

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 non-degeneracy condition 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 [15]. 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, $\mathfrak{g} = \mathfrak{w}_{\text{hor}} \oplus \mathfrak{w}_{\text{vert}}$, where $\dim \mathfrak{w}_{\text{hor}} = d$, $\dim \mathfrak{w}_{\text{vert}} = m$, and $\mathfrak{w}_{\text{vert}} \subseteq \mathfrak{z}(\mathfrak{g})$, with $\mathfrak{z}(\mathfrak{g})$ the center of the Lie algebra. We use the natural parabolic dilation structure on $\mathfrak{w}_{\text{hor}} \oplus \mathfrak{w}_{\text{vert}}$, and define for $\underline{X} \in \mathfrak{w}_{\text{hor}}$, $\overline{X} \in \mathfrak{w}_{\text{vert}}$, $\delta_t(\underline{X}, \overline{X}) = (t\underline{X}, t^2\overline{X})$. Using exponential coordinates on the group we identify G with $\mathfrak{w}_{\text{hor}} \oplus \mathfrak{w}_{\text{vert}} \equiv \mathbb{R}^d \oplus \mathbb{R}^m$. With $x = (\underline{x}, \overline{x}) \in \mathbb{R}^d \times \mathbb{R}^m$ the group law then becomes

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

where $\underline{x}^\top \vec{J} \underline{y} := \sum_{i=1}^m \overline{e}_i \underline{x}^\top J_i \underline{y}$, $\{\overline{e}_i\}_{i=1}^m$ 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.

Let Ω be a bounded open convex domain in $\mathfrak{w}_{\text{hor}} \equiv \mathbb{R}^d \times \{0\}$ 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

$$(1.2) \quad Af(x, t) = f * \mu_t(x) = \int_{\Sigma} f(\underline{x} - t\omega, \bar{x} - t\underline{x}^{\top} \vec{J}\omega) d\mu(\omega).$$

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

$$(1.3) \quad \mathfrak{M}f(x) := \sup_{t>0} |f * \mu_t(x)|.$$

The objective is to establish an $L^p(\mathbb{R}^{d+m}) \rightarrow L^p(\mathbb{R}^{d+m})$ bound for 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 [13] and independently by Narayanan and Thangavelu [14]. The paper [13] 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 [11]).

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 $p > \frac{d}{d-1}$, $d \geq 2$ by applying Stein’s result [20] (or Bourgain’s result [4] when $d = 2$) in the horizontal hyper-planes and then integrating in $\mathfrak{w}_{\text{vert}}$. 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 [13, 14, 10, 17] for the Heisenberg and Métivier cases seem to apply; this was posed as a problem in [13]. 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 [12] 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:*

- (i) *For $u \in C_c^\infty((0, \infty))$, $\lambda > 0$, $f \in L^p$ the functions $s \mapsto u(s)Af(x, \lambda s)$ are continuous and belong to the Besov space $B_{p,1}^{1/p}$, for almost every $x \in G$.*
- (ii) *The maximal operator \mathfrak{M} extends to a bounded operator on $L^p(G)$.*

Here $B_{p,1}^{1/p}(\mathbb{R})$ is the standard Besov space which is embedded in the space of bounded continuous functions. We shall in fact prove for all $\lambda > 0$

$$(1.4) \quad \left(\int_G [|u(\cdot)Af(x, \lambda \cdot)|_{B_{p,1}^{1/p}(\mathbb{R})}]^p dx \right)^{1/p} \lesssim \|f\|_p$$

where $u \in C_c^\infty((0, \infty))$. 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. By a standard argument Theorem 1.1 implies

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

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 1.3. *Let $p \geq \frac{d}{d-1}$ 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 in the above theorems is shown by modifying an example of Stein [20], see also the discussion in [12].

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 measure ν and formulate the main Proposition 3.1 to be proved for the boundedness of the local maximal operator. The arguments to extend to the global maximal operator are taken from [13]; we include a sketch 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 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 $a \leq Cb$ 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 and Brian Street for conversations on general pointwise convergence issues. 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 1.1 and 1.3 are reduced to these results.

Theorem 2.1. *Let S_1, \dots, S_m be a linearly independent set of skew symmetric matrices. Let \mathcal{U} be a neighborhood of the origin of \mathbb{R}^{d-1} and let $g : \mathcal{U} \rightarrow \mathbb{R}$ 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 an 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*

$$(2.1) \quad \mathcal{A}f(x, t) = \int f(\underline{x} - t\Gamma_g(\omega'), \bar{x} - t \sum_{i=1}^m e_{d+i} \underline{x}^\top S_i \Gamma_g(\omega')) \beta_0(\omega') d\omega'.$$

Let $u \in C_c^\infty((0, \infty))$. Then for $p > \frac{d}{d-1}$ the inequalities

$$(2.2a) \quad \sup_{\lambda > 0} \| |u\mathcal{A}f(\cdot, \lambda \cdot)| \|_{B_{p,1}^{1/p}(\mathbb{R})} \|_{L^p(\mathbb{R}^{d+m})} \leq C \|f\|_{L^p(\mathbb{R}^{d+m})},$$

$$(2.2b) \quad \left\| \sup_{t > 0} |\mathcal{A}_t f| \right\|_{L^p(\mathbb{R}^{d+m})} \leq C \|f\|_{L^p(\mathbb{R}^{d+m})},$$

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

$$(2.3) \quad \left\| \text{ess sup}_{t \in I} |\mathcal{A}_t f| \right\|_{L^{p,\infty}(\mathbb{R}^{d+m})} \leq C_I \|f\|_{L^{p,1}(\mathbb{R}^{d+m})}, \quad p = \frac{d}{d-1}.$$

We fix a point $y_\circ \in \Sigma$ and show that the operator $f \mapsto f * (\chi^{y_\circ} \mu)_t$ satisfies the asserted inequalities where χ^{y_\circ} is a C_c^∞ function supported in a neighborhood of y_\circ . Once this is achieved we can use a compactness and partition of unity argument to deduce Theorem 1.1. 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 $e_{\circ,i}$, $2 \leq i \leq d$ 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 $(\omega')^\top = (\omega_2, \dots, \omega_d)$ and the function G satisfies

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

We then have for $\nu = \chi\mu$

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

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

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

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

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

$$(2.4) \quad \mathcal{A}_t^{[1]} f(x) = \int f(\underline{x} - t\Gamma_G(\omega'), \bar{x} - t \sum_{i=1}^m e_{d+i} \underline{x}^\top \tilde{J}_i \Gamma_G(\omega')) \chi_\circ(\omega') d\omega'.$$

Then we compute

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

Since $R^{-1} = R^\top$ it suffices to prove (1.4) and the maximal bounds $\mathcal{A}^{[1]}$ in place of A .

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

We then have

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

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

$$\bar{v}_i(\underline{x}, \omega') = \underline{x}^\top \tilde{J}_i(g(\omega') + b^\top \omega')e_1 + \underline{x}^\top \tilde{J}_i P^\top \omega'.$$

Now write

$$(2.5) \quad x_1 - tg(\omega') - tb^\top \omega' = x_1 - b^\top x' - tg(\omega') + b^\top (x' - t\omega')$$

and

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

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

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

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

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

We combine the above observations, and setting

$$(2.6) \quad \mathcal{J}_i = \tilde{J}_i + P^\top b e_1^\top \tilde{J}_i P^\top P - P^\top P \tilde{J}_i^\top e_1 b^\top P$$

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

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

Now if we define

$$(2.8) \quad \mathcal{A}_t^{[2]} f(x) = \int f(\underline{x} - t\Gamma_g(\omega'), \bar{x} - t \sum_{i=1}^m e_{d+i} \underline{x}^\top \mathcal{J}_i \Gamma_g(\omega')) \chi_\circ(\omega') d\omega'$$

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

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

Hence the desired bounds for $\mathcal{A}^{[1]}$ and $\mathcal{A}^{[2]}$ are equivalent. For the case that the matrices $\mathcal{J}_1, \dots, \mathcal{J}_m$ are linearly independent (1.4) 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 (and the analogue of (1.4)) follows by an application of the spherical maximal theorems in the Euclidean case in \mathbb{R}^d ([20]) 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.

