

L^p IMPROVING BOUNDS FOR AVERAGES ALONG CURVES

TERENCE TAO AND JAMES WRIGHT

1. INTRODUCTION

Let $n \geq 2$, and let M_1 and M_2 be two smooth $n - 1$ -dimensional manifolds¹, each containing a preferred origin 0_{M_1} and 0_{M_2} . We shall abuse notation and write 0 for both 0_{M_1} and 0_{M_2} . For the purposes of integration we shall place a smooth Riemannian metric on M_1 and M_2 , although the exact choice of this metric will not be relevant. All our considerations shall be local to the origin 0 .

We are interested in the local L^p improving properties of averaging operators on curves. Before we give the rigorous description of these operators, let us first give an informal discussion.

Informally, we assume that we have a smooth assignment $x_2 \mapsto \gamma_{x_2}$ taking points in M_2 to curves in M_1 , with a corresponding dual assignment $x_1 \mapsto \gamma_{x_1}^*$ taking points in M_1 to curves in M_2 , such that

$$x_1 \in \gamma_{x_2} \iff x_2 \in \gamma_{x_1}^*.$$

We then form the operator R taking functions² on M_1 to functions on M_2 , defined by

$$(1) \quad Rf(x_2) := \int_{\gamma_{x_2}} f(x_1)a(x_1, x_2)$$

where the amplitude function $a(x_1, x_2)$ is just a smooth cutoff to a neighborhood of $(x_1, x_2) = (0, 0)$, and integration on γ_{x_2} is with respect to the induced Riemannian metric from M_2 . This operator has an adjoint

$$R^*g(x_1) := \int_{\gamma_{x_1}^*} g(x_2)a^*(x_1, x_2)$$

where the amplitude function a^* is another smooth cutoff to $(x_1, x_2) = (0, 0)$. To keep things from being vacuous, we shall assume that $0 \in \gamma_0$ (or equivalently that $0 \in \gamma_0^*$).

A basic example of such an operator is the two-dimensional *Radon transform*, in which $n = 3$, M_1 is the plane \mathbf{R}^2 , and M_2 is the space of lines in \mathbf{R}^2 (endowed with some reasonable Riemannian metric and with some line through the origin designated as the 0 line), with γ given by the tautological map $\gamma_l := l$. The dual

Received by the editors March 19, 2002.

2000 *Mathematics Subject Classification*. Primary 42B15.

Key words and phrases. Radon transforms, double fibration, L^p improving properties, averaging operators.

¹Our conventions may appear somewhat strange, but this choice of notation will be convenient to work with later on.

²All our functions in this paper will be real-valued.

curve γ_x then consists of those lines in M_2 which contain x . The Radon transform is then the familiar operator

$$Rf(l) = \int_l f.$$

Another example of such an operator is that of convolution with a fixed curve $\gamma_0^* \subset \mathbf{R}^{n-1}$, for instance $\gamma_0^* := \{(t, t^2, \dots, t^{n-1}) : t \in \mathbf{R}\}$. We then define $M_1 := M_2 := \mathbf{R}^{n-1}$ and set $\gamma_{x_2} := x_2 - \gamma_0^*$, so the dual curve is $\gamma_{x_1}^* = x_1 + \gamma_0^*$. We can then consider operators R of the form

$$Rf = f * d\sigma$$

where $d\sigma$ is a fixed smooth compactly supported measure on γ_0^* .

This class of operators was first introduced in [11]; see also [21], [24], [23]. We are interested in determining the exponents $1 \leq p, q \leq \infty$ such that R maps $L^p(M_1)$ to $L^q(M_2)$ locally in a neighborhood of the origin, perhaps shrinking the support of the cutoff a if necessary. As will become clear, such a mapping property is easy to show if $q \leq p$; the interesting case is when $q > p$. Observe that such a mapping property is essentially independent of the metric structure on M_1 , M_2 or on the cutoff a (assuming the support is sufficiently small) and is indeed invariant under diffeomorphisms of M_1 and M_2 . Thus we expect the set of exponents (p, q) with this mapping property to depend only on diffeomorphism-invariant properties of the family of curves $x_2 \mapsto \gamma_{x_2}$.

Having given an informal description of our problem of interest, we now give a formal setup (based on that in [11], [21]; a prototype of this “double fibration” formulation appeared earlier in [23]) which will be more convenient to work with. In addition to the $n - 1$ -dimensional manifolds M_1 and M_2 given earlier, we shall also work with an n -dimensional smooth Riemannian manifold Σ with another preferred origin $0 = 0_\Sigma$. We will also assume we have smooth maps $\pi_1 : \Sigma \rightarrow M_1$ and $\pi_2 : \Sigma \rightarrow M_2$ which map 0 to 0 and are submersions (i.e., the derivative has maximal rank $n - 1$) in a neighborhood of 0 .

We then define the transform R by duality as

$$(2) \quad \int_{M_2} Rf(x_2)g(x_2) dx_2 = \int_{M_1} f(x_1)R^*g(x_1) dx_1 \\ := \int_{\Sigma} f(\pi_1(x))g(\pi_2(x))a(x) dx$$

where dx is the measure on Σ induced by the Riemannian metric, and $a(x)$ is a smooth cutoff to a neighborhood of 0 . (We shall always use x to denote elements of Σ , and x_1, x_2 to denote elements of M_1, M_2 , respectively).

To connect the above formalism with the intuitive discussion given earlier, we assume that the assignment $x_2 \mapsto \gamma_{x_2}$ and the amplitude function $a(x_1, x_2)$ is given and then define Σ to be the n -dimensional manifold

$$\Sigma := \{(x_1, x_2) \in M_1 \times M_2 : x_1 \in \gamma_{x_2}\} = \{(x_1, x_2) \in M_1 \times M_2 : x_2 \in \gamma_{x_1}^*\}$$

with $0_\Sigma := (0_{M_1}, 0_{M_2})$. We then define π_1 and π_2 to be just the coordinate projections $\pi_1(x_1, x_2) := x_1$, $\pi_2(x_1, x_2) := x_2$. The reader may easily verify by the change-of-variables formula that the transform given by (1) then obeys the formula (2), if the support of the cutoffs are sufficiently small. In particular the examples of the Radon transform and convolutions with curves given previously can be put into the above framework.

