theorem 1.3 for the family $(\mathcal{A}_t^{[3]})_{t>0}$ in place of (A_t). In $\mathbb{R}^m$ we split variables $\overline{x} = (\tilde{x}, \breve{x}) \in \mathbb{R}^n \times \mathbb{R}^{m-n}$. For $h \in L^p(\mathbb{R}^{d+n})$ we define $\mathfrak{A}h(x,t) = \mathfrak{A}_t h(x)$ by

$$\mathfrak{A}_t h(\underline{x}, \tilde{x}) = \int h(\underline{x} - t\Gamma_g(\omega'), \tilde{x} - t \sum_{i=1}^{n} \overline{e}_i \underline{x}^\intercal S_i \Gamma_g(\omega')) \chi_\circ(\omega') d\omega'.$$

We get from Theorem 2.1 applied with n in place of m, that

$$\Big\| \sup_{\ell \in \mathbb{Z}} \|u(s) \mathfrak{A}_{2^\ell s} h\|_{B^{1/p}_{p,1}(\mathbb{R},ds)} \Big\|_{L^p(\mathbb{R}^{d+n})} \lesssim \|h\|_{L^p(\mathbb{R}^{d+n})}, \quad p > \frac{d}{d-1}.$$

Let $f^{\breve{x}}(\underline{x}, \tilde{x}) = f(\underline{x}, \tilde{x}, \breve{x})$, and observe that $\mathcal{A}_t^{[3]} f(\underline{x}, \tilde{x}, \breve{x}) = \mathfrak{A}_t f^{\breve{x}}(\underline{x}, \tilde{x})$. We apply the $L^p(\mathbb{R}^{d+n})$-boundedness result stated above to the functions for $h = f^{\breve{x}}$ and get

$$\iint \sup_{\ell \in \mathbb{Z}} \|u(s) \mathcal{A}^{[3]}_{2^\ell s} f(\underline{x}, \tilde{x}, \breve{x})\|^p_{B^{1/p}_{p,1}(\mathbb{R},ds)} d\underline{x} d\tilde{x} \le C^p \iint |f(\underline{x}, \tilde{x}, \breve{x})|^p d\underline{x} d\tilde{x},$$

with C independent of $\breve{x}$. Integrating over $\breve{x} \in \mathbb{R}^{m-n}$ gives the desired result.

We have now reduced the proof of Theorem 1.1 to the inequalities (2.2) in Theorem 2.1. The above arguments also reduce (after minor modifications) the proof of Theorem 1.5 to the proof of inequality (2.3). For the remainder of the paper we will be concerned with the proof of Theorem 2.1.

Remark. The shear transformation showing the equivalence of the L^p boundedness of the operators associated with $\mathcal{A}^{[1]}$ and $\mathcal{A}^{[2]}$ is not needed for the spherical case $\Sigma = S^{d-1}$, when $b = 0$. However in the general case it seems necessary, and we take this opportunity to correct an inaccuracy in [14] which deals with the case of Métivier groups (i.e. the matrices $\sum_{i=1}^m c_i J_i$ are invertible if $(c_1, \dots, c_m) \neq 0$). There it is stated that this reduction follows for more general Σ by a rotation argument which is not the case. One can use the above shear transformations instead and deduce that the arguments in [14] apply to surfaces Σ that are small perturbations of the sphere. Such a perturbation assumption would be needed for the proof in

[14] since the Métivier condition on the matrices J_i (and, equivalently, on the $\widetilde{J}_i = |y_\circ|^2 R J_i R^\intercal$) guarantees the Métivier condition on the matrices $\mathcal{J}_i$ in (2.6) only when b is sufficiently small. In the setup of this paper there is no such small smallness assumption on b needed.

For the remainder of the paper we will give the proof of Theorem 2.1 and fix linearly independent skew symmetric matrices $S_1, \dots, S_m$. For later use notice that this assumption implies that there is a $c_0 > 0$ such that

$$c_0 \le \Big\| \sum_{i=1}^m \theta_i S_i \Big\| \le c_0^{-1} \text{ for all } \theta \in \mathbb{R}^m \text{ with } 1/4 \le |\theta| \le 4. \tag{2.12}$$

This is immediate from the fact that $\theta \to \|\sum_{i=1}^m \theta_i S_i\|$ is a continuous function which takes a minimum and a maximum on the annulus $\{\theta : 1/4 \le |\theta| \le 4\}$, and by the assumed linear independence this minimum is positive.

3. Dyadic frequency decompositions

We now use the group structure on $\mathbb{R}^{d+m}$ given by (1.1) but with the J_i replaced by the skew symmetric matrices S_i, with $S_1, \dots, S_m$ linearly independent. We denote by ν the measure defined by $\langle \nu, f\rangle = \int f(g(\omega'), \omega', 0) d\omega'$ which can also be written as the pairing of the distribution

$$\beta_0(x')\beta_1(x_1, \overline{x})\delta(x_1 - g(x'), \overline{x})$$

with f; here δ is the Dirac measure in $\mathbb{R}^{m+1}$, β_0 is a C^∞ function supported on a ball of radius ϱ centered at the origin of $\mathbb{R}^{d-1}$ and β_1 is a C^∞ function supported on an ϱ^2-ball centered at $(1, 0, \dots, 0) \in \mathbb{R}^{m+1}$, here $\epsilon_\circ \ll \varrho$. We assume that ϱ is small compared with the reciprocal of the C^3 norm of g, also $\varrho \ll \|(g''(0))^{-1}\|^{-1}$ and finally $\varrho \ll c_0$ where c_0 is as in (2.12).

We use a dyadic frequency decomposition of the Fourier integral of δ to decompose $\nu = \sum_{k=0}^\infty \nu^k$ where

$$\nu^k(x) = \frac{\beta_1(x_1, \overline{x})\beta_0(x')}{(2\pi)^{m+1}} \iint \zeta_k(\sqrt{\sigma^2 + |\tau|^2}) e^{i\sigma(x_1 - g(x')) + i\langle \tau, \overline{x}\rangle} d\sigma d\tau \tag{3.1}$$

where $\zeta_0 \in C_c^\infty(\mathbb{R})$ is supported on $(-1, 1)$, $\zeta_0(s) = 1$ for $|s| < 3/4$ and $\zeta_k(s) = \zeta_0(2^{-k}s) - \zeta_0(2^{1-k}s)$ when $k \ge 1$; hence, for $k \ge 1$ the function $\zeta_k = \zeta_1(2^{-(k-1)}\cdot)$ is supported in $(-2^{k-1}, -2^{k-3}) \cup (2^{k-3}, 2^{k-1})$. For $k > 0$ we make a further decomposition in the σ-variables setting

$$\nu^{k,l}(x) = \frac{\beta_1(x_1, \overline{x})\beta_0(x')}{(2\pi)^{m+1}} \iint \zeta_{k,l}(\sigma, \tau) e^{i\sigma(x_1 - g(x')) + i\langle \tau, \overline{x}\rangle} d\sigma d\tau \tag{3.2}$$

where

$$\zeta_{k,l}(\sigma, \tau) = \begin{cases} \zeta_1(2^{1-k}\sqrt{\sigma^2 + |\tau|^2})\zeta_1(2^{l+1-k}\sigma) & \text{for } l < k \\ \zeta_1(2^{1-k}\sqrt{\sigma^2 + |\tau|^2})\zeta_0(\sigma) & \text{for } l = k \end{cases}$$

i.e., for $k \geq 1$, $l < k$ we have the restriction $|\sigma| + |\tau| \approx 2^k$ and $|\sigma| \approx 2^{k-l}$ in the frequency variables. We set $\nu_t^{k,l}(x) = t^{-d-2m}\nu^{k,l}(t^{-1}\underline{x}, t^{-2}\overline{x})$ and similarly define ν_t^k.

We state the main local estimates for $f * \nu_t^{k,l}$.

Proposition 3.1. *Let $\varepsilon > 0$. Let I be a compact subinterval of $(0,\infty)$. Then there exists a constant $C = C(\varepsilon, I) > 0$ such that the following holds.*

$$\Big(\int_I \|f * \nu_t^{k,l}\|_{L^p(\mathbb{R}^{d+m})}^p dt\Big)^{1/p} + 2^{l-k}\Big(\int_I \|\partial_t(f * \nu_t^{k,l})\|_{L^p(\mathbb{R}^{d+m})}^p dt\Big)^{1/p} \leq \begin{cases} C2^{-\frac{k(d-1)}{p'}} 2^{l(\frac{d-2}{p'}+\varepsilon)}\|f\|_{L^p(\mathbb{R}^{d+m})} & \text{if } 1 \leq p \leq 2, \\ C2^{-\frac{k(d-1)}{p}} 2^{l(\frac{d-2}{p}+\varepsilon)}\|f\|_{L^p(\mathbb{R}^{d+m})} & \text{if } 2 \leq p < \infty. \end{cases} \tag{3.3}$$

Corollary 3.2. *Let $\frac{d}{d-1} < p < \infty$, $u \in C_c^\infty((0,\infty))$. Let $f \in L^p(\mathbb{R}^{d+m})$. Then for almost every $x \in \mathbb{R}^{d+m}$ the function $t \mapsto \mathcal{A}f(x,t)$ is continuous, and for any $\lambda > 0$*

$$\Big(\int_{\mathbb{R}^{d+m}} \big\|u(\cdot)Af(x,\lambda\cdot)\big\|_{B^{1/p}_{p,1}}^p dx\Big)^{1/p} \lesssim \|f\|_{L^p(\mathbb{R}^{d+m})}. \tag{3.4}$$

Proof. By scaling we can assume that $\lambda = 1$. The first statement follows from the second, since $B^{1/p}_{p,1}$ embeds into the space of bounded continuous functions. Let $1 \leq p \leq 2$. Set $\mathcal{R}^{k,l}f(x,s) = u(s)f * \nu_s^{k,l}(x)$. We use the interpolation inequality $\|g\|_{B^\theta_{p,1}} \lesssim \|g\|_p^{1-\theta}\|g'\|_p^\theta$ $(0 < \theta < 1)$, Hölder's inequality, Fubini and the proposition to deduce that the left hand side of (3.4) is dominated by

$$\begin{aligned}\|\mathcal{R}^{k,l}f\|_{L^p(B^{1/p}_{p,1})} &\lesssim \|\mathcal{R}^{k,l}f\|_{L^p(L^p)}^{1-\theta}\|\partial_t\mathcal{R}^{k,l}f\|_{L^p(L^p)}^\theta \\ &\lesssim 2^{-k(\frac{d-1}{p'}-\theta)}2^{l(\frac{d-2}{p'}-\theta+\varepsilon)}\|f\|_p.\end{aligned}$$

The desired inequality follows by summing over $l \leq k$ and then summing over k (which is possible if $d \geq 3$ and $\frac{d}{d-1} < p \leq 2$, $\theta = 1/p$). A similar argument applies for $p > 2$. □

Proof of the restricted weak type inequality in Theorem 2.1. Let $\mathcal{R}^{k,l}f(x,t) := f * \nu_t^{k,l}$ and as in the proof of Corollary 3.2 we have that $\mathcal{R}^{k,l}$ maps L^p to $L^p(L^\infty)$ with operator norm $O(2^{k(\frac{1}{p}-\frac{d-1}{p'})}2^{-l(\frac{1}{p}-\frac{d-2}{p'}-\varepsilon)})$. If $1 \leq p < \frac{d-1}{d-2}$ we may (for sufficiently small ε) sum in l and obtain in this range

$$\|\operatorname{ess\,sup}_{t\in I}|f * \nu_t^k\|_p \lesssim 2^{k(\frac{1}{p}-\frac{d-1}{p'})}\|f\|_p.$$

We are now applying the 'Bourgain trick' in [3] to sum in k and deduce that

$$\|\operatorname{ess\,sup}_{t\in I}|f * \nu_t|\|_{L^{p,\infty}} \lesssim \|f\|_{L^{p,1}}, \quad p = \tfrac{d}{d-1}.$$ □

The most interesting part of Proposition 3.1 is the L^2-estimate. The L^p estimates follow by interpolation with L^1 estimates which we now briefly discuss.

L^1 and L^∞ estimates. In what follows $\beta(x) = \beta_1(x_1, \overline{x})\beta_0(x')$. By integration by parts with respect to σ, τ we obtain the inequality

$$|\nu^{k,l}(x)| \lesssim_N \frac{2^{k-l}}{(1+2^{k-l}|x_1 - g(x')|)^N} \frac{2^{km}}{(1+2^k|\overline{x}|)^N}; \tag{3.5}$$

moreover $2^{-k}\nabla\nu^{k,l}$, $2^{l-k}\partial_s\nu_s^{k,l}$, $2^{l-2k}\partial_s\nabla\nu_s^{k,l}$ satisfy for $|s| \approx 1$ the same pointwise bounds. Hence we obtain

$$\|\nu^{k,l}\|_1 + 2^{l-k}\|\partial_s \nu_s^{k,l}\|_1 \lesssim 1. \tag{3.6}$$

For later use we also record

$$\|\nabla\nu^{k,l}\|_1 + 2^{l-k}\|\nabla\partial_s \nu_s^{k,l}\|_1 \lesssim 2^k. \tag{3.7}$$

We will show in the next section §4 how to prove the L^p boundedness for the global maximal operator and its strengthening in Theorem 1.3, given the result of Proposition 3.1 and (3.6), (3.7). The proof of Proposition 3.1 will be given in §5 -§6.

4. The global maximal operator

We prove the global bound in Theorem 1.3, given Proposition 3.1. The reduction to Proposition 3.1 follows closely arguments in [14]; we include details for the convenience of the reader.

Let $I = (\frac{1}{4}, 4)$ and $u \in C_c^\infty(I)$. We will prove the estimate

$$\begin{aligned} &\Big\| \sup_{n\in\mathbb{Z}} \|u(s) f * \nu_{2^n s}^{k,l}\|_{B^{1/p}_{p,1}} \Big\|_p \\ &\qquad \le C_\varepsilon \begin{cases} (1+k)^{1/p} 2^{k(\frac{1}{p}-\frac{d-1}{p'})} 2^{l(\frac{d-2}{p'}-\frac{1}{p}+\varepsilon)} \|f\|_p, & 1<p\le 2 \\ (1+k)^{1/p} 2^{kd/p} 2^{l(\frac{d-1}{p}+\varepsilon)} \|f\|_p, & 2\le p<\infty, \end{cases} \end{aligned} \tag{4.1}$$

for $0 \le l \le k$; here the Besov-norm is taken with respect to the s variable. Summing in l, k for $p > \frac{d}{d-1}$ implies (1.5).

Recall that $\nu^{k,l}$ is compactly supported and that

$$\Big| \int \nu^{k,l}(x)dx \Big| \lesssim_N 2^{-kN};$$

this is seen by using (3.2) and repeated integration by parts, with respect to $(x_1, \overline{x})$, if l is small and with respect to $\overline{x}$ if l is large.

As noted in [14] we can write

$$\nu^{k,l}(\underline{x}, \overline{x}) = \mathcal{K}^{k,l}(\underline{x}, \overline{x}) + \gamma_{k,l}\rho(\underline{x}, \overline{x})$$

where $\rho \in C_c^\infty(\mathbb{R}^{d+m})$ function supported near the origin, $|\gamma_{k,l}| \le c_N 2^{-kN}$ for $0 \le l \le k$ and

$$\int \mathcal{K}^{k,l}(x)dx = 0. \tag{4.2}$$

Set $\mathcal{K}_t^{k,l}(\underline{x}, \overline{x}) = t^{-d-2m}\mathcal{K}^{k,l}(t^{-1}\underline{x}, t^{-2}\overline{x})$, $\rho_t(\underline{x}, \overline{x}) = t^{-d-2m}\rho(t^{-1}\underline{x}, t^{-2}\overline{x})$.