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 $\mathcal{A}_t^{[2]}$ as in (2.8)*

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

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

$$(2.11) \quad \mathcal{A}_t^{[3]} f(x) = \int f(\underline{x} - t\Gamma_g(\omega'), \bar{x} - t \sum_{i=1}^n e_{d+i} \underline{x}^\top S_i \Gamma_g(\omega')) \chi_\circ(\omega') d\omega'.$$

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^\top . That is, we decompose $C^\top = UDV$ 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

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

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 $V^\top = V^{-1}$)

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

which gives (2.10). \square

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 maximal estimate for $\mathcal{A}_t^{[3]}$; a similar argument shows that (1.4) can be obtained with $\mathcal{A}^{[3]}$ in place of A . In \mathbb{R}^m we split variables $\bar{x} = (\tilde{x}, \check{x}) \in \mathbb{R}^n \times \mathbb{R}^{m-n}$. For $h \in L^p(\mathbb{R}^{d+n})$ we define $\mathfrak{A}h(x, t) = \mathfrak{A}_t h(x)$ by

$$\mathfrak{A}_t h(\underline{x}, \tilde{x}) = \int h(\underline{x} - t \Gamma_g(\omega'), \tilde{x} - t \sum_{i=1}^n e_{d+i} \underline{x}^\top S_i \Gamma_g(\omega')) \chi_\circ(\omega') d\omega'$$

and we get from Theorem 2.1 applied with n in place of m , that the operator $h \mapsto \sup_t |\mathfrak{A}_t h|$ is bounded on $L^p(\mathbb{R}^{d+n})$ for $p > \frac{d}{d-1}$. Here we need to assume, as we may, that $\rho > 0$ is small enough, i.e. the support of $\omega' \mapsto \chi_\circ(\omega')$ is in a sufficiently small neighborhood of the origin.

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

$$\iint |\sup_{t>0} |\mathcal{A}_t^{[3]} f(\underline{x}, \tilde{x}, \check{x})|^p d\underline{x} d\tilde{x} \leq C^p \iint |f(\underline{x}, \tilde{x}, \check{x})|^p d\underline{x} d\tilde{x},$$

with C independent of \check{x} . Integrating over $\check{x} \in \mathbb{R}^{m-n}$ 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 1.3 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 maximal functions 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 [13] 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 [13] apply to surfaces Σ that are small perturbations of the sphere. Such a perturbation assumption would be needed for the proof in [13] since the Métivier condition on the matrices J_i (and, equivalently, on the $\tilde{J}_i = |y_\circ|^2 R J_i R^\top$) 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 smallness assumption on b needed.

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

$$(2.12) \quad c_0 \leq \left\| \sum_{i=1}^m \theta_i S_i \right\| \leq c_0^{-1} \text{ for all } \theta \in \mathbb{R}^{m-1} \text{ with } 1/4 \leq |\theta| \leq 4$$

This is immediate from the fact that $\theta \rightarrow \|\sum_{i=1}^m \theta_i S_i\|$ is a continuous function which takes a minimum and a maximum on the annulus $\{\theta : 1/4 \leq |\theta| \leq 4\}$, 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

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

with f , where δ is the Dirac measure in \mathbb{R}^{m+1} and β_0 is a C^∞ function supported in a ball of radius $\varrho \ll 1$ centered at the origin of \mathbb{R}^{d-1} and β_1 is a C^∞ function supported in an ε -ball centered at $(1, 0, \dots, 0) \in \mathbb{R}^{1+m}$. 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

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

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

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

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 $\nu_t^{k,l}(x) = t^{-d-2m}\nu^{k,l}(t^{-1}\underline{x}, t^{-2}\bar{x})$ 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 for $1 \leq p \leq 2$.*

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

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

$$(3.4) \quad \left(\int_{\mathbb{R}^{d+m}} \|u(\cdot)Af(x, \lambda \cdot)\|_{B_{p,1}^{1/p}}^p dx \right)^{1/p} \lesssim \|f\|_{L^p(\mathbb{R}^{d+m})}.$$

Proof. The first statement follows from the second, since $B_{p,1}^{1/p}$ embeds into the space of bounded continuous functions. By scaling we can assume that $\lambda = 1$. 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 $\|g\|_{B_{p,1}^\theta} \lesssim \|g\|_p^{1-\theta} \|g'\|_p^\theta$ ($0 < \theta < 1$), Hölder's inequality, Fubini and the proposition to deduce that the left hand side of (3.4) is dominated by

$$\|u Af\|_{L^p(B_{p,1}^\theta)} \lesssim \|\mathcal{R}^{k,l}f\|_{L^p(L^p)}^{1-\theta} \|\partial_t \mathcal{R}^{k,l}f\|_{L^p(L^p)}^\theta \lesssim 2^{-k(\frac{d-1}{p'} - \theta)} 2^{l(\frac{d-2}{p'} - \theta - \varepsilon)} \|f\|_p.$$

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 $p > 2$. \square

Proof of the restricted weak type inequality in Theorem 2.1. We first assume that the matrices J_1, \dots, J_m are linearly independent, in which case we can apply Proposition 3.1 with $S_i = J_i$. 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

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

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

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

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 $\beta(x) = \beta_1(x_1, \bar{x})\beta_0(x')$. By integration by parts with respect to σ, τ we obtain the inequality

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

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

$$(3.6) \quad \|\nu^{k,l}\|_1 + 2^{l-k}\|\partial_s\nu_s^{k,l}\|_1 \lesssim 1$$

and by an application of the fundamental theorem of calculus we get

$$(3.7) \quad \left\| \sup_{1 \leq s \leq 2} |f * \nu_s^{k,l}| \right\|_1 \lesssim 2^{k-l} \|f\|_1.$$

and of course we have $\|\sup_{1 \leq s \leq 2} |f * \nu_s^{k,l}|\|_\infty \lesssim \|f\|_\infty$. For later use we also record

$$(3.8) \quad \|\nabla\nu^{k,l}\|_1 + 2^{l-k}\|\nabla\partial_s\nu_s^{k,l}\|_1 \lesssim 2^k$$

We will show in the next section §4 how to prove the L^p boundedness for the global maximal operator, given the result of Proposition 3.1. The proof of Proposition 3.1 will be given in §5–§6.

4. THE GLOBAL MAXIMAL OPERATOR

The reduction of the bound for the global maximal function to Proposition 3.1 follows closely arguments in [13]; we include a sketch for the convenience of the reader. Recall that $\nu^{k,l}$ is compactly supported and that

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

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

As noted in [13] we can write

$$\nu_{k,l}(x, \bar{x}) = \mathcal{K}^{k,l}(x, \bar{x}) + \gamma_{k,l} u(x, \bar{x})$$

where u is a $C_c^\infty(\mathbb{R}^{d+m})$ function, $|\gamma_{k,l}| \leq c_N 2^{-kN}$ for $l \leq k$ and

$$(4.1) \quad \int \mathcal{K}^{k,l}(x) dx = 0.$$

Clearly if $u_t(\underline{y}, \bar{y}) = t^{-d-2m} u(t^{-1}\underline{y}, t^{-2}\bar{y})$ then the maximal operator $f \mapsto \sup_{t>0} |f * \gamma_{k,l} u_t|$ is bounded on $L^p(\mathbb{R}^{d+m})$, for $1 < p < \infty$, with operator norm $O(2^{-k})$ and so it suffices to prove that

$$(4.2) \quad \left\| \sup_{t>0} |f * \mathcal{K}_t^{k,l}| \right\|_{L^p} \lesssim_\varepsilon (1+k) 2^{k(\frac{1}{p} - \frac{d-1}{p'})} 2^{l(\frac{d-2}{p'} + \varepsilon)} \|f\|_{L^p}$$

for a suitable power c .