The class of operators of the form (2) is actually slightly larger than that given by (1), as it allows for the curves γ_{x_2} and $\gamma_{x_1}^*$ to develop cusps. In general it may happen that the kernels of $D\pi_1(0)$ and $D\pi_2(0)$ agree but this cannot be the case in the setting of (1) where Σ is a submanifold of $M_1 \times M_2$ and $\pi_j : \Sigma \rightarrow M_j$ are the restrictions of the coordinate projections from $M_1 \times M_2$. Certain examples of fractional integration along curves fall within the more general class of operators given by (2). (See the remarks section, Section 10.)

Definition 1.1. Let $1 \leq p_1, p_2 \leq \infty$. We say the sextuple $(\Sigma, M_1, M_2, \pi_1, \pi_2, 0_\Sigma)$ is of *strong-type* (p_1, p'_2) if we have an estimate of the form

$$\left| \int_{\Sigma} f_1(\pi_1(x))f_2(\pi_2(x))a(x) \, dx \right| \lesssim \|f_1\|_{L^{p_1}(M_1)}\|f_2\|_{L^{p_2}(M_2)}$$

for all f_1, f_2 and all cutoff functions a supported in a sufficiently small neighborhood of 0. (See the notation section for the definition of \lesssim .)

In the intuitive setting, the above definition (modulo irrelevant technicalities when $p_2 = \infty$) corresponds by duality to R mapping $L^{p_1}(M_1)$ to $L^{p'_2}(M_2)$, where p' is the usual dual exponent to p , $1/p + 1/p' = 1$.

We are interested in the problem of determining, for a fixed sextuple $(\Sigma, M_1, M_2, \pi_1, \pi_2, 0_\Sigma)$, the set of exponents p_1, p_2 for which the sextuple is of strong-type (p_1, p'_2) . This problem is one aspect of a much more general and difficult problem concerning smoothing estimates for (possibly singular or maximal) Radon transforms; the field is too vast to summarize here, but we refer the reader to the survey papers [48], [37], [36], [16] and to the recent papers [8], [47].

We will in fact work with a more convenient setting, that of restricted weak-type estimates.

Definition 1.2. Let $1 \leq p_1, p_2 \leq \infty$. We say the sextuple $(\Sigma, M_1, M_2, \pi_1, \pi_2, 0_\Sigma)$ is of *restricted weak-type* (p_1, p'_2) if we have an estimate of the form

$$(3) \quad \left| \int_{\Sigma} \chi_{E_1}(\pi_1(x))\chi_{E_2}(\pi_2(x))a(x) \, dx \right| \lesssim |E_1|^{1/p_1}|E_2|^{1/p_2}$$

for all measurable subsets $E_1 \subset M_1, E_2 \subset M_2$ and all cutoff functions a supported in a sufficiently small neighborhood of 0. Here we use $|E|$ to denote the measure of E with respect to the Riemannian metric.

Note that we have the trivial bounds of $O(|E_1|)$ and $O(|E_2|)$ for the left-hand side of (3) just from the hypothesis that π_1 and π_2 are local submersions. Thus we automatically have restricted weak-type when $p'_2 \leq p_1$. (A variant of this argument also gives strong-type in the same region.) Henceforth we restrict ourselves to the case $p_1 < p'_2$.

Clearly strong-type (p_1, p'_2) implies restricted weak-type (p_1, p'_2) . Conversely if $1 < p_1 < p'_2 < \infty$ and one has restricted weak-type in an open neighborhood of (p_1, p'_2) , then the Marcinkiewicz interpolation theorem gives strong-type (p_1, p'_2) . Thus if one is willing to lose endpoints, it suffices to consider the restricted weak-type problem.

We can reformulate Definition 1.2 in a more geometric manner which is more convenient to work with.

Proposition 1.3. *Let $1 \leq p_1 < p'_2 \leq \infty$. Then the sextuple $(\Sigma, M_1, M_2, \pi_1, \pi_2, 0_\Sigma)$ is of restricted weak-type (p_1, p'_2) if and only if one has the estimate*

$$(4) \quad |\Omega| \lesssim |\pi_1(\Omega)|^{1/p_1} |\pi_2(\Omega)|^{1/p_2}$$

for all sets $\Omega \subset \Sigma$ in a sufficiently small neighborhood of 0.

Proof. If (4) holds, then (3) follows by setting $\Omega := \pi_1^{-1}(E_1) \cap \pi_2^{-1}(E_2) \cap B$, where B is a sufficiently small neighborhood of 0. Conversely, if (3) holds, then (4) follows by setting $E_1 = \pi_1(\Omega)$ and $E_2 = \pi_2(\Omega)$. \square

Thus, the question is to determine to what extent the size of a set Ω in Σ is controlled by the size of its two projections $\pi_1(\Omega) \subset M_1$ and $\pi_2(\Omega) \subset M_2$. Intuitively, the answer to this question should somehow depend on how “independent” the projections π_1 and π_2 are; for instance, if $M_1 = M_2$ and $\pi_1 = \pi_2$, it is easy to see that there are no estimates of the form (4) other than the trivial ones when $p'_2 \leq p_1$.

To make the notion of “independence” more precise, we introduce two vector fields X_1 and X_2 on Σ . For $j = 1, 2$, we let X_j be a smooth vector field defined on a neighborhood of 0_Σ such that X_j is never zero, and X_j always lies parallel to the fibers of π_j , or equivalently that the push-forward $(\pi_j)_* X_j$ of X_j is identically zero. (Equivalently, $X_j(x)$ is always a non-zero element of the one-dimensional kernel of the derivative map $D\pi_j(x) : T_x \Sigma \rightarrow T_{\pi_j(x)} M_j$.) The vector field X_j is only defined up to multiplication by smooth non-zero scalar functions, but we will not be bothered by this freedom and just work with a fixed choice of X_1 and X_2 . From the classical Picard existence theorem for ODE we can see that such vector fields are guaranteed to exist in a sufficiently small neighborhood of 0.