The contribution from $\gamma_{k,l}\rho$ is harmless, indeed for all $n \in \mathbb{Z}$, $s \in [1/4, 4]$ with $j = 0, 1, 2, \dots$

$$\Big|(\frac{d}{ds})^j \big[u(s) f * \rho_{2^n s}(x)\big]\Big| \lesssim c_j Mf(x)$$

where M is the Hardy-Littlewood maximal operator associated to the Carnot-balls on the group G. This implies

$$\Big\| \sup_{n\in\mathbb{Z}} \|u(s) f * \rho_{2^n s}\gamma_{k,l}\|_{B^{1/p}_{p,1}} \Big\|_p \le C_N 2^{-kN} \|f\|_p.$$

Therefore we need to show the equivalent of (4.1) where $\nu^{k,l}$ is replaced by $\mathcal{K}^{k,l}$. This is implied by two stronger inequalities where the sup in n is replaced by an ℓ^2 norm in n when $1 < p \le 2$ and by an ℓ^p norm in n when $2 < p < \infty$.

We shall prove for $1 < p \le 2$

$$\Big\|\Big(\sum_{n\in\mathbb{Z}} \|u(s) f * \mathcal{K}^{k,l}_{2^n s}\|^2_{B^{1/p}_{p,1}}\Big)^{1/2}\Big\|_p \le C_{\varepsilon,p}(1+k)^{1/p} 2^{k(\frac{1}{p} - \frac{d-1}{p'})} 2^{l(\frac{d-2+\varepsilon}{p'} - \frac{1}{p})} \|f\|_p \tag{4.3}$$

and for $2 \le p < \infty$

$$\Big\|\Big(\sum_{n\in\mathbb{Z}} \|u(s) f * \mathcal{K}^{k,l}_{2^n s}\|^p_{B^{1/p}_{p,1}}\Big)^{1/p}\Big\|_p \le C_{\varepsilon,p}(1+k)^{1/p} 2^{-k\frac{d-2}{p}} 2^{l\frac{d-3+\varepsilon}{p}} \|f\|_p \tag{4.4}$$

We use the interpolation inequality

$$\|g\|_{B^{1/p}_{p,1}} \lesssim \|g\|_p + \|g\|_p^{1-1/p} \|g'\|_p^{1/p} \tag{4.5}$$

which is elementary and also expresses the identification of $B^{1/p}_{p,1}$ as the real interpolation space $[L^p, W^{1,p}]_{1/p,1}$.

We focus on the case $1 \le p \le 2$ and prove (4.3). The embedding inequality implies

$$\Big\|\Big(\sum_{n\in\mathbb{Z}} \|u(s) f * \mathcal{K}^{k,l}_{2^n s}\|^2_{B^{1/p}_{p,1}}\Big)^{1/2}\Big\|_p \lesssim \Big\|\Big(\sum_{n\in\mathbb{Z}} \Big(\int_I |f * \mathcal{K}^{k,l}_{2^n s}|^p ds\Big)^{2/p}\Big)^{1/2}\Big\|_p$$

$$+ \Big\|\Big(\sum_n \Big(\int_I |f * \mathcal{K}^{k,l}_{2^n s}|^p ds\Big)^{\frac{2}{pp'}} \Big(\int_I \big|f * \tfrac{d}{ds}\mathcal{K}^{k,l}_{2^n s}\big|^p ds\Big)^{\frac{2}{p^2}}\Big)^{1/2}\Big\|_p =: \mathcal{E}_1 + \mathcal{E}_2\,.$$

For the first expression on the right hand side we have by the integral Minkowski's inequality for $\ell^{2/p}$

$$\mathcal{E}_1 \le \Big\| \Big(\int_I \Big(\sum_n |f * \mathcal{K}^{k,l}_{2^n s}|^2 \Big)^{p/2} ds \Big)^{1/p} \Big\|_p$$

and for the second term we use applications of Hölder and then the integral Minkowski inequality

$$\begin{aligned} \mathcal{E}_2 &\le \Big\| \Big(\sum_n \Big(\int_I |f * \mathcal{K}^{k,l}_{2^n s}|^p ds \Big)^{\frac{2}{p}} \Big)^{\frac{1}{2p'}} \Big(\sum_n \Big(\int_I |f * \tfrac{d}{ds} \mathcal{K}^{k,l}_{2^n s}|^p ds \Big)^{\frac{2}{p}} \Big)^{\frac{1}{2p}} \Big\|_p \\ &\le \Big\| \Big(\sum_n \Big(\int_I |f * \mathcal{K}^{k,l}_{2^n s}|^p ds \Big)^{\frac{2}{p}} \Big)^{\frac{1}{2}} \Big\|_p^{\frac{1}{p'}} \Big\| \Big(\sum_n \Big(\int_I |f * \tfrac{d}{ds} \mathcal{K}^{k,l}_{2^n s}|^p ds \Big)^{\frac{2}{p}} \Big)^{\frac{1}{2}} \Big\|_p^{\frac{1}{p}} \\ &\le \Big\| \Big(\int_I \Big(\sum_n |f * \mathcal{K}^{k,l}_{2^n s}|^2 \Big)^{\frac{p}{2}} ds \Big)^{\frac{1}{p}} \Big\|_p^{\frac{1}{p'}} \Big\| \Big(\int_I \Big(\sum_n |f * \tfrac{d}{ds} \mathcal{K}^{k,l}_{2^n s}|^2 \Big)^{\frac{p}{2}} ds \Big)^{\frac{1}{p}} \Big\|_p^{\frac{1}{p}}. \end{aligned}$$

Since we may interchange the x and the s integration, everything for $p \le 2$ follows now from

$$\text{(4.6a)} \quad \Big(\iint_{G \times I} \Big(\sum_n |f * \mathcal{K}^{k,l}_{2^n s}(x)|^2 \Big)^{\frac{p}{2}} dx\, ds \Big)^{\frac{1}{p}} \lesssim_\varepsilon (1+k)^{1/p} 2^{-k \frac{d-1}{p'}} 2^{l(\frac{d-2}{p'} + \varepsilon)} \|f\|_p$$

and

$$\text{(4.6b)} \quad \Big(\iint_{G \times I} \Big(\sum_n |f * \tfrac{d}{ds} \mathcal{K}^{k,l}_{2^n s}(x)|^2 \Big)^{\frac{p}{2}} dx\, ds \Big)^{\frac{1}{p}} \lesssim_\varepsilon 2^{k-l} (1+k)^{1/p} 2^{-k \frac{d-1}{p'}} 2^{l(\frac{d-2}{p'} + \varepsilon)} \|f\|_p\,.$$

We prove (4.6) by Marcinkiewicz interpolation, using the L^2 bounds

$$\text{(4.7a)} \quad \Big(\iint_{G \times I} \sum_{n \in \mathbb{Z}} |f * \mathcal{K}^{k,l}_{2^n s}|^2 dx\, ds \Big)^{\frac{1}{2}} \lesssim_\varepsilon (1+k)^{\frac{1}{2}} 2^{-k \frac{d-1}{2} + l(\frac{d-2}{2} + \varepsilon)} \|f\|_{L^2},$$

$$\text{(4.7b)} \quad \Big(\iint_{G \times I} \sum_{n \in \mathbb{Z}} \big| f * \tfrac{d}{ds} \mathcal{K}^{k,l}_{2^n s}(x) \big|^2 dx\, ds \Big)^{\frac{1}{2}} \lesssim_\varepsilon 2^{k-l} (1+k)^{\frac{1}{2}} 2^{-k \frac{d-1}{2} + l(\frac{d-2}{2} + \varepsilon)} \|f\|_{L^2},$$

and the weak type $(1,1)$ inequalities

$$\text{(4.8a)} \quad \operatorname{meas}\Big(\{ (x,s) \in G \times I : \big(\sum_{n \in \mathbb{Z}} |f * \mathcal{K}^{k,l}_{2^n s}(x)|^2 \big)^{\frac{1}{2}} > \alpha \} \Big) \lesssim (1+k) \alpha^{-1} \|f\|_1,$$

$$\text{(4.8b)}\quad \operatorname{meas}\big(\{(x,s)\in G\times I: \big(\sum_{n\in\mathbb{Z}}|f*\frac{d}{ds}\mathcal{K}^{k,l}_{2^n s}(x)|^2\big)^{1/2}>\alpha\}\big) \lesssim (1+k)2^{k-l}\alpha^{-1}\|f\|_1\,.$$

The kernels $\mathcal{K}^{k,l}_{2^n s}$ and $2^{l-k}\frac{d}{ds}\mathcal{K}^{k,l}_{2^n s}$ enjoy similar qualitative and quantitative properties and therefore we shall only give the proofs of (4.7a) and (4.8a); the proofs of (4.7b), (4.8b) require only minor notational modifications.

Proof of (4.7a). As a consequence of Proposition 3.1 and scaling we have, for each fixed n,

$$\Big(\int_I \|f*\mathcal{K}^{k,l}_{2^n s}\|^2_{L^2}ds\Big)^{1/2}\lesssim_\varepsilon 2^{-k\frac{d-1}{2}+l(\frac{d-2}{2}+\varepsilon)}\|f\|_{L^2}.$$

Note that Proposition 3.1 was stated for $\nu^{k,l}_t$, but clearly by the above discussion we can replace $\nu^{k,l}_t$ with $\mathcal{K}^{k,l}_t$.

To combine the estimates for different n we need the following variant of the Cotlar-Stein lemma. Let H_1, H_2 be Hilbert spaces and let $T_n: H_1\to H_2$, $n\in\mathbb{Z}$ be bounded operators. Assume $B\ge 2A$, and

$$\|T_n\|_{H_1\to H_2}\le A,\quad \|T_nT^*_{n'}\|_{H_2\to H_2}\le B^2 2^{-\varepsilon|n-n'|}$$

for all $n,n'\in\mathbb{Z}$. Then for all $f\in H_1$

$$\text{(4.9)}\qquad \Big(\sum_{n\in\mathbb{Z}}\|T_nf\|^2_{H_2}\Big)^{1/2}\lesssim_\varepsilon A\sqrt{\log(\varepsilon^{-1}B/A)}\|f\|_{H_1}.$$

This is proved for the case $H_1=H_2$ in [14, Lemma 3.2] but the proof also extends to the situation of two different Hilbert spaces.

We apply this with $H_1=L^2(\mathbb{R}^{d+m})$ and $H_2=L^2(\mathbb{R}^{d+m}\times[1,2])$, for the operators $T_n: H_1\to H_2$ given by $T_nf(x,s)=f*\mathcal{K}^{k,l}_{2^n s}$. Clearly we have $\|T_nT^*_{n'}\|\lesssim_\varepsilon A^2_{k,l}$ with $A_{k,l}=2^{-k(d-1)/2+l(d-2+\varepsilon)/2}$; we use this for $|n-n'|\le 2(m+2)k$. For large $|n-n'|$ we need to establish exponential decay in $|n-n'|$ but in view of the logarithmic dependence on B in (4.9) we do not have to care about any blowup in terms of powers of 2^k in such an estimate.

As $\nu^{k,l}_s$ and $2^{-k}\nabla\nu^{k,l}_s$ are for $s\approx 1$ pointwise dominated by the right hand side of (3.5) the kernels $\mathcal{K}^{k,l}_s$, $2^{-k}\nabla\mathcal{K}^{k,l}_s$, satisfy up to a constant the same bounds. Since they are also supported on a fixed common compact set we have

$$\text{(4.10)}\qquad \|\mathcal{K}^{k,l}_s\|_1+2^{-k}\|\nabla\mathcal{K}^{k,l}_s\|_1=O(1)$$

for $|s|\approx 1$. For the orthogonality arguments it is convenient to just use the following trivial pointwise bounds with an exponential dependence on k, with $N>\max\{d,m\}$:

$$\text{(4.11)}\qquad \begin{aligned}|K^{k,l}(x)|+2^{-k}|\nabla K^{k,l}(x)| &\lesssim_N 2^{k(m+1)}(1+|x|)^{-2N}\\ &\lesssim_N 2^{k(m+1)}(1+|\underline{x}|)^{-N}(1+|\overline{x}|)^{-N}.\end{aligned}$$

Clearly, (4.11) is implied by the stronger bounds in (3.5).

A standard argument using (4.11) and the cancellation property (4.2) gives the (non-optimal) estimate

$$\|\widetilde{\mathcal{K}}^{k,l}_{2^n s} * \mathcal{K}^{k,l}_{2^{n'} s}\|_1 \lesssim 2^{k(2m+3)} 2^{-|n-n'|/2};$$

here we use the notation $\widetilde{\mathcal{K}}(x) = \overline{\mathcal{K}(-x)}$. We refer to [22, ch.XIII, §5.3] for a very similar calculation. Consequently

$$\|f * \widetilde{\mathcal{K}}^{k,l}_{2^n s} * \mathcal{K}^{k,l}_{2^{n'} s}\|_2 + 2^{l-2k} \|f * \partial_s \widetilde{\mathcal{K}}^{k,l}_{2^n s} * \partial_s \mathcal{K}^{k,l}_{2^{n'} s}\|_2 \lesssim 2^{k(2m+3)} 2^{-|n-n'|/2} \|f\|_2.$$

Hence we get $\|T_{n'} T_n^*\| \lesssim B^2 2^{-|n-n'|/2}$, with $B^2 = 2^{k(2m+3)}$. We may thus apply the almost orthogonality inequality (4.9) with $\log(B/A_{k,l}) \lesssim 1 + k$, and (4.7a) follows. □

Proof of (4.8a). We use a Calderón-Zygmund decomposition of $f \in L^1(G)$, at height α, as described in [9, Ch.3A]. The Carnot-Carathéodory balls $B(0, r)$ at the orgin are given by $\{y : |\underline{y}| \le r, |\overline{y}| \le r^2\}$, and the ball centered at some z is just the left translate, *i.e.*, $B(z, r) = \{y : z^{-1} \cdot y \in B(0, r)\}$.

One can decompose an $L^1(G)$ function f as $f = g + b$ where $\|g\|_1 \lesssim \|f\|_1$, $\|g\|_\infty \lesssim \alpha$, and $b = \sum_\nu b_\nu$ where b_ν are supported on the balls $B(y_\nu, r_\nu)$ which are explicitly given by

$$B(y_\nu, r_\nu) = \{(\underline{y}, \overline{y}) : |-\underline{y}_\nu + \underline{y}| \le r_\nu, |-\overline{y}_\nu + \overline{y} - \underline{y}_\nu^\intercal \vec{S} \underline{y}| \le r_\nu^2\}.$$

Moreover, the b_ν satisfy $\int b_\nu(y) dy = 0$ and $\sum_\nu \|b_\nu\|_1 \lesssim \|f\|_1$. Finally the $B(y_\nu, r_\nu)$ have bounded overlap and if for $A \ge 2$ we define $\Omega_\alpha := \cup_\nu B(y_\nu, A r_\nu)$, then

$$\text{meas}(\Omega_\alpha) \lesssim A^{d+2m} \alpha^{-1} \|f\|_1. \tag{4.12}$$

We set $\|\vec{S}\| := \sum_{i=1}^m \|S_i\|$ (with the matrix norm associated to the Euclidean norm in $\mathbb{R}^d$) and we will choose $A \ge 10(\|\vec{S}\| + 1)$.