We first consider the case $p = 2$, which is a consequence of

$$(4.3) \quad \left(\sum_{n \in \mathbb{Z}} \int_1^2 \|f * \mathcal{K}_{2^n s}^{k,l}\|_{L^2}^2 ds \right)^{1/2} + 2^{l-k} \left(\sum_{n \in \mathbb{Z}} \int_1^2 \|\partial_s(f * \mathcal{K}_{2^n s}^{k,l})\|_{L^2}^2 ds \right)^{1/2} \\ \lesssim_\varepsilon \sqrt{1+k} 2^{-k \frac{d-1}{2} + l(\frac{d-2}{2} + \varepsilon)} \|f\|_{L^2}$$

For each fixed n we have

$$(4.4) \quad \left(\int_1^2 \|f * \mathcal{K}_{2^n s}^{k,l}\|_{L^2}^2 ds \right)^{1/2} + 2^{l-k} \left(\int_1^2 \|\partial_s(f * \mathcal{K}_{2^n s}^{k,l})\|_{L^2}^2 ds \right)^{1/2} \\ \lesssim_\varepsilon 2^{-k \frac{d-1}{2} + l(\frac{d-2}{2} + \varepsilon)} \|f\|_{L^2}$$

This follows from Proposition 3.1 with $I = [1, 2]$ and scaling, applied to the convolution operators with convolution kernel $K_t^{k,l}$ but clearly by the above discussion we can replace $K_t^{k,l}$ with $\mathcal{K}_t^{k,l}$. To finish the proof of (4.3) we need a following variant of the Cotlar-Stein lemma. Let H_1, H_2 be Hilbert spaces and let $T_n : H_1 \rightarrow H_2$, $n \in \mathbb{Z}$ be bounded operators. Assume $B \geq 2A$, and

$$\|T_n\|_{H_1 \rightarrow H_2} \leq A, \quad \|T_n T_{n'}^*\|_{H_2 \rightarrow H_2} \leq B 2^{-\varepsilon|n-n'|}$$

for all $n, n' \in \mathbb{Z}$. Then for all $f \in H^1$

$$(4.5) \quad \left(\sum_{n \in \mathbb{Z}} \|T_n f\|_{H_2}^2 \right)^{1/2} \lesssim A(1 + \varepsilon^{-1} B/A)^{1/2} \|f\|_{H_1}.$$

This is proved for the case $H_1 = H_2$ in [13, 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 $T_n, U_n : H_1 \rightarrow H_2$ given by $T_n f(x, s) = f * \mathcal{K}_{2^n s}^{k,l}$ and $U_n f(x, s) = f * \partial_s \mathcal{K}_{2^n s}^{k,l}$. Clearly we have $\|T_n T_n^*\| \lesssim_\varepsilon A_{k,l}^2$ with $A_{k,l} = 2^{-k(d-1)/2+l(d-2+\varepsilon)/2}$, we use this for $|n - n'| \leq 2k$. As $\nu_s^{k,l}, 2^{-k} \nabla \nu_s^{k,l}, 2^{l-k} \partial_s \nu_s^{k,l}, 2^{l-2k} \nabla \partial_s \nu_s^{k,l}$ are for $s \approx 1$ pointwise dominated by the right hand side of (3.5) the kernels $\mathcal{K}_s^{k,l}, 2^{-k} \nabla \mathcal{K}_s^{k,l}, 2^{l-k} \partial_s \mathcal{K}_s^{k,l}, 2^{l-2k} \nabla \partial_s \mathcal{K}_s^{k,l}$ satisfy up to a constant the same bounds. Since they are also supported on a fixed common compact set we have

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

for $|s| \approx 1$. Using the cancellation property (4.1) a standard calculation gives

$$(4.6) \quad \|f * \tilde{\mathcal{K}}_{2^n s}^{k,l} * \mathcal{K}_{2^{n'} s}^{k,l}\|_2 + 2^{l-2k} \|f * \partial_s \tilde{\mathcal{K}}_{2^n s}^{k,l} * \partial_s \mathcal{K}_{2^{n'} s}^{k,l}\|_2 \lesssim 2^{-|n-n'|} \|f\|_2$$

so that we get $\|T_{n'} T_n^*\| + 2^{2l-4k} \|U_{n'} U_n^*\| \lesssim 2^{-|n-n'|}$. We may thus apply (4.5) with $B/A = O(2^{kC})$. Now (4.3), and in turn (4.2) for $p = 2$ follows.

We interpolate with a weak type (1, 1) inequality exactly as in [13, §6]. This is proved by Calderón-Zygmund estimates with the underlying structure of balls with the parabolic dilation structure. For $r > 0$ let $B_r(0) = \{(x, \bar{x}) : |x| \leq r, |\bar{x}| \leq r^2\}$. We then get the estimate

$$(4.7) \quad \sup_{y \in B_r} \int_{(B_{10r})^c} \sup_{1 \leq t \leq 2} |\mathcal{K}_t^{k,l}(x - y, \bar{x} - \bar{y} - x^\top \vec{J} y) - \mathcal{K}_t^{k,l}(x, \bar{x})| dx d\bar{x} \lesssim \min\{2^{k-l}, 2^{kC} 2^{-n} r, 2^{kC} 2^n r^{-1}\};$$

here the $O(2^{k-l})$ bound is used for $2^{-10Ck} \leq 2^{-n} r \leq 2^{10Ck}$ and is immediate from the above bounds for $\|\partial_s \mathcal{K}_s^{k,l}\|_1$. The bounds $2^{kC} \min\{2^{-n} r, 2^n r^{-1}\}$ follow by standard and straightforward estimates for singular integrals. The above L^2 bounds together with (4.7) (and summation in n) imply

$$(4.8) \quad \text{meas}(\{x \in \mathbb{R}^{d+m} : \sup_{t>0} |f * \mathcal{K}_t^{k,l}(x)| > \alpha\}) \lesssim (1+k) 2^{k-l} \alpha^{-1} \|f\|_1$$

and the L^p bounds for $1 < p < 2$ follows by the Marcinkiewicz interpolation theorem. Moreover the L^p bounds for $p > 2$ can be proved by interpolation with L^∞ bounds for the maximal function.

Remark. A modification of the argument above also yields the slightly stronger bound

$$\left(\int \sup_n \|u(\cdot) A f(x, 2^n \cdot)\|_{B_{p,1}^{1/p}}^p dx \right)^{1/p} \lesssim \|f\|_p, \quad p > \frac{d}{d-1}.$$

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 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 [17]). We follow [17] 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 ε^2 neighborhood of the origin we can rewrite $f * \nu_t^{k,l} = \mathfrak{A}^{k,l} f(x, t)$ where

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

and $K_t^{k,l}$ is given by

$$\begin{aligned} K_t^{k,l}(x, y) &= t^{-d-2m} \nu^{k,l}(t^{-1}(\underline{x} - \underline{y}), t^{-2}(\bar{x} - \bar{y} + \underline{x}^\top \bar{S} \underline{y})) \\ (5.1) \quad &= \mathring{a}(x, t, y) \iint \zeta_{k,l}(\sigma, \tau) e^{i \frac{\sigma}{t}(x_1 - y_1 - t g(\frac{x' - y'}{t})) + i \langle \frac{\tau}{t^2}, \bar{x} - \bar{y} + \underline{x}^\top \bar{S} \underline{y} \rangle} d\sigma d\tau \end{aligned}$$

with

$$\mathring{a}(x, t, y) = (2\pi)^{-m-1} t^{-d-2m} \beta_1\left(\frac{x_1 - y_1}{t}, \frac{\bar{x} - \bar{y} + \underline{x}^\top \bar{S} \underline{y}}{t^2}\right) \beta_0\left(\frac{x' - y'}{t}\right).$$