With these vector fields X_1, X_2 one can then define the *flow maps* $e^{tX_j} : \Sigma \rightarrow \Sigma$ in a neighborhood of 0 for sufficiently small t and $j = 1, 2$ by the ODE

$$\frac{d}{dt} e^{tX_j}(x) = X_j(e^{tX_j}(x)); \quad e^{0X_j}(x) = x.$$

These maps are smooth and form a group in a neighborhood of 0, i.e., $e^s X_j e^t X_j = e^{(s+t)X_j}$. We can interpret the X_j as first-order differentiation operators by the formula

$$X_j f(x) := \frac{d}{dt} f(e^{tX_j} x)|_{t=0}.$$

The set of first-order differential operators with smooth coefficients is closed under Lie brackets. We denote $[X_1, X_2]$ by X_{12} and observe the formula

$$(5) \quad X_{12}(x) := \frac{d^2}{dt_1 dt_2} e^{-t_1 X_1} e^{-t_2 X_2} e^{t_1 X_1} e^{t_2 X_2}(x)|_{t_1=t_2=0}.$$

More generally, we define vector fields X_w for all words w :

Definition 1.4. We define a *word* w to be any non-empty finite ordered collection of 1’s and 2’s. We define the *degree* $\deg(w) \in \mathbf{N} \times \mathbf{N}$ to be the ordered pair $(\deg(w)_1, \deg(w)_2)$, where $\deg(w)_j$ is the number of occurrences of j in w . We define the vector fields X_w for $w \in W$ recursively by $X_1 := X_1, X_2 := X_2$, and

$$X_{wj} := [X_w, X_j] \quad \text{for all } j = 1, 2 \text{ and } w \in W,$$

thus for instance

$$X_{12211} = [[[[X_1, X_2], X_2], X_1], X_1].$$

We give a partial ordering on degrees by writing $(a_1, a_2) \leq (b_1, b_2)$ when $a_1 \leq b_1$ and $a_2 \leq b_2$.

Note that there exists a neighborhood of 0 for which all the vector fields X_w exist and are smooth. At first glance it may appear that these vector fields X_w do not cover all possible Lie bracket combinations of X_1 and X_2 , but the Jacobi identity allows one to write any other Lie bracket combination as a linear combination of the X_w .

We say that X_1 and X_2 obey the *Hörmander condition* at 0 if there exist words $w_1, \dots, w_n \in W$ such that $X_{w_1}(0), \dots, X_{w_n}(0)$ span the tangent space $T_0\Sigma$ at 0. This condition has appeared several times before in work on Radon-like transforms, most notably in [8] (see also [47]) where the following result is established.

Proposition 1.5. *If X_1 and X_2 do not obey the Hörmander condition at 0, then $(\Sigma, M_1, M_2, \pi_1, \pi_2, 0)$ is not of restricted weak-type (p_1, p'_2) for any $1 \leq p_1 < p'_2 \leq \infty$. In particular, there are no non-trivial L^p mapping properties near 0, either of strong-type or of restricted weak-type.*

For completeness we give a quick proof of this result in Section 2, as a consequence of the machinery developed in [8].

It remains to consider the case when X_1 and X_2 do obey the Hörmander condition, and we shall assume this for the rest of the introduction.

Definition 1.6. We define W^n to be the space of all n -tuples (w_1, \dots, w_n) of words. If $I = (w_1, \dots, w_n)$ is an n -tuple, we define the *degree* $\deg(I) \in \mathbf{N} \times \mathbf{N}$ by

$$\deg(I) = \deg(w_1) + \dots + \deg(w_n)$$

and define the function $\lambda_I : \Sigma \rightarrow \mathbf{R}$ by

$$\lambda_I(x) := \det(X_{w_1}(x), \dots, X_{w_n}(x)).$$

There exists a neighborhood of 0 where the functions λ_I are all well-defined smooth functions. The Hörmander condition thus asserts that there exists an n -tuple I_0 such that $\lambda_{I_0}(0) \neq 0$. Henceforth we fix this n -tuple I_0 .

To relate these n -tuples I to our problem (4), we introduce two-parameter Carnot-Carathéodory balls, in the spirit of [27].

Definition 1.7. Let $x \in \Sigma$ be in a sufficiently small neighborhood of 0, and let $0 < \delta_1, \delta_2 \ll 1$ be sufficiently small numbers. We define the *two-parameter Carnot-Carathéodory ball*³ $B(x; \delta_1, \delta_2)$ to be the closure of the set

$$\{e^{t_1\delta_1 X_1} \dots e^{t_k\delta_k X_k}(x) : k \geq 0; |t_1| + \dots + |t_k| \leq 1\}$$

where we adopt the periodic convention that $\delta_{j+2} := \delta_j$ and $X_{j+2} := X_j$.

Informally, the ball $B(x; \delta_1, \delta_2)$ represents (up to constants) the set of all points which can be reached from x by flowing for an amount δ_1 in the X_1 direction and δ_2 in the X_2 direction. From (5) we heuristically expect that we can also flow by $\delta_1\delta_2$ in the X_{12} direction and still stay inside (a constant dilate of) this ball; more

³We chose this definition for the introduction as it is the most intuitive to visualize. However, in our rigorous argument we shall avoid these balls and work with an equivalent family of balls defined by exponentiating various weighted commutators of the X_i . See [27] for a detailed comparison of the two types of balls.

generally, we expect to flow by $\delta^{\deg(w)}$ in the X_w direction, where we adopt the notation

$$\delta^{(n_1, n_2)} := \delta_1^{n_1} \delta_2^{n_2}.$$

Because of this, we heuristically expect $B(x; \delta_1, \delta_2)$ to look something like the convex hull of the points $x \pm \delta^{\deg(w)} X_w$. Following these heuristics, we will eventually be able to obtain a volume estimate (inspired by a similar formula in [27]) which is roughly of the form

$$(6) \quad |B(x; \delta_1, \delta_2)| \sim \sum_I \delta^{\deg(I)} |\lambda_I(x)|$$

where I ranges over a finite set depending on I_0 and on certain restrictions on δ_1 and δ_2 .