We now turn to the estimation of (4.8a). By Chebyshev's inequality and then (4.7a)

$$\begin{aligned} &\text{meas}\big(\{(x,s) \in G \times I : \big(\sum_{n \in \mathbb{Z}} |g * \mathcal{K}^{k,l}_{2^n s}(x)|^2\big)^{\frac12} > \alpha/2\}\big) \\ &\lesssim \alpha^{-2} \iint_{G \times I} \sum_{n \in \mathbb{Z}} |g * \mathcal{K}^{k,l}_{2^n s}|^2 dx\, ds \lesssim \alpha^{-2} \|g\|_2^2 \lesssim \alpha^{-1} \|f\|_1. \end{aligned}$$

Moreover $\text{meas}(\Omega_\alpha \times I) \le 4|\Omega_\alpha| \lesssim \alpha^{-1}\|f\|_1$. It remains to estimate the contribution involving b, namely

$$\begin{aligned}
&\text{meas}\big(\{(x,s)\in\Omega_\alpha^\complement\times I : \big(\sum_{n\in\mathbb{Z}}|b*\mathcal{K}^{k,l}_{2^ns}(x)|^2\big)^{\frac12}>\alpha/2\}\big)\\
&\lesssim \alpha^{-1}\int_I\int_{\Omega_\alpha^\complement}\big(\sum_{n\in\mathbb{Z}}|b*\mathcal{K}^{k,l}_{2^ns}(x)|^2\big)^{\frac12}\,dx\,ds\\
&\lesssim \alpha^{-1}\sum_\nu\int_I\int_{(B(y_\nu,Ar_\nu))^\complement}\big(\sum_{n\in\mathbb{Z}}|b_\nu*\mathcal{K}^{k,l}_{2^ns}(x)|^2\big)^{\frac12}\,dx\,ds\\
&\lesssim \alpha^{-1}\sum_\nu\sup_{s\in I}\sum_{n\in\mathbb{Z}}\int_{(B(y_\nu,Ar))^\complement}|b_\nu*\mathcal{K}^{k,l}_{2^ns}(x)|\,dx
\end{aligned}$$

We claim that for $s\in I$ and fixed ν,

(4.13)
$$\int_{(B(y_\nu,Ar))^\complement}|b_\nu*\mathcal{K}^{k,l}_{2^ns}(x)|\,dx \lesssim \begin{cases}1 & \text{for all } n\in\mathbb{Z}\\ 2^{k(m+1)}2^nr_\nu^{-1} & \text{for } n\le\log_2 r_\nu\\ 2^{k(m+2)}(2^{-n}r_\nu)^{1/2} & \text{for } n\ge\log_2 r_\nu\end{cases}$$

We use the first bound when $\log_2 r_\nu - 10km \le n \le \log_2 r_\nu + 10km$, the second bound when $n < \log_2 r_\nu - 10km$ and the third when $n > \log_2 r_\nu + 10km$. Summing these yields

$$\sup_{s\in I}\sum_{n\in\mathbb{Z}}\int_{(B(y_\nu,Ar))^\complement}|b_\nu*\mathcal{K}^{k,l}_{2^ns}(x)|\,dx \lesssim (1+k)\|b_\nu\|_1.$$

If we then sum over ν and use $\sum_\nu\|b_\nu\|_1\lesssim\|f\|_1$ above we get

$$\text{meas}\big(\{(x,s)\in G\times I : \big(\sum_{n\in\mathbb{Z}}|g*\mathcal{K}^{k,l}_{2^ns}(x)|^2\big)^{\frac12}>\alpha/2\}\big)\lesssim(1+k)\alpha^{-1}\|f\|_1.$$

We now prove (4.13). The $O(1)$ bound in (4.13) is immediate and follows from $\|\mathcal{K}^{k,l}_{2^ns}\|_1 = O(1)$ which is a consequence of (3.6). For the second and third case in (4.13) we will just use the trivial pointwise bounds in (4.11).

We now assume that $n\le\log_2 r_\nu$ to prove the second bound in (4.13). Here we will strongly use that the integration is extended over the complement of $(B(y_\nu,Ar_\nu))^\complement$ and split it as $X_1\cup X_2$ where

$$\begin{aligned}
X_1 &= \{(\underline{x},\overline{x}) : |\underline{x}-\underline{y}_\nu|\ge Ar_\nu\},\\
X_2 &= \{(\underline{x},\overline{x}) : |\underline{x}-\underline{y}_\nu|\le Ar_\nu, |-\overline{y}_\nu+\overline{x}-\underline{y}_\nu^\intercal\vec{S}\underline{x}|\ge A^2r_\nu^2\}.
\end{aligned}$$

Then

$$\begin{aligned}
\int_{X_1}|b_\nu*\mathcal{K}^{k,l}_{2^ns}(x)|dx \;\lesssim\; & 2^{k(m+1)}\int_{B(y_\nu,r_\nu)}|b_\nu(y)|\,\times\\
&\int_{|\underline{x}-\underline{y}_\nu|\ge Ar_\nu}\frac{(2^ns)^{-d}}{(\frac{|\underline{x}-\underline{y}|}{2^ns})^N}\int_{\overline{x}}\frac{(2^ns)^{-2m}}{(1+\frac{|\overline{x}-\overline{y}+\underline{x}^\intercal\vec{S}\underline{y}|}{(2^ns)^2})^N}d\overline{x}\,d\underline{x}\,dy\,.
\end{aligned}$$

The $\overline{x}$-inner integral is $O(1)$ and in the $\underline{x}$-integral we can use $|\underline{x}-\underline{y}| \approx |\underline{x}-\underline{y}_\nu|$ and get a bound $(Ar_\nu(2^n s)^{-1})^{d-N}$ here. Hence

$$\int_{X_1} |b_\nu * \mathcal{K}^{k,l}_{2^n s}(x)|dx \lesssim 2^{k(m+1)}(2^{-n}r_\nu)^{d-N} \lesssim 2^{k(1+m)}2^n r_\nu^{-1}\|b_\nu\|_1. \tag{4.14}$$

For the X_2-contribution

$$\int_{X_2} |b_\nu * \mathcal{K}^{k,l}_{2^n s}(x)|dx \lesssim 2^{k(m+1)}\int_{B(y_\nu,r_\nu)} |b_\nu(y)| \times \tag{4.15}$$
$$\int_{|\underline{x}-\underline{y}_\nu|\le Ar_\nu} \frac{(2^n s)^{-d}}{(1+\frac{|\underline{x}-\underline{y}|}{2^n s})^N} \int_{|\overline{x}-\overline{y}_\nu+\underline{x}^\intercal\vec{S}\underline{y}_\nu|\ge A^2r_\nu^2} \frac{(2^n s)^{-2m}}{(\frac{|\overline{x}-\overline{y}+\underline{x}^\intercal\vec{S}\underline{y}|}{(2^n s)^2})^N} d\overline{x}\, d\underline{x}\, dy\,.$$

Here we gain in the $\overline{x}$-integral. For this observe for $x\in X_2$, $y\in B(y_\nu,r_\nu)$

$$\begin{aligned}
&\Big||\overline{x}-\overline{y}+\underline{x}^\intercal\vec{S}\underline{y}| - |\overline{x}-\overline{y}_\nu+\underline{x}^\intercal\vec{S}\underline{y}_\nu|\Big| \le |\overline{y}_\nu-\overline{y}+\underline{x}^\intercal\vec{S}(\underline{y}-\underline{y}_\nu)|\\
&\le |\overline{y}_\nu-\overline{y}+\underline{y}_\nu^\intercal\vec{S}(\underline{y}-\underline{y}_\nu)| + |(\underline{x}-\underline{y}_\nu)^\intercal\vec{S}(\underline{y}-\underline{y}_\nu)|\\
&\le |\overline{y}-\overline{y}_\nu+\underline{y}^\intercal\vec{S}\underline{y}_\nu| + |\underline{x}-\underline{y}_\nu|\|\vec{S}\||\underline{y}-\underline{y}_\nu| \le r_\nu^2 + A\|\vec{S}\|r_\nu^2 \le (Ar_\nu)^2/2
\end{aligned}$$

(here we used that $|\underline{x}-\underline{y}_\nu| \le Ar_\nu$ in X_2 and $A \gg \|\vec{S}\|$). The displayed inequality tells us that we can replace $|\overline{x}-\overline{y}+\underline{x}^\intercal\vec{S}\underline{y}|$ with $|\overline{x}-\overline{y}_\nu+\underline{x}^\intercal\vec{S}\underline{y}_\nu|$ in the integrand of the inner integral in (4.15). Then we get the bound $O\big(\big(Ar_\nu(2^n s)^{-1}\big)^{-(2N-2m)}\big)$ for this inner integral. This proves the second bound in (4.13).

Finally, we prove the third bound in (4.13); assume $2^n \ge r_\nu$ and extend the integration over all of $\mathbb{R}^{d+2m}$. Let, with $\delta\in(0,1)$,

$$X_3 = \{x : |\underline{x}-\underline{y}_\nu| \ge r_\nu\,(\frac{2^n}{r_\nu})^{1+\delta}\}, \quad X_4 = X_3^\complement.$$

Clearly for $y\in B(y_\nu,r_\nu)$ we have $|\underline{x}-\underline{y}|\approx|\underline{x}-\underline{y}_\nu|$ and thus integrating first in $\overline{x}$,

$$\begin{aligned}
\int_{X_3} |b_\nu * \mathcal{K}^{k,l}_{2^n s}(x)|dx &\lesssim 2^{k(m+1)}\int_{|\underline{x}-\underline{y}_\nu|\ge r_\nu(2^n/r_\nu)^{1+\delta}} \frac{(2^n s)^{-d}}{(\frac{|\underline{x}-\underline{y}_\nu|}{2^n s})^N} d\underline{x}\,\|b_\nu\|_1\\
&\lesssim 2^{k(m+1)}(2^n r_\nu^{-1})^{-\delta(N-d)}\|b_\nu\|_1.
\end{aligned}$$

For $x\in X_4$ we use the mean value zero property of b_ν and write

$$\begin{aligned}
&b_\nu * \mathcal{K}^{k,l}_{2^n s}(x) =\\
&\int b_\nu(y)\Big[\mathcal{K}^{k,l}_{2^n s}(\underline{x}-\underline{y},\overline{x}-\overline{y}+\underline{x}^\intercal\vec{S}\underline{y}) - \mathcal{K}^{k,l}_{2^n s}(\underline{x}-\underline{y}_\nu,\overline{x}-\overline{y}_\nu+\underline{x}^\intercal\vec{S}\underline{y}_\nu)\Big]dy
\end{aligned}$$

and from this

$$\int_{X_4} |b_\nu * \mathcal{K}^{k,l}_{2^n s}(x)|dx \lesssim 2^{k(m+2)} \|b_\nu\|_1 \times \sup_{y\in B(y_\nu,r_\nu)} \Big[(2^n s)^{-1}|\underline{y}_\nu - \underline{y}| + (2^{2n}s^2)^{-1} \sup_{x\in X_4} |-\overline{y} + \overline{y}_\nu + \underline{x}^\intercal \vec{S}(\underline{y} - \underline{y}_\nu)|\Big].$$

For $x \in X_4$ and $y \in B(y_\nu, r_\nu)$,

$$\begin{aligned} |-\overline{y} + \overline{y}_\nu + \underline{x}^\intercal \vec{S}(\underline{y} - \underline{y}_\nu)| &\le |(\underline{x} - \underline{y}_\nu)^\intercal \vec{S}(\underline{y} - \underline{y}_\nu)| + |-\overline{y} + \overline{y}_\nu + \underline{y}_\nu^\intercal \vec{S}(y - y_\nu)| \\ &\le |\underline{x} - \underline{y}_\nu| \|\vec{S}\| |\underline{y} - \underline{y}_\nu| + |\overline{y} - \overline{y}_\nu - \underline{y}_\nu^\intercal \vec{S} y| \\ &\le 2r_\nu \big(\frac{2^n}{r_\nu}\big)^{1+\delta} \|\vec{S}\| r_\nu + r_\nu^2 \lesssim 2^{n(1+\delta)} r_\nu^{1-\delta} + r_\nu^2. \end{aligned}$$

Using this in the estimate above and combining with the integral over X_3 yields

$$\int |b_\nu * \mathcal{K}^{k,l}_{2^n s}(x)|dx \lesssim \|b_\nu\|_1 \Big[2^{k(m+1)} \big(\frac{r_\nu}{2^n}\big)^{\delta(N-d)} + 2^{k(m+2)} \Big(\frac{r_\nu}{2^n} + \big(\frac{r_\nu}{2^n}\big)^{1-\delta} + \frac{r_\nu^2}{2^{2n}}\Big)\Big]$$

and choosing $\delta = 1/2$ gives the third estimate in (4.13), for $2^n \ge r_\nu$. □

The case $2 \le p < \infty$. We prove (4.4). We are again using the Sobolev embedding inequality (4.5), now for $p > 2$. We proceed similarly as in the $p < 2$ case (however the proof is now simpler since we are working with ℓ^p-valued functions and not with ℓ^2-valued functions and this allows to use trivial L^∞ bounds in place of the previously used Calderón-Zygmund estimates). After uses of Hölder's inequality we get

$$\begin{aligned} \Big\| \Big(\sum_{n\in\mathbb{Z}} \|u(s) f * \mathcal{K}^{k,l}_{2^n s}\|^p_{B^{1/p}_{p,1}} \Big)^{\frac1p} \Big\|_p &\lesssim \Big(\sum_{n\in\mathbb{Z}} \int_I \big\| f * \mathcal{K}^{k,l}_{2^n s} \big\|_p^p ds \Big)^{\frac1p} \\ &\quad + \Big(\sum_{n\in\mathbb{Z}} \int_I \big\| f * \mathcal{K}^{k,l}_{2^n s} \big\|_p^p ds \Big)^{\frac{1}{pp'}} \Big(\sum_{n\in\mathbb{Z}} \int_I \big\| f * \tfrac{d}{ds} \mathcal{K}^{k,l}_{2^n s} \big\|_p^p ds \Big)^{\frac{1}{p^2}}. \end{aligned}$$

Hence (4.4) follows from

$$\Big(\sum_{n\in\mathbb{Z}} \int_I \big\| f * \mathcal{K}^{k,l}_{2^n s} \big\|_p^p ds \Big)^{\frac1p} \lesssim_\varepsilon (1+k)^{\frac1p} 2^{-k\frac{d-1}{p} + l(\frac{d-2}{p} + \varepsilon)} \|f\|_p \tag{4.16a}$$

and

$$\Big(\sum_{n\in\mathbb{Z}} \int_I \big\| f * \tfrac{d}{ds} \mathcal{K}^{k,l}_{2^n s} \big\|_p^p ds \Big)^{\frac1p} \lesssim_\varepsilon 2^{k-l} (1+k)^{\frac1p} 2^{-k\frac{d-1}{p} + l(\frac{d-2}{p} + \varepsilon)} \|f\|_p. \tag{4.16b}$$

We now have for $p = \infty$ the inequalities

$$\sup_n \sup_{s\in I} \sup_{x\in G} |f * \mathcal{K}^{k,l}_{2^n s}(x)| \lesssim \|f\|_\infty \tag{4.17a}$$

and

$$\sup_n \sup_{s\in I} \sup_{x\in G} |f * \tfrac{d}{ds}\mathcal{K}^{k,l}_{2^n s}(x)| \lesssim 2^{k-l}\|f\|_\infty \tag{4.17b}$$

which are immediate consequences of $\|\mathcal{K}^{k,l}_{2^n s}\| = O(1)$, $\|\frac{d}{ds}\mathcal{K}^{k,l}_{2^n s}\| = O(2^{k-l})$ (see (3.6)). Inequality (4.16a) follows from (4.7a), (4.17a) by interpolation, and likewise (4.16b) follows from (4.7b), (4.17b). This finishes the proof of (4.4). □

Comment on Remark 1.4. An examination of the proof above allows, for fixed p, inclusions of factors of $2^{(k-l)\beta}$ on the left hand sides of the inequalities (4.3) for $p \le 2$ and (4.4) for $p > 2$. Specifically we can have $\beta < \frac{d-1}{p'} - \frac{1}{p}$ for $\frac{d}{d-1} < p \le 2$ and $\beta < \frac{d-2}{p}$ for $p > 2$. This observation can be used to justify replacing the Besov space $B^{1/p}_{p,1}$ in the s variable with $B^{\beta+1/p}_{p,1}$ in those ranges. For the cases where $\beta + 1/p \ge 1$ one needs to use that also for $j > 1$ the terms $2^{(l-k)j}(\frac{d}{ds})^j \nu^{k,l}_s$ behave like $\nu^{k,l}_s$ (in particular this requires a straightforward extension of calculations for $j = 1$ at the end of §5.1 below).

5. Basic considerations for the L^2 estimate in Proposition 3.1

It suffices to prove the proposition for functions that are supported in a small neighborhood of the origin of diameter $\ll \varrho^2 \ll 1$ since one can use a standard argument using a tiling via the group translations to reduce to the general case (for more details we refer to §2 of [18]). We follow [18] to discuss further reductions which will simplify the forthcoming L^2 bounds.