Notice that \mathring{a} lives, for $|t - 1| \leq \varepsilon^2$ where $|x_1 - 1| \lesssim \varepsilon^2$, $|x'|, |\bar{x}|, |y| \lesssim \varepsilon^2$. Introducing the frequency variables $\vartheta = (\vartheta_1, \bar{\vartheta}) \in \mathbb{R}^{m+1}$, with $\vartheta_1 = 2^{1-k} t^{-1} \sigma$, $\bar{\vartheta}_i = 2^{1-k} t^{-2} \tau_i$ we can rewrite the integral as

$$\begin{aligned} (5.2) \quad K_t^{k,l}(x, y) &= 2^{(k-1)(1+m)} \iint \zeta_1(2^l t \vartheta_1) t^{1+2m} \zeta_1((t^2 \vartheta_1^2 + t^4 |\bar{\vartheta}|^2)^{1/2}) \times \\ &\quad \mathring{a}(x, y, t) e^{i 2^{k-1}(\vartheta_1(x_1 - y_1 - t g(\frac{x' - y'}{t})) + \langle \bar{\vartheta}, \bar{x} - \bar{y} + \underline{x}^\top \bar{S} \underline{y} \rangle)} d\vartheta_1 d\bar{\vartheta} \end{aligned}$$

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

Following [17] we rewrite the phase and verify that

$$\begin{aligned} (5.3) \quad &\vartheta_1(x_1 - y_1 - t g(\frac{x' - y'}{t})) + \sum_{i=1}^m \bar{\vartheta}_i(\bar{x}_i - \bar{y}_i + \underline{x}^\top S_i \underline{y}) \\ &= (\vartheta_1 - \sum_{i=1}^m \bar{\vartheta}_i \underline{x}^\top S_i e_1)(x_1 - y_1 - t g(\frac{x' - y'}{t})) \\ &\quad + \sum_{i=1}^m \bar{\vartheta}_i(\bar{x}_i + x_1 \underline{x}^\top S_i e_1 - \bar{y}_i + \underline{x}^\top S_i P^\top y' - \underline{x}^\top S_i e_1 t g(\frac{x' - y'}{t})) \end{aligned}$$

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

$$(5.4) \quad \Psi(x, t, y, \theta) = \theta_1(x_1 - y_1 - tg(\frac{x' - y'}{t})) + \sum_{i=1}^m \bar{\theta}_i(\bar{x}_i - \bar{y}_i + \underline{x}^\top S_i P^\top y' - \underline{x}^\top S_i e_1 tg(\frac{x' - y'}{t})).$$

Based on the above calculations it is convenient to consider a variant $\mathcal{A}^{k,l}$ which is related to $\mathfrak{A}^{k,l}$ via a shear transformation (see (5.8) below). 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

$$(5.5) \quad \mathcal{K}_t^{k,l}(x, y) = 2^{(k-1)(1+m)} \iint e^{i2^{k-1}\Psi(x,t,y,\theta)} \mathring{a}(x, t, y) t^{1+2m} \times \\ \zeta_1(2^l t(\theta_1 + \underline{x}^\top S^{\bar{\theta}} e_1)) \zeta_1((t^2(\theta_1 + \underline{x}^\top S^{\bar{\theta}} e_1)^2 + t^4 |\bar{\theta}|^2)^{1/2}) d\theta_1 d\bar{\theta},$$

with the modification that for $k = l$ we replace $\zeta_1(2^l t(\theta_1 + \underline{x}^\top S^{\bar{\theta}} e_1))$ with $\zeta_0(2^l t(\theta_1 + \underline{x}^\top S^{\bar{\theta}} e_1))$. Here we used the notation $S^{\bar{\theta}} = \sum_{i=1}^m \bar{\theta}_i S_i$.

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*

$$(5.6) \quad \|\mathcal{A}^{k,l} f\|_{L^2(\mathbb{R}^{d+m} \times [\frac{1}{2}, 2])} \leq C_\varepsilon 2^{-\frac{k(d-1)}{2}} 2^{l(\frac{d-2}{2} + \varepsilon)} \|f\|_{L^2(\mathbb{R}^{d+m})},$$

with C_ε bounded as $\zeta, \zeta^\circ, \mathring{a}$ are varying over bounded subsets of C_c^∞ (with the above support assumptions).

Proof that Proposition 5.1 implies Proposition 3.1. Using (5.3) in (5.2) we see that (for $l < k$)

$$(5.7) \quad K_t^{k,l}(x, y) = 2^{(k-1)(1+m)} \iint e^{i2^{k-1}\tilde{\Psi}(x,t,y,\theta)} \mathring{a}(x, t, y) \times \\ \zeta_1(2^l t(\theta_1 + \underline{x}^\top S^\theta e_1)) \zeta_1((t^2(\theta_1 + \underline{x}^\top S^\theta e_1)^2 + t^4 |\bar{\theta}|^2)^{1/2}) d\theta_1 d\bar{\theta}$$

with $\tilde{\Psi}(x, t, y, \theta) = \Psi(\underline{x}, \bar{x} + x_1 \underline{x}^\top \vec{S} e_1, t, y, \theta)$. When $k = l$ replace $\zeta_1(2^l t(\theta_1 + \underline{x}^\top S^\theta e_1))$ by $\zeta_0(2^l t(\theta_1 + \underline{x}^\top S^\theta e_1))$. With this choice of $\tilde{\Psi}$ we have the identity $K_t^{k,l}(x, y) = \mathcal{K}_t^{k,l}(\underline{x}, \bar{x} + \underline{x}^\top \vec{S} e_1, t, y)$ and thus

$$(5.8) \quad \mathfrak{A}^{k,l} f(\underline{x}, \bar{x}, t) = \mathcal{A}_t^{k,l} f(\underline{x}, \bar{x} + \underline{x}^\top \vec{S} e_1, t).$$

A similar observation holds for $k = l$. Hence 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

$$\partial_t \tilde{\Psi}(x, t, y, \theta) = (\theta_1 + \underline{x}^\top S^{\bar{\theta}} e_1) \left(\langle \frac{x' - y'}{t}, \nabla g(\frac{x' - y'}{t}) \rangle - g(\frac{x' - y'}{t}) \right).$$

From (5.7) we calculate that the Schwartz kernel of $\partial_t \mathfrak{A}_t^{k,l}$ is given by

$$(5.9) \quad \partial_t K_t^{k,l} = K_{t,1}^{k,l} + K_{t,2}^{k,l} + 2^{k-l} K_{t,3}^{k,l}$$

where the kernels on the right hand side are given as in (5.7) but with ζ, \mathring{a}_0 replaced by $\zeta^{[i]}, \mathring{a}^{[i]}$ for $i = 1, 2, 3$, resp., with the specific definitions (for $l < k$)

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

and

$$\begin{aligned} \mathring{a}^{[1]}(x, t, y) &= t^{-1}\mathring{a}(x, t, y) \\ \mathring{a}^{[2]}(x, t, y) &= \partial_t \mathring{a}(x, t, y) \\ \mathring{a}^{[3]}(x, t, y) &= t^{-1}\mathring{a}(x, t, y) \left(\left\langle \frac{x'-y'}{t}, \nabla g\left(\frac{x'-y'}{t}\right) \right\rangle - g\left(\frac{x'-y'}{t}\right) \right) \end{aligned}$$

For $k = l$ replace ζ by ζ_0 . These formulas show that the derivative bound in (3.3) follows from Proposition 5.1 as well. \square

5.2. A family of oscillatory integral operators. It remains to prove Proposition 5.1 for $p = 2$. 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

$$(5.10) \quad c_0 \leq \|S\| \leq c_0^{-1}$$

with $c_0 > 0$; in particular the rank of S is at least 2.