The balls $B(x; \delta_1, \delta_2)$ form excellent examples of sets Ω to test (4) with. Indeed, one heuristically expects

$$|\pi_j(B(x; \delta_1, \delta_2))| \sim \frac{|B(x; \delta_1, \delta_2)|}{\delta_j}$$

for $j = 1, 2$. Inserting this into (4) with $\Omega := B(x; \delta_1, \delta_2)$, we obtain the inequality⁴

$$|B(x; \delta_1, \delta_2)| \gtrsim \delta_1^{c_1} \delta_2^{c_2}$$

as a special case of (4), where the positive numbers c_1, c_2 are given by

$$(7) \quad c_1 := p_2 / (p_1 + p_2 - p_1 p_2), \quad c_2 := p_1 / (p_1 + p_2 - p_1 p_2).$$

Comparing this with (6), we are thus led to the following conjecture.

Definition 1.8. We define the *Newton polytope* P of X_1 and X_2 at 0 to be the closed convex hull of the set

$$\{x \in \mathbf{R}^+ \times \mathbf{R}^+ : x \geq \deg(I), \lambda_I(0) \neq 0 \text{ for some } I \in W^n\}.$$

Note that P is indeed the Newton polytope of the right-hand side of (6), if one thinks of this as a Taylor series in δ_1 and δ_2 .

Conjecture 1.9. Let $(\Sigma, M_1, M_2, \pi_1, \pi_2, 0)$ and X_1, X_2 obey the Hörmander condition, and let $1 \leq p_1 < p'_2 \leq \infty$. Then we have strong-type (p_1, p'_2) if and only if the point (c_1, c_2) defined by (7) lies in the closed convex polytope P .

The main result of this paper is that this conjecture is true away from endpoint cases.

Theorem 1.10. Let the notation be as in Conjecture 1.9. Then we have strong-type (p_1, p'_2) when (c_1, c_2) lies in the interior of P . Conversely, if (c_1, c_2) lies in the exterior of P , then one does not even have restricted weak-type (p_1, p'_2) .

As one can verify from elementary algebra, the statement that (c_1, c_2) lies in P is equivalent to $(1/p_1, 1/p'_2)$ lying in the closed convex hull of

$$(8) \quad \{(0, 0), (1, 1)\} \cup \left\{ \left(\frac{\deg(I)_1}{\deg(I)_1 + \deg(I)_2 - 1}, \frac{\deg(I)_1 - 1}{\deg(I)_1 + \deg(I)_2 - 1} \right) : I \in W^n, \lambda_I(0) \neq 0 \right\}.$$

⁴Equivalently, one can obtain this bound by applying the original averaging operator R to the characteristic function of $\pi_1(B(x; \delta_1, \delta_2))$. These sets (which tend to look like squashed neighborhoods of curve arcs) are well known as test sets for these averaging operators; see, e.g., [6], [14], [47].

Similarly, the statement that (c_1, c_2) lies in the interior of P is equivalent to $(1/p_1, 1/p'_2)$ lying in the interior of the convex hull of (8).

For curves in the plane ($n = 3$), Seeger [47] has obtained Theorem 1.10 previously for the operators R in (1) (here Σ is a submanifold of $M_1 \times M_2$ and the π_j 's are the induced coordinate projections from $M_1 \times M_2$). In this case, X_1 and X_2 are necessarily linearly independent at 0 and so the 3-tuples $I \in W^3$ with $\lambda_I(0) \neq 0$ defining (8) reduce to looking at iterated commutators $X_w, w \in W$, which do not lie in the subspace spanned by X_1 and X_2 at 0. See [47] for details.

We now give some examples of Theorem 1.10.

Convolution with curves. We consider operators A_n mapping functions on \mathbf{R}^{n-1} to \mathbf{R}^{n-1} defined by

$$A_n f(x) = \int f(x - \gamma(t))a(t) dt$$

where a is a smooth cutoff in a small neighborhood of 0 and γ is a smooth curve in \mathbf{R}^{n-1} with $\gamma(0) = 0$ and $\gamma'(0) \neq 0$. This operator is of the above type with

$$M_1 := M_2 := \mathbf{R}^{n-1}; \quad \Sigma := \{(x, t) : x \in \mathbf{R}^{n-1}, t \in \mathbf{R}\}; \quad 0_{M_1} := 0_{M_2} := 0_\Sigma := 0$$

and

$$\pi_1(x, t) := x; \quad \pi_2(x, t) := x + \gamma(t).$$

We can select the vector fields X_1, X_2 as

$$X_1 := \partial_t; \quad X_2 := \partial_t - \gamma'(t) \cdot \nabla_x.$$

A routine calculation shows that

$$X_{12} = -\gamma''(t) \cdot \nabla_x$$

and more generally that

$$X_{w_1 w_2 \dots w_k} = \epsilon \gamma^{(k)}(t) \cdot \nabla_x$$

for any word $w_1 \dots w_k$, where $\epsilon = -1$ if $w_1 w_2 = 12$, $\epsilon = +1$ if $w_1 w_2 = 21$, and $\epsilon = 0$ otherwise.

Let us now specialize to the maximally curved case, when

$$\gamma(0), \gamma'(0), \dots, \gamma^{(n-1)}(0) \text{ span } \mathbf{R}^{n-1}.$$

This for instance is the case when

$$(9) \quad \gamma(t) := (t, t^2, \dots, t^{n-1}).$$

Then we see that the vector fields

$$X_1, X_2, X_{12}, X_{12w_3}, \dots, X_{12w_3 \dots w_{n-1}}$$

span $\mathbf{R}^{n-1} \times \mathbf{R}$ at zero, for arbitrary $w_3 \dots w_{n-1} \in \{1, 2\}$. Furthermore, any other n -tuples of vector fields X_w which span have equal or larger degree (with respect to the partial ordering \geq on degrees). A simple computation then shows that the convex hull of (8) is the trapezoid with vertices

$$(0, 0), \left(\frac{n^2 - 3n + 4}{n^2 - n}, \frac{n - 2}{n}\right), \left(\frac{2}{n}, \frac{2n - 4}{n^2 - n}\right), (1, 1).$$

Thus our theorem gives strong-type $(L^{p_1}, L^{p'_2})$ in the interior of this trapezoid and failure of restricted type $(L^{p_1}, L^{p'_2})$ outside this trapezoid (when $1 \leq p_1 < p'_2 \leq \infty$). This result was already obtained by Christ [6] (with earlier work in lower dimensions by [25], [29], [31], [32], [14], [17]), at least in the polynomial case (9); in fact these

papers go further and prove various strong-type or restricted-type estimates on the boundary.