5.1. *A shear transformation.* When acting on functions f supported in an ϱ^2-ball centered at the origin we can rewrite $f * \nu^{k,l}_t = \mathfrak{A}^{k,l} f(x,t)$ where

$$\mathfrak{A}^{k,l} f(x,t) = \int K^{k,l}_t(x,y) f(y)dy$$

and $K^{k,l}_t$ is given by

$$\begin{aligned} K^{k,l}_t(x,y) &= t^{-d-2m}\nu^{k,l}(t^{-1}(\underline{x}-\underline{y}), t^{-2}(\overline{x}-\overline{y}+\underline{x}^\intercal \vec{S}\underline{y})) \\ &= \mathring{a}(x,t,y) \iint \zeta_{k,l}(\sigma,\tau) e^{i\frac{\sigma}{t}(x_1-y_1-tg(\frac{x'-y'}{t}))+i\langle \frac{\tau}{t^2}, \overline{x}-\overline{y}+\underline{x}^\intercal \vec{S}\underline{y}\rangle} d\sigma d\tau \end{aligned} \tag{5.1}$$

with

$$\mathring{a}(x,t,y) = (2\pi)^{-m-1} t^{-d-2m} \beta_1(\tfrac{x_1-y_1}{t}, \tfrac{\overline{x}-\overline{y}+\underline{x}^\intercal \vec{S}\underline{y}}{t^2})\beta_0(\tfrac{x'-y'}{t})\chi_\varrho(y)$$

and $y \mapsto \chi_\varrho(y)$ supported in a ϱ^2 neighborhood of the origin. Notice that $\mathring{a}$ lives, where $|t-1| \le \varrho^2$, $|x_1 - 1| \lesssim \varrho^2$, $|x'|, |\overline{x}|, |y| \lesssim \varrho^2$. Introducing the frequency variables $\vartheta = (\vartheta_1, \overline{\vartheta}) \in \mathbb{R}^{m+1}$, with $\vartheta_1 = 2^{1-k}t^{-1}\sigma$, $\overline{\vartheta}_i = 2^{1-k}t^{-2}\tau_i$

we can rewrite the integral as

$$K_t^{k,l}(x,y) = 2^{(k-1)(1+m)}\mathring{a}(x,t,y)\iint \zeta_1(2^l t\vartheta_1)\upsilon_1(t,\vartheta_1,\overline{\vartheta}) \times e^{i2^{k-1}(\vartheta_1(x_1-y_1-tg(\frac{x'-y'}{t}))+\langle\overline{\vartheta},\overline{x}-\overline{y}+\underline{x}^\intercal\vec{S}\underline{y}\rangle)}d\vartheta_1 d\overline{\vartheta}. \tag{5.2}$$

where we have abbreviated

$$\upsilon_1(t,\vartheta_1,\overline{\vartheta}) = t^{1+2m}\zeta_1((t^2\vartheta_1^2+t^4|\overline{\vartheta}|^2)^{1/2}).$$

When $l=k$ we get a similar formula where $\zeta_1(2^k t\vartheta_1)$ is replaced with $\zeta_0(2^k t\vartheta_1)$.

Following [18] we rewrite the phase and verify that

$$\begin{aligned}\vartheta_1\big(x_1-y_1-tg(\tfrac{x'-y'}{t})\big)&+\sum_{i=1}^m \overline{\vartheta}_i(\overline{x}_i-\overline{y}_i+\underline{x}^\intercal S_i\underline{y})\\ &= \big(\vartheta_1-\sum_{i=1}^m\overline{\vartheta}_i\underline{x}^\intercal S_i e_1\big)\big(x_1-y_1-tg(\tfrac{x'-y'}{t})\big)\\ &+\sum_{i=1}^m\overline{\vartheta}_i\big(\overline{x}_i+x_1\underline{x}^\intercal S_i e_1-\overline{y}_i+\underline{x}^\intercal S_i P^\intercal y'-\underline{x}^\intercal S_i e_1 tg(\tfrac{x'-y'}{t})\big).\end{aligned}\tag{5.3}$$

Setting $\theta_1 = \vartheta_1 - \sum_{i=1}^m \overline{\vartheta}_i\underline{x}^\intercal S_i e_1$, $\overline{\theta}_i = \overline{\vartheta}_i$, we can write the Schwartz kernel using the $(\theta_1,\overline{\theta})$ frequency variables. Define the phase function Ψ by

$$\begin{aligned}&\Psi(x,t,y,\theta) =\\ &\theta_1(x_1-y_1-tg(\tfrac{x'-y'}{t}))+\sum_{i=1}^m\overline{\theta}_i(\overline{x}_i-\overline{y}_i+\underline{x}^\intercal S_i P^\intercal y'-\underline{x}^\intercal S_i e_1 tg(\tfrac{x'-y'}{t}))\,.\end{aligned}\tag{5.4}$$

Making the substitution $(\vartheta_1,\overline{\vartheta}) = (\theta_1+\underline{x}^\intercal S^{\overline{\theta}}e_1,\overline{\theta})$, here using the notation $S^{\overline{\theta}} = \sum_{i=1}^m\overline{\theta}_i S_i$ we obtain

$$K_t^{k,l}(x,y) = 2^{(k-1)(1+m)}\mathring{a}(x,t,y)\iint e^{i2^{k-1}\Psi(\underline{x},\overline{x}+x_1\underline{x}^\intercal\vec{S}e_1,t,y,\theta)}\times \zeta_1(2^l t(\theta_1+\underline{x}^\intercal S^{\overline{\theta}}e_1))\upsilon_1(t,\theta_1+\underline{x}^\intercal S^{\overline{\theta}}e_1,\overline{\theta})d\theta_1 d\overline{\theta}; \tag{5.5}$$

here note the nonlinear shear transformation

$$(\underline{x},\overline{x})\mapsto(\underline{x},\overline{x}+x_1\underline{x}^\intercal\vec{S}e_1)$$

which is present in the phase function. It is now natural to consider a variant $\mathcal{A}^{k,l}$ which is related to $\mathfrak{A}^{k,l}$ via this shear transformation. Let

$$\mathcal{A}^{k,l}f(x,t)\equiv\mathcal{A}_t^{k,l}f(x) = \int\mathcal{K}_t^{k,l}(x,y)f(y)dy$$

where the Schwartz kernel is given by

$$\mathcal{K}_t^{k,l}(x,y) = 2^{(k-1)(1+m)} a(x,t,y) \iint e^{i2^{k-1}\Psi(x,t,y,\theta)} \times \zeta(2^l t(\theta_1 + \underline{x}^\intercal S^{\overline{\theta}} e_1)) \upsilon(t, \theta_1 + \underline{x}^\intercal S^{\overline{\theta}} e_1, \overline{\theta}) d\theta_1 d\overline{\theta}, \tag{5.6}$$

here $s \mapsto \zeta(s)$ is supported where $|s| \approx 1$, and we use the modification that for $k = l$ we replace $\zeta(2^k t(\theta_1 + \underline{x}^\intercal S^{\overline{\theta}} e_1))$ with $\zeta_0(2^k t(\theta_1 + \underline{x}^\intercal S^{\overline{\theta}} e_1))$. We are still assuming that a is supported where

$$\operatorname{supp}(a) \subset \{(x,t,y) : |t-1| \le \varrho^2,\ |x_1 - 1| \lesssim \varrho^2,\ |x'|, |\overline{x}|, |y| \lesssim \varrho^2\}.$$

Notice that the nonlinear shear transformation does not essentially change this support assumption since by the skew-symmetry of the S_i we have $|x_1 \underline{x}^\intercal S_i e_1| \lesssim \varrho^2$.

With the choice of

$$a(x,t,y) = \mathring{a}(\underline{x}, \overline{x} - x_1 \underline{x}^\intercal \vec{S} e_1, t, y), \quad \zeta = \zeta_1, \quad \upsilon = \upsilon_1$$

we get for $l < k$

$$\mathfrak{A}^{k,l} f(\underline{x}, \overline{x}, t) = \mathcal{A}_t^{k,l} f(\underline{x}, \overline{x} + x_1 \underline{x}^\intercal \vec{S} e_1, t) \tag{5.7}$$

(and the same with $\zeta = \zeta_0$ if $k = l$).

We deduce the L^2-estimate in Proposition 3.1 from the following variant.

Proposition 5.1. *Let $\varepsilon > 0$. Then there exists a constant $C = C(\varepsilon) > 0$ such that*

$$\|\mathcal{A}^{k,l} f\|_{L^2(\mathbb{R}^{d+m} \times [\frac12, 2])} \le C 2^{-\frac{k(d-1)}{2}} 2^{l(\frac{d-2}{2} + \varepsilon)} \|f\|_{L^2(\mathbb{R}^{d+m})}, \tag{5.8}$$

with C bounded as ζ, υ, a are varying over bounded subsets of C_c^∞ (with the above support assumptions).

Proof that Proposition 5.1 implies Proposition 3.1. By (5.7) Proposition 5.1 immediately implies the first half of (3.3), by a change of variable. To prove the derivative bound in (3.3) first observe

$$\partial_t \Psi(\underline{x}, \overline{x} + x_1 \underline{x}^\intercal \vec{S} e_1, t, y, \theta) = \big(\theta_1 + \underline{x}^\intercal S^{\overline{\theta}} e_1\big) \big(\langle \tfrac{x'-y'}{t}, \nabla g(\tfrac{x'-y'}{t}) \rangle - g(\tfrac{x'-y'}{t}) \big).$$

From (5.5) we calculate that

$$\partial_t \mathfrak{A}^{k,l} f(x,t) = \sum_{i=1,2,3} \mathfrak{A}_t^{k,l,[i]} f(x) + 2^{k-l} \mathfrak{A}_t^{k,l,[4]} f(x)$$

where the Schwartz kernel of $\mathfrak{A}_t^{k,l,[i]}$ is given by $K_t^{k,l,[i]}$, defined as in (5.5) but with ζ, a, υ replaced by $\zeta^{[i]}$, $\mathring{a}^{[i]}$, $\upsilon^{[i]}$ for $i = 1,2,3,4$, resp., with the following definitions (for $l < k$)

$$\zeta^{[1]}(s) = s\zeta_1'(s), \quad \zeta^{[2]}(s) = \zeta^{[3]}(s) = \zeta_1(s), \quad \zeta^{[4]}(s) = \frac{is}{2}\zeta_1(s),$$

$$\upsilon^{[1]} = \upsilon^{[2]} = \upsilon^{[4]} = \upsilon_1, \quad \upsilon^{[3]} = \partial_t \upsilon_1,$$

$$\mathring{a}^{[1]} = t^{-1}\mathring{a}, \quad \mathring{a}^{[2]} = \partial_t \mathring{a}, \quad \mathring{a}^{[3]} = \mathring{a},$$

and
$$\mathring{a}^{[4]}(x,t,y) = t^{-1}\mathring{a}(x,t,y)\big(\langle \tfrac{x'-y'}{t}, \nabla g(\tfrac{x'-y'}{t})\rangle - g(\tfrac{x'-y'}{t})\big).$$

For $l=k$ replace ζ_1 by ζ_0. These formulas show that the derivative bound in (3.3) follows from Proposition 5.1 as well, as we have
$$\mathfrak{A}_t^{k,l,[i]} f(\underline{x}, \overline{x}) = \mathcal{A}_t^{k,l,[i]} f(\underline{x}, \overline{x} + x_1 \underline{x}^\intercal \vec{S} e_1)$$
where, with $a^{[i]}(x,t,y) = \mathring{a}^{[i]}(\underline{x}, \overline{x} - x_1\underline{x}^\intercal \vec{S} e_1, t, y)$, the operator $\mathcal{A}_t^{k,l,[i]}$ has Schwartz kernel
$$\mathcal{K}_t^{k,l,[i]}(x,y) = 2^{(k-1)(1+m)} a^{[i]}(x,t,y) \iint e^{i2^{k-1}\Psi(x,t,y,\theta)} \times \zeta^{[i]}(2^l t(\theta_1 + \underline{x}^\intercal S^{\overline{\theta}} e_1)) v^{[i]}(t, \theta_1 + \underline{x}^\intercal S^{\overline{\theta}} e_1, \overline{\theta}) d\theta_1 d\overline{\theta}.$$
Now we can use a change of variables and apply Proposition 5.1 to $\mathcal{A}_t^{k,l,[i]}$, for $i=1,2,3,4$ to complete the proof of Proposition 5.8. □

5.2. *A family of oscillatory integral operators.* It remains to prove Proposition 5.1. We reduce it to a result on oscillatory integrals acting on functions on $\mathbb{R}^d$. Here we write, $x=(x_1,x')$, $y=(y_1,y')$ for the vectors in $\mathbb{R}^d$, omitting the underbar. In what follows we are given a skew-symmetric $d\times d$ matrix S and assume that its matrix norm satisfies
$$c_0 \le \|S\| \le c_0^{-1} \tag{5.9}$$
with $0 < c_0 \le 1$; in particular the rank of S is at least 2.

We define the phase function ψ by
$$\psi(x,t,y) = y_1(x_1 - tg(\tfrac{x'-y'}{t})) + x^\intercal S(P^\intercal y' - tg(\tfrac{x'-y'}{t})e_1) \tag{5.10}$$
and set
$$\sigma(x', y_1) = y_1 + (x')^\intercal P S e_1. \tag{5.11}$$
The function ζ_1 can be split as $\zeta_1 = \zeta_1^+ + \zeta_1^-$ where $\mathrm{supp}(\zeta_1^+) \subset (\frac12, 2)$ and $\mathrm{supp}(\zeta_1^-) \subset (-2, -\frac12)$.

Setting $\lambda = 2^{k-1}$ and letting $l \le k$ we define, for functions $f \in L^2(\mathbb{R}^d)$,
$$T^{\lambda,l} f(x,t) = \int e^{i\lambda\psi(x,t,y)} \chi_l(x,t,y) f(y) dy \tag{5.12}$$
where
$$\chi_l(x,t,y) = \begin{cases} \chi(x,t,y)\zeta(2^l t\sigma(x',y_1)), & l \le k-1 \\ \chi(x,t,y)\zeta_0(2^l t\sigma(x',y_1)), & l = k. \end{cases} \tag{5.13}$$
Here χ is C_c^∞-function supported where $t\approx 1$, $|x'|, |y'| \le \varrho$, and the diameter of $\mathrm{supp}(\chi)$ does not exceed ϱ. For $l \le k-1$ we use the convention for ζ to be either ζ_1^+ or ζ_1^-. Note then that for $l\le k-1$ we have $|\sigma| \approx 2^{-l}$ on $\mathrm{supp}(\chi_l)$ and in addition the sign of σ is the same for all (x,t,y) in the support.

Proposition 5.2. *Suppose* $c_0 \le \|S\| \le c_0^{-1}$. *For* $\varepsilon > 0$,

$$\|T^{\lambda,l} f\|_{L^2(\mathbb{R}^d \times [1/2,2])} \lesssim C_\varepsilon 2^{l(\frac{d-2}{2}+\varepsilon)} \lambda^{-\frac{d}{2}} \|f\|_{L^2(\mathbb{R}^d)}\,. \tag{5.14}$$

The constant C_ε *depends on* c_0 *but not on the specific matrix* S, *and stays bounded if* $\zeta_0, \zeta_1^{\pm}, \chi$ *range over a bounded set of* C_c^∞ *functions.*

For $l \gg 1$ this is the key technical result of this paper; see §6.

5.3. *Reduction of Proposition 5.1 to oscillatory integral operators.* We will use Proposition 5.2 to deduce Proposition 5.1. The estimate is more straightforward if a can be written as a tensor product of functions of each of the variables x_i, t, y_j. To reduce to this situation we choose functions $x_i \mapsto \alpha_i(x_i)$, $t \mapsto \gamma(t)$, $y_j \mapsto \beta_j(y_j)$, $1 \le i, j \le d+m$, all with compact support such that

$$\breve{a}(x,t,y) := \gamma(t) \prod_{i=1}^{d+m} \alpha_i(x_i) \prod_{j=1}^{d+m} \beta_j(y_j)$$

equals 1 on $\mathrm{supp}(a)$, so that the support of each factor is contained in an interval of length less than 2π.