We define the phase function ψ by

$$(5.11) \quad \psi(x, t, y) = y_1(x_1 - tg(\frac{x'-y'}{t})) + x^\top S(P^\top y' - tg(\frac{x'-y'}{t})e_1)$$

and set

$$(5.12) \quad \sigma(x', y_1) = y_1 + (x')^\top P S e_1.$$

The function ζ_1 can be split as $\zeta_1 = \zeta_1^+ + \zeta_1^-$ where $\text{supp}(\zeta_1^+) \subset (\frac{1}{2}, 2)$ and $\text{supp}(\zeta_1^-) \subset (-2, -\frac{1}{2})$.

Setting $\lambda = 2^{k-1}$ and letting $l \leq k$ we define, for functions $f \in L^2(\mathbb{R}^d)$,

$$(5.13) \quad T^{\lambda,l} f(x, t) = \int e^{i\lambda\psi(x,t,y)} \chi_l(x, t, y) f(y) dy$$

where

$$(5.14) \quad \chi_l(x, t, y) = \begin{cases} \chi(x, t, y) \zeta(2^l t \sigma(x', y_1)), & l \leq k-1 \\ \chi(x, t, y) \zeta_0(2^l t \sigma(x', y_1)), & l = k. \end{cases}$$

Here χ is C_c^∞ -function supported where $t \approx 1$, $|x'|, |y'| \leq \varrho$, and $\text{diam}(\text{supp}(\chi)) \leq \varrho$. For $l \leq k-1$ we use the convention for ζ to be either ζ_1^+ or ζ_1^- . Note then that for $l \leq k-1$ we have $|\sigma| \approx 2^{-l}$ on $\text{supp}(\chi_l)$ and in addition the sign of σ is the same for all (x, t, y) in the support.

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

$$(5.15) \quad \|T^{\lambda,l} f\|_{L^2(\mathbb{R}^d \times [1/2,2])} \lesssim C_\varepsilon 2^{l(\frac{d-2}{2}+\varepsilon)} \lambda^{-\frac{d}{2}} \|f\|_{L^2(\mathbb{R}^d)}.$$

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

For $l \gg 1$ this is the main 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 \hat{a} can be written as a tensor product of functions of each of the variables x_i, t, y_k . To reduce to this situation we choose functions $x_i \mapsto \alpha_i(x_i)$, $t \mapsto \gamma(t)$, $y_i \mapsto \beta_i(y_i)$, $i = 1, \dots, d$, all with compact support such that

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

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

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

$$\hat{a}(x, t, y) t^{1+2m} = \sum_{(n, \nu, \mu) \in \mathbb{Z} \times \mathbb{Z}^{d+m} \times \mathbb{Z}^{d+m}} c_{n, \nu, \mu} e^{itn} \prod_{i=1}^{d+m} e^{ix_i \nu_i} \prod_{j=1}^{d+m} e^{iy_j \mu_j}$$

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

$$(5.16) \quad \mathcal{A}^{k,l} f(x, t) = \sum_{n, \nu, \mu} c_{n, \nu, \mu} e^{itn} \prod_{i=1}^{d+m} e^{ix_i \nu_i} \mathcal{A}_\mu^{k,l} f(x, t)$$

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

$$(5.17) \quad \mathcal{A}_\mu^{k,l} f(x, t) = 2^{(k-1)(m+1)} \mathcal{F}_k^m \mathcal{G}^{k,l} \mathcal{F}_{k,\mu}^{m+1} f(x, t);$$

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

$$(5.18) \quad \mathcal{F}_k^m G(x, \bar{x}, t) = \prod_{i=d+1}^{d+m} \alpha_i(x_i) \int_{\mathbb{R}^m} G(x, \bar{\theta}, t) e^{i2^{k-1} \langle \bar{x}, \bar{\theta} \rangle} d\bar{\theta},$$

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

$$(5.19) \quad \begin{aligned} \mathcal{G}^{k,l} F(x, \bar{\theta}, t) &= \gamma(t) \prod_{i=1}^d \alpha_i(x_i) \int_{\theta_1, y', \bar{\theta}} e^{i2^{k-1} \psi^{\bar{\theta}}(x_1, x', t, \theta_1, y')} \zeta_1(2^l t (\theta_1 + x^\top S^{\bar{\theta}} e_1)) \\ &\quad \times \zeta_1((t^2(\theta_1 + x^\top S^{\bar{\theta}} e_1)^2 + t^4 |\bar{\theta}|^2)^{1/2}) \prod_{j=2}^d \beta_j(y_j) F(\theta_1, y', \bar{\theta}) d\theta_1 dy' d\bar{\theta} \end{aligned}$$

with

$$(5.20) \quad \psi^{\bar{\theta}}(x_1, x', t, \theta_1, y') = \theta_1(x_1 - tg(\frac{x' - y'}{t})) + \underline{x}^\top S^{\bar{\theta}}(P^\top y' - tg(\frac{x' - y'}{t})e_1),$$

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

$$(5.21) \quad \mathcal{F}_{k,\mu}^{m+1} f(\theta_1, y', \bar{\theta}) = \int e^{i2^{k-1}(y_1\theta_1 + \langle \bar{y}, \bar{\theta} \rangle)} e^{i\langle \mu, y \rangle} \beta_1(y_1) \prod_{j=d+1}^{d+m} \beta_j(y_j) f(y_1, y', \bar{y}) dy_1 d\bar{y}.$$

We have the estimates

$$(5.22) \quad \|\mathcal{F}_k^m G\|_{L^2(\mathbb{R}^{d+m+1})} \lesssim 2^{-km/2} \|G\|_{L^2(\mathbb{R}^{d+m+1})}$$

$$(5.23) \quad \|\mathcal{G}^{k,l} F\|_{L^2(\mathbb{R}^{d+m+1})} \leq C_\varepsilon 2^{l(\frac{d-2}{2} + \varepsilon)} 2^{-kd/2} \|F\|_{L^2(\mathbb{R}^{d+m})}$$

$$(5.24) \quad \|\mathcal{F}_{k,\mu}^{m+1} f\|_{L^2(\mathbb{R}^{d+m})} \lesssim 2^{-k(m+1)/2} \|f\|_{L^2(\mathbb{R}^{d+m})}$$

and clearly the desired estimate (5.6) follows from (5.22), (5.23), (5.24) in conjunction with (5.16), (5.17) and the rapid decay of the $c_{n,\mu,\nu}$.

We justify the L^2 estimates. (5.22) is an immediate consequence of Plancherel's theorem in \mathbb{R}^m and likewise (5.24) is a consequence of Plancherel's theorem in \mathbb{R}^{m+1} . It remains to consider (5.23); here we rely on Proposition 5.2. With $\psi^{\bar{\theta}}$ as in (5.20) define for functions $(\theta_1, y') \mapsto g(\theta_1, y')$

$$(5.25) \quad \mathcal{T}_{\bar{\theta}}^{\lambda,l} g(\underline{x}, t) = \int_{\theta_1, y'} \exp(i\lambda \psi^{\bar{\theta}}(\underline{x}, t, \theta_1, y')) \chi^{\bar{\theta}}(\underline{x}, t, \theta_1, y') \zeta_1(2^l t \sigma^{\bar{\theta}}(x', \theta_1)) g(\theta_1, y') d\theta_1 dy'$$

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

$$\chi^{\bar{\theta}}(\underline{x}, t, \theta_1, y') = \gamma(t) \prod_{i=1}^d \alpha_i(x_i) \prod_{j=2}^d \beta_j(y_j) \zeta_1((t^2(\theta_1 + \underline{x}^\top S^{\bar{\theta}} e_1)^2 + t^4 |\bar{\theta}|^2)^{1/2}).$$