Restricted x-ray transforms. We consider operators of the form

$$R_n f(x', s) := \int f(x' + t\gamma(s), t) a(x', s, t) dt$$

where $n \geq 3$, $x' \in \mathbf{R}^{n-2}$, $\gamma(s) = (s, \dots, s^{n-2})$, $s, t \in \mathbf{R}$ and $a(x', s, t)$ is a bump function. This is the X-ray transform on \mathbf{R}^{n-1} restricted to the line complex of lines whose direction lies in the curve $\{(\gamma(s), 1) : s \in \mathbf{R}\}$. This is a model example of operators studied for instance in [10], [22], [18], [19], [20], [7], [14], [17], [33], and elsewhere. This operator can be placed into the above setting with $M_1 := \{(x', t) \in \mathbf{R}^{n-1}\}$, $M_2 := \{(x', s) \in \mathbf{R}^{n-1}\}$, and

$$\Sigma := \{(x', t, s) \in \mathbf{R}^n\}$$

with 0 being the usual origin and the projection maps π_1, π_2 defined by

$$\pi_1(x', t, s) := (x' + t\gamma(s), t); \quad \pi_2(x', t, s) := (x', s).$$

The associated vector fields X_1, X_2 can thus be chosen to be

$$X_1 = \partial_s - t\gamma'(s) \cdot \nabla_{x'}; \quad X_2 = \partial_t.$$

A computation then shows that

$$\begin{aligned} X_{12} &= -X_{21} = \gamma'(s) \cdot \nabla_{x'}, \\ X_{121} &= -X_{211} = -\gamma''(s) \cdot \nabla_{x'}, \\ X_{1211} &= -X_{2111} = \gamma'''(s) \cdot \nabla_{x'}, \end{aligned}$$

and so forth. Furthermore, all other commutators are zero. Thus, we essentially only have one spanning set of vector fields:

$$X_1, X_2, X_{12}, X_{121}, \dots, X_{121\dots 1},$$

where the last vector field has $n - 3$ consecutive ones. The total degree of this n -tuple can then be computed to be

$$\left(\frac{n^2 - 3n + 4}{2}, n - 1\right).$$

The convex hull of (8) is thus the triangle with vertices

$$(0, 0), (1, 1), \left(\frac{n^2 - 3n + 4}{n(n - 1)}, \frac{n - 2}{n}\right),$$

and so our theorem gives $L^p \rightarrow L^q$ mapping properties in the interior of this triangle and failure of $L^p \rightarrow L^q$ outside this triangle. This agrees with previous results in [13], [14], [17], [33], [7], although the results there include certain endpoints which are not obtained by our methods.

Folds and cusps. When $\Sigma \subset M_1 \times M_2$ is a submanifold and the π_j 's are the induced coordinate projections, the underlying geometry is sometimes best expressed in terms of the microlocal picture

$$\begin{array}{ccc} & C & \\ & \swarrow & \searrow \\ \tilde{\pi}_1 & & \tilde{\pi}_2 \\ T^*M_1 \setminus 0 & & T^*M_2 \setminus 0 \end{array}$$

where $C = N^*\Sigma' = \{(x_1, \xi_1; x_2, \xi_2) \in T^*(M_1 \times M_2) \setminus 0 : (\xi_1, -\xi_2) \perp T_{(x_1, x_2)}\Sigma\}$, the twisted conormal bundle of Σ , is the canonical relation in $T^*M_1 \times T^*M_2$ (we are using perhaps nonstandard notation here). When $n \geq 4$, there must be points in C where $D\tilde{\pi}_2$ and $D\tilde{\pi}_1$ drop rank. The simplest singularity that can occur is a Whitney fold ($S_{1,0}$ in the Thom-Boardman description) and the second simplest stable singularity for $\tilde{\pi}_1$ or $\tilde{\pi}_2$, after folds, is a Whitney cusp; singularity class $S_{1,1,0} := S_{1_2,0}$.

For the restricted x-ray transform R_n above, $\tilde{\pi}_1$ has only singularities that belong to the Morin singularity classes $S_{1_k,0}$, $1 \leq k \leq n-3$ (see [15] for a discussion of these singularity classes and their connections with the operators considered here), while both projections $\tilde{\pi}_1$ and $\tilde{\pi}_2$ for the convolution example A_n (when γ is maximally curved) have singularities at most of type $S_{1_k,0}$, $1 \leq k \leq n-3$. We refer to situations as *one-sided* when conditions are imposed on one projection while no assumptions are imposed on the other projection. On the other hand, situations for which both $\tilde{\pi}_1$ and $\tilde{\pi}_2$ belong to a given singularity class are referred to as *two-sided*. Thus the geometry underlying A_4 is a two-sided fold while that of A_5 is a two-sided cusp. See [22] for the microlocal analysis of the restricted x-ray transform. Also see [13], [14], [16], [39] for more details.

To see examples of how this can be expressed in our setting, consider operators of the form

$$(10) \quad Rf(x) = \int f(\gamma(x, t))a(x, t) dt$$

where $n \geq 3$, $\gamma : \mathbf{R}^{n-1} \times \mathbf{R} \rightarrow \mathbf{R}^{n-1}$ is a smooth map with $\gamma(x, t) \equiv x$ and $a(x, t)$ is a bump function. Thus $x \mapsto \gamma_t(x) := \gamma(x, t)$ is a family of diffeomorphisms for small t and a result in [8] associates to $\{\gamma_t\}$ a unique sequence of smooth vector fields Z_1, Z_2, \dots such that $\gamma_t(x) \sim \exp(\sum t^j Z_j)(x)$ to infinite order as $t \rightarrow 0$. This operator can be put into the above setting with $M_1 = M_2 = \mathbf{R}^{n-1}$, $\Sigma = \{(x, t) : x \in \mathbf{R}^{n-1}, t \in \mathbf{R}\}$, with 0 being the usual origin and