On the support of $\breve{a}$ we have the following Fourier series expansion

$$a(x,t,y) = \sum_{(n,\nu,\mu)\in\mathbb{Z}\times\mathbb{Z}^{d+m}\times\mathbb{Z}^{d+m}} c_{n,\nu,\mu} e^{itn} \prod_{i=1}^{d+m} e^{ix_i\nu_i} \prod_{j=1}^{d+m} e^{iy_j\mu_j}$$

where the coefficients $c_{n,\nu,\mu}$ are rapidly decreasing. This yields a decomposition

$$\mathcal{A}^{k,l} f(x,t) = \sum_{n,\nu,\mu} c_{n,\nu,\mu} e^{itn} \prod_{i=1}^{d+m} e^{ix_i\nu_i} \mathcal{A}_\mu^{k,l} f(x,t) \tag{5.15}$$

where $\mathcal{A}_\mu^{k,l}$ is factorized as a composition of three operators,

$$\mathcal{A}_\mu^{k,l} f(x,t) = 2^{(k-1)(m+1)} \mathcal{F}_k^m \mathcal{G}^{k,l} \mathcal{F}_{k,\mu}^{m+1} f(x,t); \tag{5.16}$$

here $\mathcal{F}_k^m$ is defined on functions $(\underline{x}, \overline{\theta}, t) \mapsto G(\underline{x}, \overline{\theta}, t)$ by

$$\mathcal{F}_k^m G(\underline{x}, \overline{x}, t) = \prod_{i=d+1}^{d+m} \alpha_i(x_i) \int_{\mathbb{R}^m} G(\underline{x}, \overline{\theta}, t) e^{i2^{k-1}\langle \overline{x}, \overline{\theta}\rangle} d\overline{\theta}, \tag{5.17}$$

$\mathcal{G}^{k,l}$ is defined on functions $(\theta_1, y', \overline{\theta}) \mapsto F(\theta_1, y', \overline{\theta})$ by

$$\begin{aligned} \mathcal{G}^{k,l} F(\underline{x}, \overline{\theta}, t) = \gamma(t) \prod_{i=1}^{d} \alpha_i(x_i) \int_{\theta_1, y'} & e^{i2^{k-1}\psi^{\overline{\theta}}(x_1, x', t, \theta_1, y')} \\ & \times \zeta_1(2^l t(\theta_1 + \underline{x}^\intercal S^{\overline{\theta}} e_1)) \upsilon(t, \theta_1, \overline{\theta}) \prod_{j=2}^{d} \beta_j(y_j) F(\theta_1, y', \overline{\theta}) d\theta_1 dy' \end{aligned} \tag{5.18}$$

with

$$\psi^{\overline{\theta}}(x_1,x',t,\theta_1,y') = \theta_1(x_1 - tg(\tfrac{x'-y'}{t})) + \underline{x}^\intercal S^{\overline{\theta}}(P^\intercal y' - tg(\tfrac{x'-y'}{t})e_1), \tag{5.19}$$

and finally $\mathcal{F}^{m+1}_{k,\mu}$ is defined on functions $(y_1,y',\overline{y}) \mapsto f(y_1,y',\overline{y})$ by

$$\begin{aligned} &\mathcal{F}^{m+1}_{k,\mu} f(\theta_1,y',\overline{\theta}) \\ &\quad = \int e^{-i2^{k-1}(y_1\theta_1 + \langle \overline{y},\overline{\theta}\rangle)} e^{i\langle\mu,y\rangle}\beta_1(y_1)\prod_{j=d+1}^{d+m}\beta_j(y_j) f(y_1,y',\overline{y})dy_1 d\overline{y}. \end{aligned} \tag{5.20}$$

We have the estimates

$$\|\mathcal{F}^m_k G\|_{L^2(\mathbb{R}^{d+m+1})} \lesssim 2^{-km/2}\|G\|_{L^2(\mathbb{R}^{d+m+1})} \tag{5.21}$$

$$\|\mathcal{G}^{k,l}F\|_{L^2(\mathbb{R}^{d+m+1})} \le C_\varepsilon 2^{l(\frac{d-2}{2}+\varepsilon)}2^{-kd/2}\|F\|_{L^2(\mathbb{R}^{d+m})} \tag{5.22}$$

$$\|\mathcal{F}^{m+1}_{k,\mu} f\|_{L^2(\mathbb{R}^{d+m})} \lesssim 2^{-k(m+1)/2}\|f\|_{L^2(\mathbb{R}^{d+m})} \tag{5.23}$$

and clearly the desired estimate (5.8) follows from (5.21), (5.22), (5.23) in conjunction with (5.15), (5.16) and the rapid decay of the $c_{n,\mu,\nu}$.

We justify the L^2 estimates. (5.21) is an immediate consequence of Plancherel's theorem in $\mathbb{R}^m$ and likewise (5.23) is a consequence of Plancherel's theorem in $\mathbb{R}^{m+1}$. It remains to consider (5.22); here we rely on Proposition 5.2. With $\psi^{\overline{\theta}}$ as in (5.19) define for functions $(\theta_1,y') \mapsto g(\theta_1,y')$

$$\begin{aligned} &\mathcal{T}^{\lambda,l}_{\overline{\theta}} g(\underline{x},t) = \\ &\int_{\theta_1,y'} \exp(i\lambda\psi^{\overline{\theta}}(\underline{x},t,\theta_1,y'))\chi^{\overline{\theta}}(\underline{x},t,\theta_1,y')\zeta_1(2^l t\sigma^{\overline{\theta}}(x',\theta_1))g(\theta_1,y')d\theta_1 dy' \end{aligned} \tag{5.24}$$

where $\sigma^{\overline{\theta}}(x',\theta_1) = \theta_1 + \underline{x}^\intercal S^{\overline{\theta}} e_1 = \theta_1 + (x')^\intercal P S^{\overline{\theta}} e_1$; moreover

$$\chi^{\overline{\theta}}(\underline{x},t,\theta_1,y') = \gamma(t)\upsilon(t,\theta_1,\overline{\theta})\prod_{i=1}^d \alpha_i(x_i)\prod_{j=2}^d \beta_j(y_j)\,.$$

By Proposition 5.2 we have with $\lambda \approx 2^k$

$$\|\mathcal{T}^{\lambda,l}_{\overline{\theta}} g\|_{L^2(\mathbb{R}^{d+1})} \lesssim 2^{l(\frac{d-2}{2}+\varepsilon)}2^{-kd/2}\|g\|_{L^2(\mathbb{R}^d)} \tag{5.25}$$

uniformly in $\overline{\theta}$; note that we have exactly the setup in (5.12), except there we use the notation x for $\underline{x}$, y_1 for θ_1, and S for $S^{\overline{\theta}}$. For the estimate (5.25) the uniformity assertion in Proposition 5.2 is crucial and so is the assumption of the S_i being linearly independent and therefore satisfying (2.12). We have $\mathcal{G}^{k,l}F(x,\overline{\theta},t) = \mathcal{T}^{2^{k-1},l}_{\overline{\theta}}[F(\cdot,\overline{\theta})]$ and thus applying (5.25) gives (5.22). This covers the case $l<k$, and the case $l=k$ is analogous, requiring a minor notational modification. This finishes the proof of Proposition 5.1, given Proposition 5.2.

6. Proof of Proposition 5.2

For small l we shall rely on a standard T^*T argument from [10]. The main part of the proof concerns the case of large l; here we rely on almost orthogonality arguments based on the Cotlar-Stein lemma, in the following version. Consider a finite set $\mathcal{V}$ indexing bounded operators $T_\nu : H_1 \to H_2$ where H_1, H_2 are Hilbert spaces. Then we have the following bound for the operator norm of the sum:

$$\Big\| \sum_{\nu\in\mathcal{V}} T_\nu \Big\|_{H_1\to H_2} \lesssim \sup_\nu \sum_{\nu'} \|T_\nu^* T_{\nu'}\|_{H_1\to H_1}^{1/2} + \sup_\nu \sum_{\nu'} \|T_\nu T_{\nu'}^*\|_{H_2\to H_2}^{1/2}. \tag{6.1}$$

This well known version follows by a simple modification of the proof in [22, ch. VII.2] (*cf.* also [7, p.223]).

6.1. *The case of small l.* This is the regime where one can use a standard T^*T argument (cf. [10]). Recall that $g(0) = 1$, $\nabla g(0) = 0$, $\mathrm{diam}(\mathrm{supp}(\chi)) \le \varrho \ll 1$, in particular $|x'|, |y'| \le \varrho \ll 1$ for $(x, t, y) \in \mathrm{supp}(\chi)$. Denote as before

$$\sigma = \sigma(x', y_1) = y_1 + (x')^\intercal PSe_1. \tag{6.2}$$

Let the $(d+1)\times d$ matrix $\partial_y^\intercal \partial_{x,t}\psi$ be defined by $(\partial_y^\intercal \partial_{x,t}\psi)_{i,j} = \partial_{x_i}\partial_{y_j}\psi$ for $1 \le i, j \le d$ and $(\partial_y^\intercal \partial_{x,t}\psi)_{d+1,j} = \partial_t \partial_{y_j}\psi$ for $1 \le j \le d$. One calculates ([18])

$$\partial_y^\intercal \partial_{x,t}\psi\big|_{(x,t,y)} = \tag{6.3}$$
$$\begin{pmatrix} 1 & e_1^\intercal SP^\intercal \\ -g'(\frac{x'-y'}{t}) & t^{-1}\sigma(x', y_1) g''(\frac{x'-y'}{t}) + PSP^\intercal + PSe_1(g'(\frac{x'-y'}{t}))^\intercal \\ -1 + \widetilde{g}(x, t, y) & -t^{-2}\sigma(x', y_1)(x'-y')^\intercal g''(\frac{x'-y'}{t}) \end{pmatrix}$$

where $\widetilde{g}(x, t, y) := 1 - g(\frac{x'-y'}{t}) + t^{-1}g'(\frac{x'-y'}{t})(x'-y')$.

Using (6.3) we obtain for the determinant of the $d\times d$ submatrix $\partial_y^\intercal \partial_x \psi$

$$\det(\partial_y^\intercal \partial_x \psi(x, t, y)) = \det\big(t^{-1}\sigma(x', y_1) g''(\tfrac{x'-y'}{t}) + PSP^\intercal\big) + O(\varrho).$$

From [14, Lemma 5.3], it follows that the matrix $t^{-1}\sigma g''(\frac{x'-y'}{t}) + PSP^\intercal$ is invertible. This says that $\partial_y^\intercal \partial_x \psi(x, t, y)$ is invertible for all $(x, t, y) \in \mathrm{supp}(\chi)$. Also, the derivatives of the amplitude $\chi(x, t, y)\zeta(2^l t\sigma(x', y_1))$ are bounded when $2^l \approx 1$. Thus the standard oscillatory integral theorem from [10] applies and we may conclude the bound $\|T^{\lambda,l} f(\cdot, t)\|_2 \le C(l)\lambda^{-d/2}\|f\|_2$ which one uses for $2^l \lesssim \varrho^{-1}$.

6.2. *The case of large l.* We may assume that $2^{-l} \ll \varrho \ll 1$ (recall from the beginning of §3 the specifications of the parameter ϱ). Choose an orthonormal basis $\mathfrak{e}_1, \dots, \mathfrak{e}_d$ with $\mathfrak{e}_1 = e_1$, and $Se_1 \in \mathrm{span}(\mathfrak{e}_2)$. Set

$$\delta_l = \max\{|Se_1|, 2^{-l}\} \tag{6.4}$$

To prepare for almost orthogonality arguments we tile $\mathbb{R}^d$ into boxes with sidelengths $(2^{l\varepsilon/d}\lambda^{-1}, 2^{l\varepsilon/d}\lambda^{-1}\delta_l^{-1}, 2^{l(1+\varepsilon/d)}\lambda^{-1}, \dots, 2^{l(1+\varepsilon/d)}\lambda^{-1})$, with the sides parallel to the $\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3, \dots, \mathfrak{e}_d$. The family of boxes $\mathfrak{s}$ can be parametrized by $\mathbb{Z}^d$; we define the lower corners by $c(\mathfrak{z}) = 2^{l\varepsilon/d}\big(\lambda^{-1}\mathfrak{z}_1\mathfrak{e}_1 + \lambda^{-1}\delta_l^{-1}\mathfrak{z}_2\mathfrak{e}_2 + \sum_{i=3}^d 2^l\lambda^{-1}\mathfrak{z}_i\mathfrak{e}_i\big)$, and let

$$\mathfrak{s}(\mathfrak{z}) = \{y : \langle c(\mathfrak{z}_1, \dots, \mathfrak{z}_d), \mathfrak{e}_i\rangle \le \langle y, \mathfrak{e}_i\rangle < \langle c(\mathfrak{z}_1{+}1, \dots, \mathfrak{z}_d{+}1), \mathfrak{e}_i\rangle,\ i = 1, \dots, d\}.$$

We also write $\mathfrak{c}_\mathfrak{s} = c(\mathfrak{z})$ if $\mathfrak{s} = \mathfrak{s}(\mathfrak{z})$. Denote by $\mathfrak{S}$ the (finite) family of those boxes which intersect $\{y : (x,t,y) \in \operatorname{supp}(\chi) \text{ for some } (x,t)\}$. We then decompose

$$T^{\lambda,l} = \sum_{\mathfrak{s}\in\mathfrak{S}} T^{\lambda,l}_\mathfrak{s}, \text{ with } T^{\lambda,l}_\mathfrak{s}[f] = T^{\lambda,l}[f\mathbb{1}_\mathfrak{s}]. \tag{6.5}$$

Note that

$$T^{\lambda,l}_\mathfrak{s}(T^{\lambda,l}_{\mathfrak{s}'})^* = 0 \text{ if } \mathfrak{s} \ne \mathfrak{s}'. \tag{6.6}$$

Notice that we have $|T^{\lambda,l}_\mathfrak{s} f(x,t)| \lesssim |\mathfrak{s}|^{1/2}\|f\|_2$. Because of the compact support of the kernel in the (x,t) variable we see that the L^2 operator norm $\|T^{\lambda,l}_\mathfrak{s}\|_{2\to2}$ is $O(|\mathfrak{s}|^{1/2})$. It is crucial for our analysis that this can be improved by a factor of $\delta_l^{1/2}$:

Lemma 6.1. *There exists a constant $C > 0$ independent of $\mathfrak{s} \in \mathfrak{S}$ such that the estimate*

$$\big\|T^{\lambda,l}_\mathfrak{s} f(\cdot,t)\big\|_{L^2(\mathbb{R}^d)} \le C\lambda^{-\frac{d}{2}} 2^{l(\frac{d-2}{2}+\varepsilon)}\|f\|_{L^2(\mathbb{R}^d)}$$

holds for every $\mathfrak{s} \in \mathfrak{S}$, with C independent of $t \in [\frac14, 4]$ and $\mathfrak{s}$.

Proof of Lemma 6.1. We freeze $t \in [\frac14, 4]$ for this proof and write $\mathcal{T}^{\lambda,l}_\mathfrak{s} f(x) = T^{\lambda,l}_\mathfrak{s} f(x,t)$, all estimates will be uniform in t.