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

$$(5.26) \quad \|\mathcal{T}_{\bar{\theta}}^{\lambda,l} g\|_{L^2(\mathbb{R}^{d+1})} \lesssim 2^{l(\frac{d-2}{2} + \varepsilon)} 2^{-kd/2} \|g\|_{L^2(\mathbb{R}^d)}$$

uniformly in $\bar{\theta}$; note that we have exactly the setup in (5.13), except there we use the notation x for \underline{x} , y_1 for θ_1 , and S for $S^{\bar{\theta}}$. For the estimate (5.26) the uniformity assertion in Proposition 5.2 is crucial and so is the assumption of the S_i being linearly independent, resulting in the uniform bound (2.12). We have $\mathcal{G}^{k,l} F(x, \bar{\theta}, t) = \mathcal{T}_{\bar{\theta}}^{2^{k-1},l} [F(\cdot, \bar{\theta})]$ and thus applying (5.26) gives (5.23). 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 [9]. 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 $T_\nu : H_1 \rightarrow H_2$ where H_1, H_2 are Hilbert spaces. Then we have the following bound for the operator norm of the sum:

$$(6.1) \quad \left\| \sum_{\nu \in \mathcal{V}} T_\nu \right\|_{H_1 \rightarrow H_2} \lesssim \sup_{\nu} \sum_{\nu'} \|T_\nu^* T_{\nu'}\|_{H_1 \rightarrow H_1}^{1/2} + \sup_{\nu} \sum_{\nu'} \|T_\nu T_{\nu'}^*\|_{H_2 \rightarrow H_2}^{1/2}.$$

This well known version follows by a simple modification of the proof in [21, 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. [9]). Recall that $g(0) = 1$, $\nabla g(0) = 0$, $\text{diam}(\text{supp}(\chi)) \leq \varrho \ll 1$, in particular $|x'|, |y'| \leq \varrho \ll 1$ for $(x, t, y) \in \text{supp}(\chi)$. Denote as before

$$(6.2) \quad \sigma = \sigma(x', y_1) = y_1 + (x')^\top P S e_1.$$

Let the $(d+1) \times d$ matrix $\partial_y^\top \partial_{x,t} \psi$ be defined by $(\partial_y^\top \partial_{x,t} \psi)_{i,j} = \partial_{x_i} \partial_{y_j} \psi$ for $1 \leq i, j \leq d$ and $(\partial_y^\top \partial_{x,t} \psi)_{d+1,j} = \partial_t \partial_{y_j} \psi$ for $1 \leq j \leq d$. One calculates ([17])

$$(6.3) \quad \partial_y^\top \partial_{(x,t)} \psi \Big|_{(x,t,y)} = \begin{pmatrix} 1 & e_1^\top S P^\top \\ -g'(\frac{x'-y'}{t}) & t^{-1} \sigma(x', y_1) g''(\frac{x'-y'}{t}) + P S P^\top + P S e_1 (g'(\frac{x'-y'}{t}))^\top \\ -1 + \tilde{g}(x, t, y) & -t^{-2} \sigma(x', y_1) (x' - y')^\top g''(\frac{x'-y'}{t}) \end{pmatrix}$$

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

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

$$\det(\partial_y^\top \partial_x \psi(x, t, y)) = \det(t^{-1} \sigma(x', y_1) g''(\frac{x'-y'}{t}) + P S P^\top) + O(\varrho).$$

From [13, Lemma 5.3], it follows that the matrix $t^{-1} \sigma g''(\frac{x'-y'}{t}) + P S P^\top$ is invertible. This says that $\partial_y^\top \partial_x \psi(x, t, y)$ is invertible for all $(x, t, y) \in \text{supp}(\chi)$. 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 [9] applies and we may conclude the bound $\|T^{\lambda,l} f(\cdot, t)\| \leq C(l) \lambda^{-d/2} \|f\|_2$ 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 $\mathbf{e}_1, \dots, \mathbf{e}_d$ with $\mathbf{e}_1 = e_1$, and $S e_1 \in \text{span}(\mathbf{e}_2)$. Set

$$(6.4) \quad \delta_l = \max\{|S e_1|, 2^{-l}\}$$

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 $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \dots, \mathbf{e}_d$. The family of boxes \mathfrak{s} can be parametrized by \mathbb{Z}^d ; we define the lower corners by $c(\mathfrak{z}) = 2^{l\varepsilon/d}(\lambda^{-1}\mathfrak{z}_1\mathbf{e}_1 + \lambda^{-1}\delta_l^{-1}\mathfrak{z}_2\mathbf{e}_2 + \sum_{i=3}^d 2^l\lambda^{-1}\mathfrak{z}_i\mathbf{e}_i)$, and let

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

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

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

Note that

$$(6.6) \quad T_{\mathfrak{s}}^{\lambda, l}(T_{\mathfrak{s}'}^{\lambda, l})^* = 0 \text{ if } \mathfrak{s} \neq \mathfrak{s}'$$

Notice that we have $|T_{\mathfrak{s}}^{\lambda, l}f(x, t)| \lesssim |\mathfrak{s}|^{1/2}\|f\|_2$. Because of the compact support of the kernel in the (x, t) variable we see that the L^2 operator norm $\|T_{\mathfrak{s}}^{\lambda, l}\|_{2 \rightarrow 2}$ 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*

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

holds for every $\mathfrak{s} \in \mathfrak{S}$, with C independent of $t \in [1/4, 4]$ and \mathfrak{s} .

Proof of Lemma 6.1. We have $|\mathfrak{s}| \lesssim 2^{l(d-2+\varepsilon)}\delta_l^{-1}\lambda^{-d}$ and therefore obtain

$$\|T^{\lambda, l}f(\cdot, t)\|_{L^2 \rightarrow L^2} \lesssim 2^{l(d-2+\varepsilon)/2}\delta_l^{-1/2}\lambda^{-d/2}.$$

Let $c_1 \ll c_0$ be a small constant, and the displayed estimate is already sufficient if $|Se_1| \geq c_1$. In what follows we consider the case $|Se_1| \leq c_1$. We freeze t for this proof and write $\mathcal{T}_{\mathfrak{s}}^{\lambda, l}f(x) = T_{\mathfrak{s}}^{\lambda, l}f(x, t)$, all estimates will be uniform in $t \in [1/4, 4]$.

As $Sy = (e_1^{\top}SP^{\top}y', y_1PSe_1 + PSP^{\top}y')$ and $|e_1^{\top}SP^{\top}y'| + |y_1PSe_1| \lesssim c_1$, moreover $c_1 \ll c_0$ we see using (5.10) that

$$(6.7) \quad \sup_{y' \in \mathbb{R}^{d-1}, 1/4 \leq |y'| \leq 4} |PSP^{\top}y'| \geq c_0/2.$$

Let d_0 be the smallest integer greater than or equal to $(d-1)/2$. Since PSP^{\top} is skew-symmetric, there exists nonnegative numbers $s_1 \geq \dots \geq s_{d_0}$ and orthonormal vectors $\vec{u}_1, \dots, \vec{u}_{d-1} \in \mathbb{R}^{d-1}$ such that

$$(6.8a) \quad \begin{aligned} PSP^{\top}\vec{u}_{2i-1} &= s_i\vec{u}_{2i}, \\ PSP^{\top}\vec{u}_{2i} &= -s_i\vec{u}_{2i-1}, \end{aligned}$$

for $1 \leq i \leq d_o$ if $d - 1$ is even, and

$$(6.8b) \quad \begin{aligned} PSP^\top \vec{u}_{2i-1} &= s_i \vec{u}_{2i}, & PSP^\top \vec{u}_{2d_o-1} &= 0 \\ PSP^\top \vec{u}_{2i} &= -s_i \vec{u}_{2i-1}, \end{aligned}$$

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

To estimate $\mathcal{T}_s^{\lambda,l} f$, we further decompose the slab \mathfrak{s} into smaller pieces. Since PSe_1 is orthogonal to e_1 we can decompose $PSe_1 = \sum_{i=1}^{d-1} \alpha_i \vec{u}_i$ 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 $\text{span}(\vec{u}_1, \vec{u}_2)$ with the property that $PSP^\top b = -\beta_2 s_1 \vec{u}_1 + \beta_2 s_1 \vec{u}_2$ is perpendicular to PSe_1 . For later use notice that $|PSP^\top b| = s_1$.