$$\pi_1(x, t) = \gamma_t(x), \quad \pi_2(x, t) = x.$$

(Equivalently, if $\frac{\partial \gamma}{\partial t} \neq 0$, one could take $\Sigma = \{(\gamma_t(x), x) : x \in \mathbf{R}^{n-1}, t \in \mathbf{R}\}$ so that $\Sigma \subset M_1 \times M_2$ is a submanifold containing the diagonal.) The associated vector fields X_1 and X_2 can be chosen to be

$$X_1 = W(x, t) \cdot \nabla_x + \partial_t : \quad X_2 = \partial_t$$

where $W(x, t) = \frac{d}{dh} \gamma_{t+h}^{-1} \circ \gamma_t(x)|_{h=0}$. Observe that $X_{12} = [X_1, X_2] = W'(x, t) \cdot \nabla_x$ where $'$ denotes differentiation with respect to t . Similarly $X_{122\dots 2} = (-1)^k W^{(k)}(x, t) \cdot \nabla_x$ where k is the number of 2's.

In this setting, when $n = 4$, a result of Phong and Stein [39] states that $\tilde{\pi}_2$ (resp. $\tilde{\pi}_1$) has at most Whitney fold singularities iff Z_1, Z_2 , and $Z_3 + \frac{1}{6}[Z_1, Z_2]$ (resp. Z_1, Z_2 and $Z_3 - \frac{1}{6}[Z_1, Z_2]$) are linearly independent. (A similar result holds for cusps; see [14].) A simple computation, using the Baker-Campbell-Hausdorff formula, shows that Z_1, Z_2 and $Z_3 + \frac{1}{6}[Z_1, Z_2]$ are linearly independent at x iff $W(x, 0), W'(x, 0)$ and $W''(x, 0)$ are linearly independent and this is equivalent to

$$\lambda_I(x, 0) := \det(X_1, X_2, X_{12}, X_{122})(x, 0) \neq 0.$$

(More generally, for any $n \geq 4$, a theorem of Greenleaf and Seeger [15] gives the equivalence between the linear independence of the $\{W^{(k)}\}_{0 \leq k \leq n-2}$, the linear independence of some combination of the Z_j 's, and $\tilde{\pi}_2$ having only (strong) Morin singularities, $S_{1_k,0}^+$, $k \leq n - 3$.) Since $I = (1, 2, 12, 122)$ has degree $(3, 4)$, Theorem 1.10 implies that when $n = 4$ and $\tilde{\pi}_2$ has at most Whitney fold singularities, the operator R maps L^{p_1} to $L^{p'_2}$ whenever $(1/p_1, 1/p'_2)$ lies in the interior of the triangle with vertices $(0, 0)$, $(1, 1)$, $(1/2, 1/3)$. This is a particular case of a theorem of Greenleaf and Seeger [13] which requires an additional hypothesis but contains estimates on the boundary of the triangle. A similar result holds when $\tilde{\pi}_1$ has at most Whitney fold singularities.

The restricted x-ray transform provides an example of the sharpness of this one-sided fold theorem. Another interesting example is the following one considered by Secco [45]. Let

$$\gamma_t(x) = x \cdot (t, t^2, at^3)$$

where \cdot denotes the 3-dimensional Heisenberg group multiplication. Then $W(x, t) = (-1, -2t, -(3a + 1/2)t^2 + 1/2x_2 - x_1t)$ and so

$$X_{12} = 2\partial_{x_2} + ((6a + 1)t + x_1)\partial_{x_3},$$

$$X_{122} = -(6a + 1)\partial_{x_3}, \quad X_{121} = (1 - 6a)\partial_{x_3}.$$

All other iterated commutators are zero. When $a = 1/6$, Theorem 1.10 shows that the corresponding averaging operator (10) does not map L^{p_1} to $L^{p'_2}$ whenever $(1/p_1, 1/p'_2)$ lies outside the triangle with vertices $(0, 0)$, $(1, 1)$, $(\frac{1}{2}, \frac{1}{3})$. See [45] for these results including endpoint estimates.

A nondegenerate example \mathbf{R}^5 . As the final example, we consider an operator of the form (10) when $n = 5$ and where

$$\gamma(x, t) = (x_1 + t, x_2 + t^2, x_3 + t^3, x_4 + t^4 + x_2t).$$

One easily computes that $W(x, t) = (-1, -2t, -3t^2, -4t^3 - x_2 + 2t^2)$ and so at the origin we have

$$X_1 = -\partial_{x_1} + \partial_t, \quad X_2 = \partial_t, \quad X_{12} = 2\partial_{x_2},$$

$$X_{121} = -6\partial_{x_3} + 2\partial_{x_4}, \quad X_{122} = -6\partial_{x_3} + 4\partial_{x_4},$$

$$X_{1211} = X_{1212} = X_{1221} = X_{1222} = 24\partial_{x_4};$$

all other commutators are zero. Therefore the only 5-tuples of iterated commutators that contribute have degrees $(5, 5)$, $(7, 4)$, $(6, 5)$, $(5, 6)$ and $(4, 7)$. The convex hull of (8) is then the pentagon with vertices

$$(0, 0), \left(\frac{4}{10}, \frac{3}{10}\right), \left(\frac{5}{9}, \frac{4}{9}\right), \left(\frac{7}{10}, \frac{6}{10}\right), (1, 1).$$

Our theorem thus gives strong-type $(L^{p_1}, L^{p'_2})$ in the interior of this pentagon and failure of restricted weak-type $(L^{p_1}, L^{p'_2})$ outside. This pentagon is a larger region in $(\frac{1}{p_1}, \frac{1}{p'_2})$ space than the corresponding one for the operator A_5 considered above.

In the remarks section (Section 10) we explain the difficulties in extending our arguments for Theorem 1.10 to the endpoint cases. After some geometric preliminaries in Sections 3 and 4, we will prove the first part of Theorem 1.10 in Section 5 and the second part in Sections 6–9.