We have $|\mathfrak{s}| \lesssim 2^{l(d-2+\varepsilon)}\delta_l^{-1}\lambda^{-d}$ and therefore obtain from the Cauchy-Schwarz inequality

$$\|\mathcal{T}^{\lambda,l}_\mathfrak{s}\|_{L^2\to L^2} \lesssim 2^{l(d-2+\varepsilon)/2}\delta_l^{-1/2}\lambda^{-d/2}.$$

Let $c_1 \ll c_0$ be a small constant, and the displayed estimate is already sufficient if $|Se_1| \ge c_1$. In what follows we consider the case $|Se_1| \le c_1$. We note that in this case

$$\sup_{w'\in\mathbb{R}^{d-1}, |w'|=1} |PSP^\intercal w'| \ge c_0/2. \tag{6.7}$$

Indeed, write $w = (w_1, w')$ and $Sw = \big(e_1^\intercal SP^\intercal w', w_1 PSe_1 + PSP^\intercal w'\big)$; we have $|e_1^\intercal SP^\intercal w'| + |w_1 PSe_1| \lesssim c_1|w|$ with $c_1 \ll c_0$ and (6.7) holds by (5.9).

Let $d_\circ$ be the smallest integer greater than or equal to $(d-1)/2$. Since $PSP^\intercal$ is skew-symmetric, there exists nonnegative numbers $s_1 \ge \dots \ge s_{d_\circ}$ and orthonormal vectors $\vec{u}_1, \dots \vec{u}_{d-1} \in \mathbb{R}^{d-1}$ such that

$$\begin{aligned} PSP^\intercal \vec{u}_{2i-1} &= s_i\vec{u}_{2i}, \\ PSP^\intercal \vec{u}_{2i} &= -s_i\vec{u}_{2i-1}, \end{aligned} \tag{6.8a}$$

for $1 \le i \le d_\circ$ if $d-1$ is even, and

$$\begin{aligned} PSP^\intercal \vec{u}_{2i-1} &= s_i \vec{u}_{2i}, \\ PSP^\intercal \vec{u}_{2i} &= -s_i \vec{u}_{2i-1}, \end{aligned} \qquad PSP^\intercal \vec{u}_{2d_\circ - 1} = 0 \tag{6.8b}$$

for $1 \le i \le d_\circ - 1$, if $d-1$ is odd. By (6.7), we have $s_1 \gtrsim c_0$.

To estimate $\mathcal{T}_{\mathfrak{s}}^{\lambda,l} f$, we further decompose the slab $\mathfrak{s}$ into smaller pieces. We may write $PSe_1 = \sum_{i=1}^{d-1} \alpha_i \vec{u}_i$ and let $b = \beta_1 \vec{u}_1 + \beta_2 \vec{u}_2$ where $\beta_1^2 + \beta_2^2 = 1$ and $\alpha_2 \beta_1 - \alpha_1 \beta_2 = 0$. Then b is a unit vector in $\mathrm{span}(\vec{u}_1, \vec{u}_2)$ with the property that $PSP^\intercal b = -\beta_2 s_1 \vec{u}_1 + \beta_1 s_1 \vec{u}_2$ is perpendicular to PSe_1. For later use notice that $|PSP^\intercal b| = s_1$.

We now decompose $\mathfrak{s}$ into subsets $\mathfrak{r}_n(\mathfrak{s})$ defined for $n \in \mathbb{Z}$ by

$$\mathfrak{r}_n(\mathfrak{s}) = \{ y = (y_1, y') \in \mathfrak{s} : 2^{l\varepsilon/d} \lambda^{-1} n \le \langle b, y' \rangle < 2^{l\varepsilon/d} \lambda^{-1}(n+1) \}. \tag{6.9}$$

Define $\mathcal{T}_{\mathfrak{s},n}^{\lambda,l} f = \mathcal{T}_{\mathfrak{s}}^{\lambda,l}[f \mathbb{1}_{\mathfrak{r}_n(\mathfrak{s})}]$ so that $\mathcal{T}_{\mathfrak{s}}^{\lambda,l} = \sum_n \mathcal{T}_{\mathfrak{s},n}^{\lambda,l}$. As $\langle P^\intercal b, e_1 \rangle = 0$ we have

$$|\mathfrak{r}_n(\mathfrak{s})| \lesssim 2^{l(d-2+\varepsilon)} \lambda^{-d},$$

and by the Cauchy-Schwarz inequality we get

$$\| \mathcal{T}_{\mathfrak{s},n}^{\lambda,l} \|_{L^2 \to L^2} \lesssim 2^{l(d-2+\varepsilon)/2} \lambda^{-d/2}. \tag{6.10}$$

Since in view of the disjointness of the sets $\mathfrak{r}_n(\mathfrak{s})$ we have $\mathcal{T}_{\mathfrak{s},n}^{\lambda,l} (\mathcal{T}_{\mathfrak{s},n'}^{\lambda,l})^* = 0$ for $n \ne n'$ it suffices, by the Cotlar-Stein Lemma, to show

$$\| (\mathcal{T}_{\mathfrak{s},n}^{\lambda,l})^* \mathcal{T}_{\mathfrak{s},n'}^{\lambda,l} \|_{L^2 \to L^2} \lesssim 2^{l(d-2+\varepsilon)} \lambda^{-d} |n - n'|^{-N} \quad \text{if } |n - n'| \ge C_1 \tag{6.11}$$

for some large C_1.

We now assume that $y \in \mathfrak{r}_n(\mathfrak{s})$, $z \in \mathfrak{r}_{n'}(\mathfrak{s})$; since both y, z belong to $\mathfrak{s}$ this means that $|n - n'| \le C 2^l$. The Schwartz kernel of $(\mathcal{T}_{\mathfrak{s},n}^{\lambda,l})^* \mathcal{T}_{\mathfrak{s},n'}^{\lambda,l}$ is given by

$$H_{n,n'}(y,z) = \mathbb{1}_{\mathfrak{r}_n(\mathfrak{s})}(y) \mathbb{1}_{\mathfrak{r}_{n'}(\mathfrak{s})} \int e^{-i\lambda \phi(x,t,y,z)} \overline{\chi_l(x,t,y)} \chi_l(x,t,y) \, dx \tag{6.12}$$

where

$$\phi(x,t,y,z) = \psi(x,t,y) - \psi(x,t,z). \tag{6.13}$$

The argument will rely on an integration by parts using the directional derivative

$$\langle v, \partial_{x'} \rangle = \sum_{i=2}^d v_i \frac{\partial}{\partial x_i} \text{ with } v = \frac{PSP^\intercal b}{|PSP^\intercal b|}. \tag{6.14}$$

Note that

$$\langle v, \partial_{x'} \rangle \phi(x,t,y,z) = \sum_{i=2}^d v_i \int_0^1 \partial_y^\intercal \partial_{x_i} \psi(x,t,w^\tau(y,z)) \, d\tau \, (y - z) \tag{6.15}$$

$$\text{where } w^\tau \equiv w^\tau(y,z) := (1-\tau) y + \tau z. \tag{6.16}$$

Using (6.3), we write

$$\partial_y^\intercal \partial_{x'}\psi\big|_{(x,t,w^\tau)}(y-z) = \langle y'-z', b\rangle PSP^\intercal b + PSP^\intercal \Pi_{b^\perp}(y'-z') + \tag{6.17}$$
$$t^{-1}\sigma(x',w_1^\tau)g''(\tfrac{x'-w^{\tau\prime}}{t})(y'-z') - g'(\tfrac{x'-w^{\tau\prime}}{t})(y_1-z_1) + PSe_1(g'(\tfrac{x'-w^{\tau\prime}}{t}))^\intercal(y'-z').$$

Since $PSP^\intercal b$ and thus v is perpendicular to both PSe_1, $PSP^\intercal\Pi_{b^\perp}(y'-z')$, and $|PSP^\intercal b| = s_1$ we have

$$\partial_y^\intercal \langle v, \partial_{x'}\rangle\psi\big|_{(x,t,w^\tau)}(y-z) = s_1\langle y'-z', b\rangle + \tag{6.18}$$
$$(ts_1)^{-1}\sigma(x',w_1^\tau)(PSP^\intercal b)^\intercal g''(\tfrac{x'-w^{\tau\prime}}{t})(y'-z') - s_1^{-1}(y_1-z_1)(PSP^\intercal b)^\intercal g'(\tfrac{x'-w^{\tau\prime}}{t}).$$

Notice from (6.2) that

$$\sigma(x',(1-\tau)y_1+\tau z_1) = (1-\tau)\sigma(x',y_1) + \tau\sigma(x',z_1). \tag{6.19}$$

Thus if $\chi_l(x,t,y)\neq 0$ and $\chi_l(x,t,z)\neq 0$ then $\sigma(x',w_1^\tau(y,z)) = O(2^{-l})$. Since $y,z\in\mathfrak{s}$, we also have $|y_1-z_1|\lesssim \lambda^{-1}2^{l\varepsilon/d}$ and $|y'-z'|\lesssim\lambda^{-1}2^{(1+\varepsilon/d)l}$. Hence the expression in the second line of display (6.18) is $O(\lambda^{-1}2^{l\varepsilon/d})$. Finally $|\langle y'-z',b\rangle|\approx|n-n'|\lambda^{-1}2^{l\varepsilon/d}$ because $(y,z)\in\mathfrak{r}_n(\mathfrak{s})\times\mathfrak{r}_{n'}(\mathfrak{s})$. Thus, we may use these observations in (6.15), (6.18) to conclude that

$$\big|\langle v,\partial_{x'}\rangle\phi(x,t,y,z)\big| \gtrsim |n-n'|\lambda^{-1}2^{l\varepsilon/d}, \quad \text{if } |n-n'|\geq C_1 \tag{6.20}$$

for a large constant C_1. This lower bound allows us to integrate by parts in the integral (6.12).

Let $\mathcal{L}$ be the formal adjoint of $g\mapsto(-\langle v,\partial_{x'}\rangle\phi)^{-1}\langle v,\partial_{x'}\rangle g$, i.e.

$$\mathcal{L}g = \langle v,\partial_{x'}\rangle\Big(\frac{g}{\langle v,\partial_{x'}\rangle\phi}\Big) = \frac{\langle v,\partial_{x'}\rangle g}{\langle v,\partial_{x'}\rangle\phi} - \frac{g\langle v,\partial_{x'}\rangle^2\phi}{(\langle v,\partial_{x'}\rangle\phi)^2}\,.$$

Setting

$$\eta_l(x,t,y,z) := \overline{\chi_l(x,t,y)}\chi_l(x,t,z) \tag{6.21}$$

we have

$$H_{n,n'}(y,z) = \mathbb{1}_{\mathfrak{r}_n(\mathfrak{s})}(y)\mathbb{1}_{\mathfrak{r}_{n'}(\mathfrak{s})}(z)\int e^{i\lambda\phi(x,t,y,z)}\frac{\mathcal{L}^N\eta_l(x,t,y,z)}{(-i\lambda)^N}\,dx. \tag{6.22}$$

In order to estimate $\mathcal{L}^N\eta_l$ we first observe that because v and PSe_1 are perpendicular we have $\langle v,\partial_{x'}\rangle\sigma(x',y_1)\equiv 0$ and $\langle v,\partial_{x'}\rangle\partial_y\sigma(x',y_1)\equiv 0$. This implies that the functions $\langle v,\partial_{x'}\rangle^j\partial_{y_i}\psi(x,t,w^\tau)$, $2\leq i\leq d$, $j\geq 2$, belong to ideal generated by $\sigma(x',y_1)$, a quantity which is $O(2^{-l})$. This in turn implies that for $(x,t,y,z)\in\operatorname{supp}(\eta_l)$, $y,z\in\mathfrak{s}$

$$|\langle v,\partial_{x'}\rangle^j\phi(x,t,y,z)|\lesssim|y_1-z_1|+2^{-l}|y'-z'|\lesssim 2^{l\varepsilon/d}\lambda^{-1}.$$

A straightforward calculation together with (6.20) shows

$$|\mathcal{L}^N\eta_l(x,t,y,z)|\lesssim\lambda^N(2^{l\varepsilon/d}|n-n'|)^{-N} \text{ for } y\in\mathfrak{r}_n(\mathfrak{s}),\ z\in\mathfrak{r}_{n'}(\mathfrak{s})$$

and from (6.22) we get

$$\sup_z \int |H_{n,n'}(y,z)|\, dy + \sup_y \int |H_{n,n'}(y,z)| dz \lesssim 2^{l(d-2+\varepsilon-N\varepsilon/d)} \lambda^{-d} |n-n'|^{-N}$$

for $|n-n'| \geq C_1$. Hence we get (6.11) by Schur's test. □

In order to finish the proof of Proposition 5.2 using Lemma 6.1 and (6.1), it remains to show that the operator norms of $(T^{\lambda,l}_{\mathfrak{s}})^* T^{\lambda,l}_{\mathfrak{s}'}$ are small if $\mathfrak{s}$, $\mathfrak{s}'$ are far apart. In order to quantify this we decompose the set of pairs $(\mathfrak{s},\mathfrak{s}')$ in families $\mathcal{U}_{\kappa_1,\kappa_2,\kappa_3}$ with $\kappa_i \in \{0,1,2,\dots\}$ which we now define. For $\mathfrak{s} \in \mathfrak{S}$, we write $c^i_{\mathfrak{s}} = \langle c_{\mathfrak{s}}, \mathfrak{e}_i \rangle$, $i=1,2$, $c^{\perp}_{\mathfrak{s}} = \Pi_{\mathrm{span}(\mathfrak{e}_1,\mathfrak{e}_2)^{\perp}} = \sum_{k=3}^{d} c^k_{\mathfrak{s}} \mathfrak{e}_k$.

Let $\kappa_1,\kappa_2,\kappa_3 \in \mathbb{N}_0 \equiv \{0,1,2,\dots\}$ such that $2^{\kappa_i} \leq 4\lambda$. To parse the following definition note that $2\lfloor 2^{\kappa-1} \rfloor = 2^{\kappa}$ if $\kappa \in \mathbb{N}$ and $2\lfloor 2^{\kappa-1}\rfloor = 0$ if $\kappa = 0$. We define $\mathcal{U}_{\kappa_1,\kappa_2,\kappa_3}$ as the set of pairs $(\mathfrak{s},\mathfrak{s}') \in \mathfrak{S} \times \mathfrak{S}$ such that

$$2\lfloor 2^{\kappa_1-1}\rfloor \lambda^{-1} \leq 2^{-l\varepsilon/d} |c^1_{\mathfrak{s}} - c^1_{\mathfrak{s}'}| \leq 2^{\kappa_1+1}\lambda^{-1}, \tag{6.23a}$$

$$2\lfloor 2^{\kappa_2-1}\rfloor 2^{\kappa_1} \lambda^{-1} \delta_l^{-1} \leq 2^{-l\varepsilon/d} |c^2_{\mathfrak{s}} - c^2_{\mathfrak{s}'}| < 2^{\kappa_2+\kappa_1+1}\delta_l^{-1}\lambda^{-1}, \tag{6.23b}$$

$$2\lfloor 2^{\kappa_3-1}\rfloor 2^{\kappa_2+\kappa_1} \lambda^{-1} 2^l \leq 2^{-l\varepsilon/d} |c^{\perp}_{\mathfrak{s}} - c^{\perp}_{\mathfrak{s}'}| \leq 2^{\kappa_3+\kappa_2+\kappa_1+1}\lambda^{-1}2^l\,. \tag{6.23c}$$

We let $\mathcal{U}^{\mathfrak{s}}_{\kappa_1,\kappa_2,\kappa_3} = \{\mathfrak{s}' \in \mathfrak{S} : (\mathfrak{s},\mathfrak{s}') \in \mathcal{U}_{\kappa_1,\kappa_2,\kappa_3}\}$. It is easy to see that for every $\mathfrak{s} \in \mathfrak{S}$

$$\mathfrak{S} = \bigcup_{\kappa_1,\kappa_2,\kappa_3 \geq 0} \mathcal{U}^{\mathfrak{s}}_{\kappa_1,\kappa_2,\kappa_3}.$$

When all κ_i are small we can use Lemma 6.1. The following lemma gives improved bounds if at least one of $\kappa_1,\kappa_2,\kappa_3$ is large.