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

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

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

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

and by the Cauchy-Schwarz inequality we get

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

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

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

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 $|n - n'| \leq C 2^l$. The Schwartz kernel of $(\mathcal{T}_{s,n}^{\lambda,l})^* \mathcal{T}_{s,n'}^{\lambda,l}$ is given by

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

where

$$(6.13) \quad \phi(x, t, y, z) = -\psi(x, t, y) + \psi(x, t, z).$$

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

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

Note that

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

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

Using (6.3), we write

$$(6.17) \quad \partial_y^\top \partial_{x'} \psi|_{(x,t,w^\tau)}(y-z) = \langle y' - z', b \rangle PSP^\top b + PSP^\top \Pi_{b^\perp}(y' - z') + t^{-1} \sigma(x', w_1^\tau) g''\left(\frac{x' - w^{\tau'}}{t}\right)(y' - z') - g'\left(\frac{x' - w^{\tau'}}{t}\right)(y_1 - z_1) - PSe_1(g'\left(\frac{x' - w^{\tau'}}{t}\right))^\top (y' - z').$$

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

$$(6.18) \quad \partial_y^\top \langle v, \partial_{x'} \rangle \psi|_{(x,t,w^\tau)}(y-z) = s_1 \langle y' - z', b \rangle + (ts_1)^{-1} \sigma(x', w_1^\tau) (PSP^\top b)^\top g''\left(\frac{x' - w^{\tau'}}{t}\right)(y' - z') - s_1^{-1} (y_1 - z_1) (PSP^\top b)^\top g'\left(\frac{x' - w^{\tau'}}{t}\right).$$

Notice from (6.2) that

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

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 $|\langle y' - z', b \rangle| \sim |n - n'| \lambda^{-1} 2^{l\varepsilon/d}$ 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

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

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.

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

Setting

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

we have

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

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 $(x, t, y, z) \in \text{supp}(\eta_l)$, $y, z \in \mathfrak{s}$

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

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. \square

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 $(T_{\mathfrak{s}}^{\lambda,l})^* T_{\mathfrak{s}'}^{\lambda,l}$ 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 $c_{\mathfrak{s}}^i = \langle c_{\mathfrak{s}}, \mathfrak{e}_i \rangle$, $i = 1, 2$, $c_{\mathfrak{s}}^\perp = \Pi_{\text{span}(\mathfrak{e}_1, \mathfrak{e}_2)^\perp} = \sum_{k=3}^d c_{\mathfrak{s}}^k \mathfrak{e}_k$.

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

$$(6.23a) \quad 2 \lfloor 2^{\kappa_1-1} \rfloor \lambda^{-1} \leq 2^{-l\varepsilon/d} |c_{\mathfrak{s}}^1 - c_{\mathfrak{s}'}^1| \leq 2^{\kappa_1+1} \lambda^{-1},$$

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

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

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

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

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 $N > 0$*

$$(6.24) \quad \|(T_{\mathfrak{s}}^{\lambda,l})^* T_{\mathfrak{s}'}^{\lambda,l}\|_{L^2 \rightarrow L^2} \lesssim_N 2^{-(l\varepsilon/d+\kappa_1)N} 2^{l(d-2+\varepsilon)} \delta_l^{-1} \lambda^{-d}.$$

(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 $N > 0$*

$$(6.25) \quad \|(T_{\mathfrak{s}}^{\lambda,l})^* T_{\mathfrak{s}'}^{\lambda,l}\|_{L^2 \rightarrow L^2} \lesssim_N 2^{-(l\varepsilon/d+\kappa_1+\kappa_2)N} 2^{l(d-2+\varepsilon)} \delta_l^{-1} \lambda^{-d}.$$

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

$$(6.26) \quad \|(T_{\mathfrak{s}}^{\lambda,l})^* T_{\mathfrak{s}'}^{\lambda,l}\|_{L^2 \rightarrow L^2} \lesssim_N 2^{-(l\varepsilon/d+\kappa_1+\kappa_2+\kappa_3)N} 2^{l(d-2+\varepsilon)} \delta_l^{-1} \lambda^{-d}.$$

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 $\mathcal{K}_{\mathfrak{s}_1, \mathfrak{s}_2} \equiv \mathcal{K}_{\mathfrak{s}_1, \mathfrak{s}_2}^{\lambda,l}$ of $(T_{\mathfrak{s}}^{\lambda,l})^* T_{\mathfrak{s}'}^{\lambda,l}$ given by

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

with η_l as in (6.21). Note that whenever $l \leq k-1$ the definition (6.21) of η_l via (5.14) 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 $\kappa_1 \geq 5$, $\kappa_2, \kappa_3 \leq 10$, integration by parts with respect to x_1 , when $|Se_1| \geq 2^{-l}$ and $\kappa_2 \geq 5$, $\kappa_3 \leq 10$, and integration by parts using the directional derivative $\langle \frac{y' - z'}{|y' - z'|}, \partial_{x'} \rangle$, either when $\kappa_3 \geq 5$ or when $\kappa_2 \geq 5$, $\kappa_3 \leq 10$, $|Se_1| \leq 2^{-l}$ (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}

$$(6.28) \quad \sum_{\mathfrak{s}'} \|(T_{\mathfrak{s}}^{\lambda, l})^* T_{\mathfrak{s}'}^{\lambda, l}\|^{1/2} \lesssim 2^{l(\frac{d-2+\varepsilon}{2})} \lambda^{-d/2}$$

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

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

From Lemma 6.1 we have

$$\|(T_{\mathfrak{s}}^{\lambda, l})^* T_{\mathfrak{s}'}^{\lambda, l}\| \lesssim \|T_{\mathfrak{s}}^{\lambda, l}\| \|T_{\mathfrak{s}'}^{\lambda, l}\| \lesssim 2^{l(d-2+\varepsilon)} \lambda^{-d}$$

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

$$(6.30) \quad \sup_{\mathfrak{s}} \sum_{\kappa_1, \kappa_2, \kappa_3 \leq 10} \sum_{\mathfrak{s}' \in \mathcal{U}_{\kappa_1, \kappa_2, \kappa_3}^{\mathfrak{s}}} \|(T_{\mathfrak{s}}^{\lambda, l})^* T_{\mathfrak{s}'}^{\lambda, l}\|^{1/2} \lesssim 2^{l(d-2+\varepsilon)/2} \lambda^{-d/2}.$$

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

$$\begin{aligned} \sup_{\mathfrak{s}} \sum_{\max\{\kappa_1, \kappa_2, \kappa_3\} \geq 5} \sum_{\mathfrak{s}' \in \mathcal{U}_{\kappa_1, \kappa_2, \kappa_3}^{\mathfrak{s}}} \|(T_{\mathfrak{s}}^{\lambda, l})^* T_{\mathfrak{s}'}^{\lambda, l}\|^{1/2} \\ \lesssim 2^{l\frac{d-2+\varepsilon}{2}} \lambda^{-\frac{d}{2}} 2^{l(\frac{1}{2} - \frac{\varepsilon N}{2d})} \sum_{(\kappa_1, \kappa_2, \kappa_3) \in \mathbb{N}_0^3} 2^{-(\kappa_1 + \kappa_2 + \kappa_3)(\frac{N}{2} - d)}. \end{aligned}$$

For $N > 2d$ we can sum in $\kappa_1, \kappa_2, \kappa_3$, and 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. \square

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

Proof of (6.24). We are now in the case $\kappa_3 \geq 5$, and $\kappa_1, \kappa_2 \leq 10$ 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 $|y_1 - z_1| \approx 2^{\kappa_1} \lambda^{-1} 2^{l\varepsilon/d}$, $|\langle y - z, \mathfrak{e}_2 \rangle| \lesssim \delta_l^{-1} \lambda^{-1} 2^{l\varepsilon/d}$, $|\langle y - z, \mathfrak{e}_i \rangle| \lesssim \lambda^{-1} 2^{l(1+\varepsilon/d)}$, $i = 3, \dots, d$.