2. PROOF OF PROPOSITION 1.5

We first give a proof of Proposition 1.5. The idea is to use a quantitative version of the proof of Frobenius’s theorem given by Theorem 8.8 in [8]. In our language, this theorem asserts (among other things) that if X_1 and X_2 do not obey the Hörmander condition at 0, then for any $N > 0$ there exists a submanifold $S \subset \Sigma$ containing 0 of positive codimension such that for all $0 < \delta \ll 1$ and all $x \in S$, $t \in \mathbf{R}$ with $\text{dist}(x, 0) \lesssim \delta$ and $|t| \lesssim \delta$ we have

$$(11) \quad \text{dist}(e^{tX_1}x, S), \text{dist}(e^{tX_2}x, S) \lesssim \delta^N.$$

In other words, S is “almost invariant” under the flow of X_1 and X_2 . The proof of this claim in [8] uses a quantitative form of the Baker-Campbell-Hausdorff formula; it is also possible to proceed using the machinery in this paper but we will not do so here.

Assuming this statement, we test (4) with the following neighborhood of S :

$$\Omega := \{x \in \Sigma : \text{dist}(x, 0) \lesssim \delta; \text{dist}(x, S) \lesssim \delta^N\}.$$

If S has codimension k , then we have

$$|\Omega| \sim \delta^{n-k} \delta^{Nk}.$$

On the other hand, from (11) and the smoothness of X_1, X_2 we see that the set

$$\{e^{tX_1}x : x \in \Omega; |t| \lesssim \delta\}$$

is contained in a set which is essentially the same as Ω (but with different implicit constants in the definition) and so also has measure $\sim \delta^{n-k} \delta^{Nk}$. From Fubini’s theorem and the fact that π_1 is a submersion we thus have

$$|\pi_1(\Omega)| \lesssim |\Omega|/\delta.$$

Similarly we have

$$|\pi_2(\Omega)| \lesssim |\Omega|/\delta.$$

Inserting these facts into (4), we obtain the condition

$$\delta^{n-k} \delta^{Nk} \lesssim (\delta^{n-k} \delta^{Nk} \delta^{-1})^{1/p_1+1/p_2},$$

since δ can be arbitrarily small, this forces

$$(n - k + Nk) \geq (n - k + Nk - 1) \left(\frac{1}{p_1} + \frac{1}{p_2} \right).$$

Since N can be arbitrarily large and $k > 0$, we thus obtain

$$\frac{1}{p_1} + \frac{1}{p_2} \leq 1$$

and the claim follows.

3. NOTATION

In our argument the sextuplet $(\Sigma, M_1, M_2, \pi_1, \pi_2, 0)$ shall be fixed. We assume that X_1, X_2 obey the Hörmander condition. By continuity we may find a fixed n -tuple I_0 and a fixed open neighborhood $V \subset \Sigma$ of 0 such that the vector fields X_w are all smooth and well defined on V , and λ_{I_0} is bounded away from zero on V . Unless otherwise specified, the symbol x will always be assumed to denote a point in V .

In our argument we will need a large parameter $N \gg 1$, which will eventually depend on I_0, p_1, p_2 , and we will also need a small parameter $0 < \varepsilon \ll 1$, which will eventually depend on N and I_0, p_1, p_2 . For the first part of the paper, in which we develop the theory of Carnot-Carathéodory balls, we shall only use the parameter ε ; the N parameter will be used later when we begin to prove (4).

We use C to denote various large numbers which depend on I_0, p_1, p_2, V , and the sextuplet $(\Sigma, M_1, M_2, \pi_1, \pi_2, 0)$. We use C_N to denote various large numbers which depend on the above parameters as well as N , and $C_{N,\varepsilon}$ to denote various large numbers which depend on the above parameters as well as N , and ε . We will always assume implicitly that these constants have been chosen sufficiently large.

Let A, B be positive quantities. We use $A \lesssim B$ or $A = O(B)$ to denote the estimate $A \leq C_{N,\varepsilon}B$, and $A \ll B$ to denote the estimate $A \leq C_{N,\varepsilon}^{-1}B$. We use $A \sim B$ to denote the estimate $A \lesssim B \lesssim A$. Later on we will introduce a variant notation $A \lesssim\lesssim B$.

Thus for instance we have

$$(12) \quad \lambda_{I_0}(x) \sim 1 \quad \text{for all } x \in V.$$

We also let K be a large constant (depending only on $I_0, p_1, p_2, N, \varepsilon$, and the sextuplet) to be chosen later; our constants $C, C_N, C_{N,\varepsilon}$ will *not* be allowed to depend on K unless explicitly subscripted.

4. TWO-PARAMETER CARNOT-CARATHÉODORY BALLS

We now set up the machinery needed to study the Carnot-Carathéodory balls. Our arguments here are very much inspired by the beautiful paper of [27]. However the results in [27] are phrased for one-parameter balls and in order to exploit them, one would have to reproduce many of the arguments for the two-parameter setting. This would have added no new insights beyond those already in [27], so we have adopted a different approach—based more on Gronwall's inequality than the Baker-Campbell-Hausdorff formula—to give the same type of results.

In this section we fix an $x_0 \in V$, which we assume to be sufficiently close to 0. We also assume $\delta = (\delta_1, \delta_2)$ to be an arbitrary pair of numbers which obey the smallness condition

$$0 < \delta_1, \delta_2 \leq C_{N,\varepsilon,K}^{-1}$$

and the non-degeneracy condition

$$\delta_1^{1/\varepsilon} \lesssim \delta_2 \lesssim \delta_1^\varepsilon.$$

These will be the two radii for our two-parameter Carnot-Carathéodory balls. As in the introduction we define

$$\delta^{(n_1, n_2)} := \delta_1^{n_1} \delta_2^{n_2}$$

and

$$(K\delta)^{(n_1, n_2)} := (K\delta_1)^{n_1} (K\delta_2)^{n_2}$$

for any pair (n_1, n_2) .

We let $\mathbf{I} \subset W^n$ denote the set

$$\mathbf{I} := \left\{ I \in W^n : \deg(I) \leq \frac{2}{\varepsilon} \deg(I_0) \right\}.$$

Observe that \mathbf{I} is a finite set containing I_0 , and that from the smallness and non-degeneracy conditions of δ_1, δ_2 ,

$$(13) \quad (K\delta)^{\deg(I)} \lesssim (K\delta)^{\deg(I_0)} \quad \text{for all } I \notin \mathbf{I}.$$

Define the vector-valued function Λ by

$$(14) \quad \Lambda(x) := ((K\delta)^{\deg(I)}\lambda_I(x))_{I \in \mathbf{I}}.$$

We shall see later that $|\Lambda(x)|$ controls the volume of $B(x; \delta_1, \delta_2)$, if x is sufficiently close to 0.