Lemma 6.2. *For $\kappa_1,\kappa_2,\kappa_3 \in \mathbb{N}_0$, $(\mathfrak{s},\mathfrak{s}') \in \mathfrak{S}\times\mathfrak{S}$ we have the following estimates:*

(i) If $\kappa_1 \geq 5$, $\kappa_2,\kappa_3 \leq 10$ and $(\mathfrak{s},\mathfrak{s}') \in \mathcal{U}_{\kappa_1,\kappa_2,\kappa_3}$ then for all $N > 0$

$$\|(T^{\lambda,l}_{\mathfrak{s}})^* T^{\lambda,l}_{\mathfrak{s}'}\|_{L^2\to L^2} \lesssim_N 2^{-(\frac{l\varepsilon}{d}+\kappa_1)N} 2^{l(d-2+\varepsilon)} \delta_l^{-1} \lambda^{-d}\,. \tag{6.24}$$

(ii) If $\kappa_2 \geq 5$, $\kappa_3 \leq 10$ and $(\mathfrak{s},\mathfrak{s}') \in \mathcal{U}_{\kappa_1,\kappa_2,\kappa_3}$ then for all $N > 0$

$$\|(T^{\lambda,l}_{\mathfrak{s}})^* T^{\lambda,l}_{\mathfrak{s}'}\|_{L^2\to L^2} \lesssim_N 2^{-(\frac{l\varepsilon}{d}+\kappa_1+\kappa_2)N} 2^{l(d-2+\varepsilon)} \delta_l^{-1} \lambda^{-d}\,. \tag{6.25}$$

(iii) If $\kappa_3 \geq 5$ and $(\mathfrak{s},\mathfrak{s}') \in \mathcal{U}_{\kappa_1,\kappa_2,\kappa_3}$ then for all $N > 0$

$$\|(T^{\lambda,l}_{\mathfrak{s}})^* T^{\lambda,l}_{\mathfrak{s}'}\|_{L^2\to L^2} \lesssim_N 2^{-(\frac{l\varepsilon}{d}+\kappa_1+\kappa_2+\kappa_3)N} 2^{l(d-2+\varepsilon)} \delta_l^{-1} \lambda^{-d}\,. \tag{6.26}$$

Lemma 6.2 will be proved in §6.3. In each case, we will analyze for $y \in \mathfrak{s}$ and $z \in \mathfrak{s}'$ the size of the Schwartz kernel $\mathcal{K}_{\mathfrak{s},\mathfrak{s}'} \equiv \mathcal{K}^{\lambda,l}_{\mathfrak{s},\mathfrak{s}'}$ of $(T^{\lambda,l}_{\mathfrak{s}})^* T^{\lambda,l}_{\mathfrak{s}'}$ given by

$$\mathcal{K}_{\mathfrak{s},\mathfrak{s}'}(y,z) = \mathbb{1}_{\mathfrak{s}}(y)\mathbb{1}_{\mathfrak{s}'}(z) \int e^{i\lambda\phi(x,t,y,z)} \eta_l(x,t,y,z)\, dxdt \tag{6.27}$$

with η_l as in (6.21). Note that whenever $l \le k-1$ the definition (6.21) of η_l via (5.13) implies that $\sigma(x', y_1)$ and $\sigma(x', z_1)$ have the same sign, and absolute value comparable to 2^{-l}. Our proof will then rely on various integration by parts in the integral (6.27). Specifically for $(\mathfrak{s}, \mathfrak{s}') \in \mathcal{U}_{\kappa_1,\kappa_2,\kappa_3}$ we use integration by parts with respect to t, when $\kappa_1 \ge 5$, $\kappa_2, \kappa_3 \le 10$, integration by parts with respect to x_1, when $|Se_1| \ge 2^{-l}$ and $\kappa_2 \ge 5$, $\kappa_3 \le 10$, and integration by parts using the directional derivative $\langle \frac{y'-z'}{|y'-z'|}, \partial_{x'} \rangle$, either when $\kappa_3 \ge 5$ or when $\kappa_2 \ge 5$, $\kappa_3 \le 10$, $|Se_1| \le 2^{-l}$ (see §6.3 below). Assuming Lemma 6.2 we can now give the

Proof of Proposition 5.2. We verify (6.1). In view of (6.6) it suffices to prove for each $\mathfrak{s}$

$$\sum_{\mathfrak{s}'} \|(T^{\lambda,l}_{\mathfrak{s}})^* T^{\lambda,l}_{\mathfrak{s}'}\|^{1/2} \lesssim 2^{l(\frac{d-2+\varepsilon}{2})} \lambda^{-d/2} \tag{6.28}$$

with implicit constant independent of $\mathfrak{s}$. From the definition of $\mathcal{U}_{\kappa_1,\kappa_2,\kappa_3}$ it is easy to see that

$$\sup_{\mathfrak{s}} \#(\mathcal{U}^{\mathfrak{s}}_{\kappa_1,\kappa_2,\kappa_3}) \lesssim 2^{\kappa_1 d + \kappa_2(d-1) + \kappa_3(d-2)}. \tag{6.29}$$

From Lemma 6.1 we have

$$\|(T^{\lambda,l}_{\mathfrak{s}})^* T^{\lambda,l}_{\mathfrak{s}'}\| \lesssim \|T^{\lambda,l}_{\mathfrak{s}}\| \|T^{\lambda,l}_{\mathfrak{s}'}\| \lesssim 2^{l(d-2+\varepsilon)} \lambda^{-d}$$

and thus by (6.29) for $\kappa_i \le 10$, $i = 1, 2, 3$ we have

$$\sup_{\mathfrak{s}} \sum_{\kappa_1,\kappa_2,\kappa_3 \le 10} \sum_{\mathfrak{s}' \in \mathcal{U}^{\mathfrak{s}}_{\kappa_1,\kappa_2,\kappa_3}} \|(T^{\lambda,l}_{\mathfrak{s}})^* T^{\lambda,l}_{\mathfrak{s}'}\|^{1/2} \lesssim 2^{l(d-2+\varepsilon)/2} \lambda^{-d/2}. \tag{6.30}$$

Moreover using that $\delta_l^{-1} \le 2^l$ we obtain from Lemma 6.2 and (6.29)

$$\sup_{\mathfrak{s}} \sum_{\max\{\kappa_1,\kappa_2,\kappa_3\} \ge 5} \sum_{\mathfrak{s}' \in \mathcal{U}^{\mathfrak{s}}_{\kappa_1,\kappa_2,\kappa_3}} \|(T^{\lambda,l}_{\mathfrak{s}})^* T^{\lambda,l}_{\mathfrak{s}'}\|^{1/2}$$
$$\lesssim 2^{l\frac{d-2+\varepsilon}{2}} \lambda^{-\frac{d}{2}} 2^{l(\frac{1}{2} - \frac{\varepsilon N}{2d})} \sum_{(\kappa_1,\kappa_2,\kappa_3) \in \mathbb{N}_0^3} 2^{-(\kappa_1+\kappa_2+\kappa_3)(\frac{N}{2} - d)}.$$

For $N > 2d$ we can sum in $\kappa_1, \kappa_2, \kappa_3$, and if in addition we also choose $N > 1 + d/\varepsilon$ we get the bound $O(2^{l\frac{d-2+\varepsilon}{2}} \lambda^{-\frac{d}{2}})$ for the last display and (6.28) follows. □

6.3. *Proof of Lemma 6.2.* We verify first (6.24), then (6.25) in the case $|Se_1| \ge 2^{-l}$ and then give a unified treatment of (6.26) and the case $|Se_1| \le 2^{-l}$ in (6.25).

Proof of (6.24). We are now in the case $\kappa_1 \geq 5$, and $\kappa_2, \kappa_3 \leq 10$ in (6.23).

We examine the Schwartz kernel $\mathcal{K}_{\mathfrak{s},\mathfrak{s}'}$ of $(T_{\mathfrak{s}}^{\lambda,l})^* T_{\mathfrak{s}'}^{\lambda,l}$ given in (6.27); in the case under consideration we have $|y_1 - z_1| \approx 2^{\kappa_1}\lambda^{-1}2^{l\varepsilon/d}$, $|\langle y-z, \mathfrak{e}_2\rangle| \lesssim \delta_l^{-1}\lambda^{-1}2^{\kappa_1}2^{l\varepsilon/d}$, $|\langle y-z, \mathfrak{e}_i\rangle| \lesssim \lambda^{-1}2^{\kappa_1}2^{l(1+\varepsilon/d)}$, $i = 3, \dots, d$. We now integrate by parts with respect to t; for this observe that (with w^τ as in (6.16))

(6.31)
$$\partial_t\phi(x,t,y,z) = \partial_t\psi(x,t,y) - \partial_t\psi(x,t,z) = -(y_1 - z_1) + \int_0^1 \big[\widetilde{g}(x,t,w^\tau)(y_1 - z_1) - t^{-2}\sigma(x', w_1^\tau)(x' - w^{\tau\prime})^\intercal g''(\tfrac{x'-w^{\tau\prime}}{t})(y' - z')\big]d\tau.$$

Since $|x' - y'| \leq \varrho \ll 1$ we have $|\widetilde{g}(x,t,w^\tau)| \ll 1$, and from, (6.31) and $|\sigma| \lesssim 2^{-l}$ we see that

$$|\partial_t\phi(x,t,y,z)| \approx |y_1 - z_1| \approx 2^{\kappa_1}\lambda^{-1}2^{l\varepsilon/d}\,. \tag{6.32}$$

Observe that the higher t-derivatives of $\widetilde{g}$ are $\lesssim \varrho \ll 1$. Moreover σ does not depend on t and we see that

$$|\partial_t^N\phi(x,t,y,z)| \lesssim_N |y_1 - z_1| + 2^{-l}|y' - z'| \lesssim 2^{\kappa_1}2^{l\varepsilon/d}\lambda^{-1},$$
$$|\partial_t^N[\eta_l(x,t,y,z)]| \lesssim_N 1.$$

Hence integration by parts with respect to t yields the pointwise bound $|\mathcal{K}_{\mathfrak{s},\mathfrak{s}'}(y,z)| \lesssim (2^{\kappa_1}2^{l\varepsilon/d})^{-N}$ which gives

$$\sup_y \int |\mathcal{K}_{\mathfrak{s},\mathfrak{s}'}(y,z)|dz + \sup_z \int |\mathcal{K}_{\mathfrak{s},\mathfrak{s}'}(y,z)|dy \lesssim_N \frac{\lambda^{-d}2^{l(d-2+\varepsilon)}\delta_l^{-1}}{(2^{\kappa_1}2^{l\varepsilon/d})^N}.$$

As $\delta_l^{-1} \leq 2^l$ we obtain (6.24), by Schur's test.

Proof of (6.25) *in the case* $|Se_1| \geq 2^{-l}$. This now concerns the case $\kappa_2 \geq 5$. We will integrate by parts with respect to x_1 in (6.27) and observe

$$\partial_{x_1}\phi(x,t,y,z) = \partial_{x_1}\psi(x,t,y) - \partial_{x_1}\psi(x,t,z)$$
$$= y_1 - z_1 + e_1^\intercal SP^\intercal(y' - z') = y_1 - z_1 - |Se_1|\langle y' - z', \mathfrak{e}_2\rangle.$$

In the present case $|Se_1| = \delta_l$ and $(\mathfrak{s},\mathfrak{s}') \in \mathcal{U}_{\kappa_1,\kappa_2,\kappa_3}$ with $\kappa_2 \geq 5$ and thus for $y \in \mathfrak{s}$, $z \in \mathfrak{s}'$

$$|\langle y-z, \mathfrak{e}_2\rangle| \geq 2^{\kappa_2-1+\kappa_1}\lambda^{-1}\delta_l^{-1}2^{l\varepsilon/d}, \qquad |y_1 - z_1| \leq 2^{\kappa_1+1}\lambda^{-1}2^{l\varepsilon/d};$$

hence

$$|\partial_{x_1}\phi(x,t,y,z)| \approx |Se_1||\langle y' - z', \mathfrak{e}_2\rangle| \approx 2^{\kappa_1+\kappa_2}\lambda^{-1}2^{l\varepsilon/d}. \tag{6.33}$$

Note that σ does not depend on x_1 and $\partial_{x_1}^N\phi = 0$ for $N \geq 2$. After N-fold integration by parts with respect to x_1 we get $|\mathcal{K}_{\mathfrak{s},\mathfrak{s}'}(y,z)| \lesssim (2^{\kappa_1+\kappa_2}2^{l\varepsilon/d})^{-N}$. As above, the asserted estimate (6.25) follows by Schur's test. □

Proof of (6.25) *in the case* $|Se_1| \le 2^{-l}$ *and proof of* (6.26). Notice that in view of the small support of χ we have in the present case $\mathcal{K}^{\lambda,l}_{\mathfrak{s},\mathfrak{s}'} = 0$ when $2^l\lambda^{-1} \ge 1$, so the case $l = k$ is trivial. In what follows we assume $l \le k-1$; it will be crucial that in this case $\sigma(x', y_1)$, $\sigma(x', z_1)$ have the same sign for $y \in \mathfrak{s}$ and $z \in \mathfrak{s}'$.

If $|Se_1| \le 2^{-l}$ we have $\delta_l = 2^{-l}$ and for the proof of (6.25) we have also $\kappa_3 \le 10$ and we shall prove the pointwise estimate

$$|\mathcal{K}_{\mathfrak{s},\mathfrak{s}'}(y,z)| \lesssim_N (2^{\kappa_1+\kappa_2}2^{l\varepsilon/d})^{-N} \tag{6.34}$$

under the assumption that $y \in \mathfrak{s}$, $z \in \mathfrak{s}'$ satisfy

$$\begin{aligned} &2^{\kappa_1+\kappa_2-1}\lambda^{-1}2^l \le 2^{-l\varepsilon/d}|\langle y-z, \mathfrak{e}_2\rangle| \le 2^{\kappa_1+\kappa_2+2}\lambda^{-1}2^l, \\ &2^{-l\varepsilon/d}|(y-z)^\perp| \lesssim 2^{\kappa_1+\kappa_2+11}\lambda^{-1}2^l, \quad |Se_1| \le 2^{-l}. \end{aligned} \tag{6.35}$$

here $(y-z)^\perp := \sum_{i=3}^d \langle y-z, \mathfrak{e}_i\rangle \mathfrak{e}_i$.