We now integrate by parts with respect to t ; for this observe that

$$(6.31) \quad \begin{aligned} \partial_t \phi(x, t, y, z) &= \partial_t \psi(x, t, y) - \partial_t \psi(x, t, z) = -(y_1 - z_1) \\ &+ \int_0^\tau \left[\tilde{g}(x, t, w^\tau)(y_1 - z_1) - t^{-2} \sigma(x', w_1^\tau)(x' - w^\tau)^\top g''\left(\frac{x' - w^\tau}{t}\right)(y' - z') \right]. \end{aligned}$$

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

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

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

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

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

$$\begin{aligned} \partial_{x_1} \phi(x, t, y, z) &= \partial_{x_1} \psi(x, t, y) - \partial_{x_1} \psi(x, t, z) \\ &= y_1 - z_1 + e_1^\top S P^\top (y' - z') = y_1 - z_1 - |Se_1| \langle y' - z', \mathfrak{e}_2 \rangle. \end{aligned}$$

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

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

hence

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

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. \square

Proof of (6.25) in the case $|Se_1| \leq 2^{-l}$ and Proof of (6.26). Notice that in view of the small support of χ we have in the present case $\mathcal{K}_{\mathfrak{s}, \mathfrak{s}'}^{\lambda, l} = 0$ when $2^l \lambda^{-1} \geq 1$, so the case $l = k$ is trivial. In what follows we assume $l \leq k - 1$; 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 $|Se_1| \leq 2^{-l}$ we have $\delta_l = 2^{-l}$ and for the proof of (6.25) we have also $\kappa_3 \leq 10$ and we shall prove the pointwise estimate

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

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

$$(6.35) \quad \begin{aligned} 2^{\kappa_1 + \kappa_2 - 1} \lambda^{-1} 2^l &\leq 2^{-l\varepsilon/d} |\langle y_2 - z_2, \mathfrak{e}_2 \rangle| \leq 2^{\kappa_1 + \kappa_2 + 2} \lambda^{-1} 2^l, \\ 2^{-l\varepsilon/d} |(y - z)^\perp| &\lesssim 2^{\kappa_1 + \kappa_2 + 11} \lambda^{-1} 2^l, \quad |Se_1| \leq 2^{-l}. \end{aligned}$$

here $(y - z)^\perp := (y_3 - z_3, \dots, y_d - z_d)$.

Moreover for (6.26) we shall prove

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

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

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

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))

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

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

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

where

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

and

$$\begin{aligned} R_2(x, t, y, z) &= \int_0^1 \left[-\langle \frac{y' - z'}{|y' - z'|}, g'(\frac{x' - w^{\tau'}}{t}) \rangle (y_1 - z_1) + \langle \frac{y' - z'}{|y' - z'|}, PSe_1 \rangle (g'(\frac{x' - w^{\tau'}}{t})^\top(y' - z')) \right] d\tau \\ &= -(y_1 - z_1 - |Se_1| \langle y' - z', \mathfrak{e}_2 \rangle) \int_0^1 (g'(\frac{x' - w^{\tau'}}{t})^\top (\frac{y' - z'}{|y' - z'|})) d\tau. \end{aligned}$$

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

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

and we use

$$\begin{aligned} 2^{-l} |y' - z'| &\approx 2^{-l} |\langle y - z, \mathbf{e}_2 \rangle| \approx 2^{l\varepsilon/d} 2^{\kappa_1 + \kappa_2} && \text{if (6.35) holds,} \\ 2^{-l} |y' - z'| &\approx 2^{-l} |(y - z)^\perp| \approx 2^{l\varepsilon/d} 2^{\kappa_1 + \kappa_2 + \kappa_3} && \text{if (6.37) holds.} \end{aligned}$$

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

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

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

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

Moreover we get this when (6.37) holds and $|Se_1| \leq 2^{-l}$. If (6.37) holds and $|Se_1| \geq 2^{-l}$ then $|Se_1| |\langle y' - z', \mathbf{e}_2 \rangle| \lesssim 2^{\kappa_1 + \kappa_2 + 2} \lambda^{-1} 2^{l\varepsilon/d}$ 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 $y \in \mathfrak{s}$, $z \in \mathfrak{s}'$,

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

and

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

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

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

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 $PSe_1 = |Se_1| \mathbf{e}_2$ and if we now assume (6.37) we have $|\langle y' - z', PSe_1 \rangle| \leq \delta_l |\langle y - z, \mathbf{e}_2 \rangle| \leq 2^{\kappa_1 + \kappa_2 + 2} \lambda^{-1} 2^{l\varepsilon/d}$ and $|y' - z'| \geq |(y - z)^\perp| \geq 2^{\kappa_1 + \kappa_2 + \kappa_3 - 1} \lambda^{-1} 2^{l\varepsilon/d}$; 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 $N \geq 2$. This implies, for all N ,

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

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

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

An integration by parts yields

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

where

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

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. \square

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 open even in the model case of the Heisenberg group \mathbb{H}^1 .

7.2. *A restricted weak type endpoint bound.* Theorem 1.3 establishes a restricted weak type $(\frac{d}{d-1}, \frac{d}{d-1})$ endpoint estimate for the local maximal operator, when $d \geq 3$. 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 $f \mapsto \sup_{1 \leq t \leq 2} |f * \mu_t|$ maps L^p to L^q for some $q > p$; this would imply corresponding problem 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 $m \geq 2$ the dependence on the matrices J_1, \dots, J_m could be quite complicated. The case of Heisenberg type groups is covered in [17].

7.4. *Restricted dilation sets.* One can also consider maximal functions with restricted dilation sets. The $L^p \rightarrow L^p$ estimates with Minkowski dimension type assumptions are rather straightforward; one can combine the methods of this paper with elementary arguments in [19, 18]. In contrast the L^p -improving estimates are harder; for the Heisenberg groups \mathbb{H}^n , with $n \geq 2$,

this problem was considered in [18]. For general dilation sets there is a large variety of possible type sets (cf. [16, 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 special case simplifies the analysis in the present paper but it has been relaxed in [1, 17] 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. ISBN 978-3527401208. MR 1608283
8. Aswin Govindan Sheri, Jonathan Hickman and James Wright. *Lacunary maximal functions on homogeneous groups*, arXiv:2305.04700
9. Lars Hörmander, *Oscillatory integrals and multipliers on FL^p* , Ark. Mat. 11 (1973), 1–11. MR 340924
10. Joonil Kim, *Annulus maximal averages on variable hyperplanes*, arXiv:1906.03797, 2019.
11. Juyoung Lee and Sanghyuk Lee, $L^p \rightarrow L^q$ estimates for the circular maximal operator on Heisenberg radial functions. Math. Ann. 385 (2023), no. 3-4, 1521–1544. MR 4566682
12. 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
13. 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
14. 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

15. 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
16. Joris Roos and Andreas Seeger, *Spherical maximal functions and fractal dimensions of dilation sets*, Amer. J. Math. 145, no. 4 (2023).
17. 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
18. ———, *Spherical maximal operators on Heisenberg groups: Restricted dilation sets*. Studia Math., to appear. arXiv 2208.02774.
19. 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
20. Elias M. Stein, *Maximal functions. I. Spherical means*, Proc. Nat. Acad. Sci. U.S.A. 73 (1976), no. 7, 2174–2175. MR 420116
21. ———, *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`