Moreover for (6.26) we shall prove

$$|\mathcal{K}_{\mathfrak{s},\mathfrak{s}'}(y,z)| \lesssim_N (2^{\kappa_1+\kappa_2+\kappa_3}2^{l\varepsilon/d})^{-N} \tag{6.36}$$

under the assumption that $\kappa_3 \ge 5$ and that $y \in \mathfrak{s}$, $z \in \mathfrak{s}'$ satisfy

$$\begin{aligned} &2^{\kappa_1+\kappa_2+\kappa_3-1}\lambda^{-1}2^l \le 2^{-l\varepsilon/d}|(y-z)^\perp| \le 2^{\kappa_1+\kappa_2+\kappa_3+2}\lambda^{-1}2^l, \\ &2^{-l\varepsilon/d}|\langle y-z, \mathfrak{e}_2\rangle| \le 2^{\kappa_1+\kappa_2+2}\delta_l^{-1}\lambda^{-1}. \end{aligned} \tag{6.37}$$

We use the directional derivative $\langle \frac{y'-z'}{|y'-z'|}, \partial_{x'}\rangle$ in our integration by parts argument. From (6.3) we get (with w^τ as in (6.16))

$$\begin{aligned} \partial_{x'}\phi(x,t,y,z) &= \int_0^1 \partial_y^\intercal \partial_{x'}\psi(x,t,w^\tau)(y-z)d\tau \\ &= \int_0^1 \Big[-g'(\tfrac{x'-w^{\tau\prime}}{t})(y_1-z_1) + \frac{\sigma(x',w_1^\tau)}{t} g''(\tfrac{x'-w^{\tau\prime}}{t})(y'-z') \\ &\qquad + PSP^\intercal(y'-z') + PSe_1(g'(\tfrac{x'-w^{\tau\prime}}{t})^\intercal(y'-z'))\Big]\, d\tau. \end{aligned} \tag{6.38}$$

Take the scalar product with $\frac{y'-z'}{|y'-z'|}$ and use that $(y'-z')^\intercal PSP^\intercal(y'-z') = 0$ to get

$$\begin{aligned} \langle \tfrac{y'-z'}{|y'-z'|}, \partial_{x'}\rangle \phi(x,t,y,z) \;&= \int_0^1 \frac{\sigma(x',w_1^\tau)}{t} d\tau\, \frac{(y'-z')^\intercal g''(0)(y'-z')}{|y'-z'|} \\ &\quad + R_1(x,t,y,z) + R_2(x,t,y,z) \end{aligned} \tag{6.39}$$

where

$$R_1(x,t,y,z) = \big(\tfrac{y'-z'}{|y'-z'|}\big)^\intercal \int_0^1 \frac{\sigma(x',w_1^\tau)}{t}\big(g''(\tfrac{x'-w^{\tau\prime}}{t}) - g''(0)\big)d\tau(y'-z')$$

and

$$R_2(x,t,y,z) =$$
$$\int_0^1 \Big[-\langle \tfrac{y'-z'}{|y'-z'|}, g'(\tfrac{x'-w^{\tau\prime}}{t})\rangle (y_1-z_1) + \langle \tfrac{y'-z'}{|y'-z'|}, PSe_1\rangle (g'(\tfrac{x'-w^{\tau\prime}}{t})^{\intercal}(y'-z'))\Big] d\tau$$
$$= -\big(y_1 - z_1 - |Se_1|\langle y'-z', \mathfrak{e}_2\rangle\big) \int_0^1 g'(\tfrac{x'-w^{\tau\prime}}{t})^{\intercal}\big(\tfrac{y'-z'}{|y'-z'|}\big) d\tau.$$

By the single-signedness of σ we have $|\int_0^1 t^{-1}\sigma(x', w_1^\tau)d\tau| \gtrsim 2^{-l}$; here we use (6.19). Hence, because of the positive definiteness of $g''(0)$ we see that the main term in (6.39) satisfies the lower bound

$$\Big|\int_0^1 \frac{\sigma(x', w_1^\tau)}{t} d\tau \, \frac{(y'-z')^\intercal g''(0)(y'-z')}{|y'-z'|}\Big| \gtrsim 2^{-l}|y'-z'|$$

and we use

$$2^{-l}|y'-z'| \approx 2^{-l}|\langle y-z, \mathfrak{e}_2\rangle| \approx 2^{l\varepsilon/d} 2^{\kappa_1+\kappa_2}\lambda^{-1} \quad \text{if (6.35) holds,}$$
$$2^{-l}|y'-z'| \approx 2^{-l}|(y-z)^\perp| \approx 2^{l\varepsilon/d} 2^{\kappa_1+\kappa_2+\kappa_3}\lambda^{-1} \quad \text{if (6.37) holds.}$$

Since $\|g''(\frac{x'-w^{\tau\prime}}{t}) - g''(0)\| = O(\varrho)$ we get

$$|R_1(x,t,y,z)| \lesssim \varrho 2^{-l}|y'-z'| \lesssim \begin{cases} \varrho 2^{\kappa_1+\kappa_2}\lambda^{-1}2^{l\varepsilon/d} & \text{if (6.35) holds} \\ \varrho 2^{\kappa_1+\kappa_2+\kappa_3}\lambda^{-1}2^{l\varepsilon/d} & \text{if (6.37) holds.} \end{cases}$$

Finally

$$|R_2(x,t,y,z)| \lesssim \varrho\big(|y_1-z_1| + |Se_1||\langle y'-z', \mathfrak{e}_2\rangle|\big)$$

and we have $|y_1 - z_1| \lesssim 2^{\kappa_1}\lambda^{-1}2^{l\varepsilon/d}$ and thus clearly

$$|R_2(x,t,y,z)| \lesssim \varrho\big(|y_1-z_1| + 2^{-l}|y'-z'|\big) \lesssim \varrho 2^{\kappa_1+\kappa_2}\lambda^{-1}2^{l\varepsilon/d} \text{ if (6.35) holds.}$$

Moreover we get this when (6.37) holds and $|Se_1| \le 2^{-l}$.

Now consider the case that (6.37) holds and $|Se_1| \ge 2^{-l}$. Then

$$|Se_1||\langle y'-z', \mathfrak{e}_2\rangle| \lesssim 2^{\kappa_1+\kappa_2+2}\lambda^{-1}2^{l\varepsilon/d}$$

and thus we also get

$$|R_2(x,t,y,z)| \lesssim \varrho 2^{\kappa_1+\kappa_2}\lambda^{-1}2^{l\varepsilon/d} \text{ if (6.37) holds.}$$

Altogether, for $y \in \mathfrak{s}$, $z \in \mathfrak{s}'$,

$$|\langle \tfrac{y'-z'}{|y'-z'|}, \partial_{x'}\rangle \phi(x,t,y,z)| \gtrsim 2^{\kappa_1+\kappa_2}\lambda^{-1}2^{l\varepsilon/d} \text{ if (6.35) holds,} \tag{6.40}$$

and

$$|\langle \tfrac{y'-z'}{|y'-z'|}, \partial_{x'}\rangle \phi(x,t,y,z)| \gtrsim 2^{\kappa_1+\kappa_2+\kappa_3}\lambda^{-1}2^{l\varepsilon/d} \text{ if (6.37) holds.} \tag{6.41}$$

We need corresponding upper bounds for the higher derivatives $\langle \frac{y'-z'}{|y'-z'|}, \partial_{x'}\rangle$. First observe that

$$\langle \tfrac{y'-z'}{|y'-z'|}, \partial_{x'}\rangle \sigma(x', y_1) = \langle \tfrac{y'-z'}{|y'-z'|}, PSe_1\rangle. \tag{6.42}$$

Clearly this is $O(2^{-l})$ when $\delta_l = 2^{-l}$, in particular under assumption (6.35). On the other hand, if $\delta_l > 2^{-l}$ then we use that $PSe_1 = |Se_1|\mathfrak{e}_2$ and if we now assume (6.37) we have $|\langle y'-z', PSe_1\rangle| \le \delta_l |\langle y-z, \mathfrak{e}_2\rangle| \le 2^{\kappa_1+\kappa_2+2}\lambda^{-1}2^{l\varepsilon/d}$ and $|y'-z'| \ge |(y-z)^{\perp}| \ge 2^{\kappa_1+\kappa_2+\kappa_3-1}\lambda^{-1}2^l 2^{l\varepsilon/d}$; hence $\langle \frac{y'-z'}{|y'-z'|}, PSe_1\rangle = O(2^{-l})$ and therefore $\langle \frac{y'-z'}{|y'-z'|}, \partial_{x'}\rangle \sigma(x', y_1) = O(2^{-l})$. Moreover, for the higher derivatives we have $\langle \frac{y'-z'}{|y'-z'|}, \partial_{x'}\rangle^N \sigma = 0$ for $N \ge 2$. This implies, for all N,

$$\big|\langle \tfrac{y'-z'}{|y'-z'|}, \partial_{x'}\rangle^N [\eta_l(x,t,y,z)]\big| \lesssim_N 1.$$

Differentiating in (6.39) and using these estimates for σ and $\langle \frac{y'-z'}{|y'-z'|}, \partial_{x'}\rangle\sigma$, also yields

$$\begin{aligned}\big|\langle \tfrac{y'-z'}{|y'-z'|}, \partial_{x'}\rangle^N \phi(x,t,y,z)\big| &\lesssim 2^{-l}|y'-z'| + |y_1-z_1| + |Se_1||\langle y'-z', \mathfrak{e}_2\rangle| \\ &\lesssim \begin{cases} 2^{\kappa_1+\kappa_2}\lambda^{-1}2^{l\varepsilon/d} & \text{if (6.35) holds} \\ 2^{\kappa_1+\kappa_2+\kappa_3}\lambda^{-1}2^{l\varepsilon/d} & \text{if (6.37) holds} \end{cases}\end{aligned}$$

An integration by parts yields

$$\mathcal{K}_{\mathfrak{s},\mathfrak{s}'}(y,z) = \mathbb{1}_{\mathfrak{s}}(y)\mathbb{1}_{\mathfrak{s}'}(z) \int e^{i\lambda\phi(x,t,y,z)} \frac{\mathcal{L}^N \eta_l(x,t,y,z)}{(-i\lambda)^N}\, dxdt \tag{6.43}$$

where

$$\mathcal{L}g(x,t,y,z) = \langle \frac{y'-z'}{|y'-z'|}, \partial_{x'}\rangle \Big(\frac{g}{\frac{y'-z'}{|y'-z'|}\partial_{x'}\phi} \Big);$$

and we have

$$|\mathcal{L}^N[\eta_l(x,t,y,z)]| \lesssim \begin{cases} (2^{\kappa_1+\kappa_2}2^{l\varepsilon/d})^{-N}\lambda^N & \text{if (6.35) holds} \\ (2^{\kappa_1+\kappa_2+\kappa_3}2^{l\varepsilon/d})^{-N}\lambda^N & \text{if (6.37) holds.} \end{cases}$$

By (6.43) this leads to the pointwise estimates (6.34) (under assumption (6.35)) and (6.36) (under assumption (6.37)). By applying Schur's test we obtain the claimed bounds in both cases. □

7. Open problems and further directions

7.1. $d = 2$. The problem of nontrivial L^p bounds for the Nevo-Thangavelu maximal operator when $d = 2$ remains currently open even in the model case of the Heisenberg group $\mathbb{H}^1$.

7.2. *A restricted weak type endpoint bound.* Theorem 1.5 establishes a restricted weak type $(\frac{d}{d-1}, \frac{d}{d-1})$ endpoint estimate for the local maximal operator, when $d \ge 3$. Does this endpoint bound also hold for the global operator? This is the case when all J_i are zero (*cf.* [3]).

7.3. *L^p-improving estimates.* One can ask whether the local operator $f \mapsto \sup_{1 \le t \le 2} |f * \mu_t|$ maps L^p to L^q for some $q > p$; this would imply corresponding sparse bounds for the global maximal operator (see [2]). As a model case for the case $m = 1$ the precise q-range for such results should depend on the rank of J_1 (and no L^p improving takes place when $J_1 = 0$). For $m \ge 2$ the dependence on the matrices $J_1, \dots, J_m$ could be quite complicated. The case of Heisenberg type groups is covered in [18].

7.4. *Restricted dilation sets.* One can also consider maximal functions with restricted dilation sets. The $L^p \to L^p$ estimates with Minkowski dimension type assumptions are rather straightforward; one can combine the methods of this paper with elementary arguments in [20, 19]. In contrast the L^p-improving estimates are harder; for the Heisenberg groups $\mathbb{H}^n$, with $n \ge 2$, this problem was considered in [19]. For general dilation sets there is a large variety of possible type sets (*cf.* [17, Thm.1.2]), and much remains open.

7.5. *Higher step groups.* It would be interesting to develop versions of our theorem which apply in the general setting of stratified groups; here only the case of lacunary dilations is well understood (see e.g. [8]).

7.6. *Averages over tilted measures.* The above problems can also be formulated for the case where the spherical measure μ is no longer supported in a subspace invariant under the automorphic dilations. The assumption of invariance under automorphic dilations is crucial for the analysis in the present paper but it has been relaxed in [1, 18] which cover results on maximal functions associated with such tilted measures on Heisenberg or Heisenberg type groups.

References

1. Theresa C. Anderson, Laura Cladek, Malabika Pramanik, and Andreas Seeger, *Spherical means on the Heisenberg group: stability of a maximal function estimate.* J. Anal. Math. 145 (2021), no. 1, 1–28. MR 4361901
2. Sayan Bagchi, Sourav Hait, Luz Roncal, and Sundaram Thangavelu. *On the maximal function associated to the spherical means on the Heisenberg group*, New York J. Math. 27 (2021), 631–675. MR 4250270.
3. Jean Bourgain, *Estimations de certaines fonctions maximales.* C. R. Acad. Sci. Paris Sér. I Math.301(1985), no.10, 499–502. MR 0812567
4. ______, *Averages in the plane over convex curves and maximal operators.* J. Analyse Math. 47 (1986), 69–85. MR 0874045
5. David Beltran, Shaoming Guo, Jonathan Hickman, and Andreas Seeger, *The circular maximal operator on Heisenberg radial functions*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) 23 (2022), no. 2, 501–568. MR 4453958
6. Chin Hung Ching, *Pseudo-differential operators with nonregular symbols.* J. Differential Equations 11 (1972), 436–447. MR 0296764
7. Andrew Comech, *Integral operators with singular canonical relations,* chapter (pp. 200–248) in Spectral Theory, Microlocal Analysis, Singular Manifolds (M. Demuth, E. Schrohe, B.-W. Schulze, J. Sjostrand, eds.). Akademie Verlag, Berlin, 1997. MR 1608283

8. Aswin Govindan Sheri, Jonathan Hickman and James Wright. *Lacunary maximal functions on homogeneous groups*, J. Funct. Anal. 286 (2024), no.3, Paper No. 110250, 25 pp. MR 4669588
9. Gerald B. Folland, Elias M. Stein. Hardy spaces on homogeneous groups. Math. Notes, 28 Princeton University Press, Princeton, NJ; University of Tokyo Press, Tokyo, 1982. xii+285 pp. MR 0657581
10. Lars Hörmander, *Oscillatory integrals and multipliers on FL^p*, Ark. Mat. **11** (1973), 1–11. MR 340924
11. Joonil Kim, *Annulus maximal averages on variable hyperplanes*, arXiv:1906.03797, 2019.
12. Juyoung Lee and Sanghyuk Lee, $L^p \to L^q$ estimates for the circular maximal operator on Heisenberg radial functions. Math. Ann. 385 (2023), no. 3-4, 1521–1544. MR 4566682
13. Naijia Liu and Lixin Yan, *Singular spherical maximal operators on a class of degenerate two-step nilpotent Lie groups*, Math. Z. 304 (2023), no. 1, 16. MR 4581163
14. Detlef Müller and Andreas Seeger, *Singular spherical maximal operators on a class of two step nilpotent Lie groups*, Israel J. Math. 141 (2004), 315–340. MR 2063040
15. E. K. Narayanan and Sundaram Thangavelu, *An optimal theorem for the spherical maximal operator on the Heisenberg group*, Israel J. Math. 144 (2004), 211–219. MR 2121541
16. Amos Nevo and Sundaram Thangavelu, *Pointwise ergodic theorems for radial averages on the Heisenberg group*, Adv. Math. 127 (1997), no. 2, 307–334. MR 1448717
17. Joris Roos and Andreas Seeger, *Spherical maximal functions and fractal dimensions of dilation sets*, Amer. J. Math. 145, no. 4 (2023). MR 4660419
18. Joris Roos, Andreas Seeger, and Rajula Srivastava, *Lebesgue space estimates for spherical maximal functions on Heisenberg groups*, Int. Math. Res. Not., IMRN 2022, no. 24, 19222–19257. MR 4523247
19. ———, *Spherical maximal operators on Heisenberg groups: Restricted dilation sets.* Studia Math., 273 (2023), no.1, 1–28. MR 4660419
20. Andreas Seeger, Stephen Wainger, and James Wright. *Pointwise convergence of spherical means.* Math. Proc. Cambridge Philos. Soc. 118 (1995), no. 1, 115–124. MR 1329463
21. Elias M. Stein, *Maximal functions. I. Spherical means*, Proc. Nat. Acad. Sci. U.S.A. 73 (1976), no. 7, 2174–2175. MR 420116
22. ———, *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*, Princeton Mathematical Series, vol. 43, Princeton University Press, Princeton, NJ, 1993, With the assistance of Timothy S. Murphy, Monographs in Harmonic Analysis, III. MR 1232192

Jaehyeon Ryu: School of Mathematics, Korea Institute for Advanced Study, Seoul 02455, Republic of Korea & Department of Mathematics, University of Wisconsin, 480 Lincoln Drive, Madison, WI, 53706, USA.

Email address: jhryu@kias.re.kr

Andreas Seeger: Department of Mathematics, University of Wisconsin, 480 Lincoln Drive, Madison, WI, 53706, USA.

Email address: seeger@math.wisc.edu