From (12) we have

$$(15) \quad |\Lambda(x)| \gtrsim (K\delta)^{\deg(I_0)};$$

from (13) and the smallness of δ_1, δ_2 we thus have

$$(16) \quad (K\delta)^{\deg(I)}|\lambda_I(x)| \lesssim |\Lambda(x)|$$

for all $I \in W^n$, with the implicit constant here allowed to depend on I .

Fix an $x_0 \in V$ and observe from (14) that we may find an n -tuple $I_{x_0} = (w_1, \dots, w_n)$ in \mathbf{I} such that

$$(17) \quad |(K\delta)^{\deg(I_{x_0})}\lambda_{I_{x_0}}(x_0)| \sim |\Lambda(x_0)|.$$

Fix this I_{x_0} . Define the map $\Phi = \Phi_{x_0}$ from \mathbf{R}^n to V by

$$(18) \quad \Phi(t_1, \dots, t_n) := \exp\left(\sum_{j=1}^n t_j K^{-1}(K\delta)^{\deg(w_j)} X_{w_j}\right)x_0$$

and consider the “balls” $\Phi(B)$, where B is a ball in \mathbf{R}^n centered at the origin with fixed radius C , which we will choose later. It turns out that these sets are roughly comparable to the balls $B(x; \delta_1, \delta_2)$ defined in the introduction (for a detailed discussion, see [27]). In fact in the proof of Theorem 1.10 we shall work exclusively with the $\Phi(B)$ and avoid using the balls $B(x; \delta_1, \delta_2)$.

Since Φ is smooth and

$$(19) \quad \partial_{t_j} \Phi(0) = K^{-1}(K\delta)^{\deg(w_j)} X_{w_j},$$

we see from (17) that

$$\det(D\Phi)(0) \sim K^{-n}|\Lambda(x_0)|$$

and in particular that Φ is locally invertible near the origin. (We will make this statement more quantitative later on.)

For each $w \in W$, let Y_w be the pull-back of the vector field $K^{-1}(K\delta)^{\deg(w)} X_w$ by the map Φ ; these are defined near 0 by the local invertibility of Φ , though we shall shortly show that they are in fact defined on all of B (there is a slight abuse of notation here since the Y_w ’s are not exactly iterated commutators of Y_1 and Y_2 but in fact a multiple of the corresponding commutator—e.g., $[Y_1, Y_2] = KY_{12}$ by the way we have just defined Y_{12}). We also consider the standard Euclidean vector fields $\partial_1, \dots, \partial_n$ on B , as well as the coordinate functions t_1, \dots, t_n . From (19) we have

$$(20) \quad Y_{w_j}(0) = \partial_j(0).$$

Thus the system of vector fields Y_{w_1}, \dots, Y_{w_n} is a perturbation of the Euclidean system $\partial_1, \dots, \partial_n$ near the origin.

We now come to the main result of this section, which gives regularity of the vector fields Y_w and controls the geometry of the ball $\Phi(B)$.

Proposition 4.1. *The vector fields Y_w are well defined and smooth on all of B and obey the estimates*

$$(21) \quad \|Y_w\|_{C^M(B)} \lesssim C_{w,M}$$

for all $w \in W$ and $M \geq 0$, if K is sufficiently large depending on M .

For the vector fields Y_{w_i} we have the more precise bounds

$$(22) \quad Y_{w_i}(t) = \partial_i + O(|t|/K)$$

for $t \in B$ and $i = 1, \dots, n$. In particular we have

$$(23) \quad \det(Y_{w_1}, \dots, Y_{w_n})(t) \sim 1 \quad \text{for all } t \in B.$$

Finally, we have the volume bounds

$$(24) \quad |\Phi(E)| \sim K^{-n} |\Lambda(x_0)| |E|$$

for any subset E of B .

Proof. Step 1. Estimate Y_{w_j} .

The first step is to bound the Y_{w_i} and in particular to prove (22) and (23).

By (18) and the definition of exponentiation we have the vector field identity

$$(25) \quad \sum_{j=1}^n t_j \partial_j = \sum_{j=1}^n t_j Y_{w_j}$$

on B .

Let $1 \leq i \leq n$. Taking the Lie bracket of (25) with Y_{w_i} , we obtain

$$\sum_{j=1}^n Y_{w_i}(t_j) \partial_j + t_j [Y_{w_i}, \partial_j] = \sum_{j=1}^n Y_{w_i}(t_j) Y_{w_j} + t_j [Y_{w_i}, Y_{w_j}]$$

which we rewrite as

$$\sum_{j=1}^n t_j [\partial_j, Y_{w_i} - \partial_i] + (Y_{w_i} - \partial_i) = - \sum_{j=1}^n (Y_{w_i} - \partial_i)(t_j)(Y_{w_j} - \partial_j) - \sum_{j=1}^n t_j [Y_{w_i}, Y_{w_j}].$$

If we write

$$(26) \quad Y_{w_i} =: \partial_i + \sum_{k=1}^n a_i^k(t) \partial_k,$$

then the previous becomes

$$(27) \quad \sum_{j=1}^n \sum_{k=1}^n t_j (\partial_j a_i^k) \partial_k + \sum_{k=1}^n a_i^k \partial_k = - \sum_{j=1}^n \sum_{k=1}^n a_i^j a_j^k \partial_k - \sum_{j=1}^n t_j [Y_{w_i}, Y_{w_j}].$$

Define the vector-valued quantity

$$A(t) := (a_i^j(t))_{1 \leq i, j \leq n};$$

from (26), (20) we observe that $A(0) = 0$. We now try to use (27) to obtain good bounds on A .

From (35) below we see that $[Y_{w_i}, Y_{w_j}]$ can be written as a linear combination of the Y_{w_k} with coefficients $O(K^{-1})$, so that the ∂_k coefficient of this vector field is $O((1 + |A|)/K)$. From (27) we thus have

$$\left| \sum_{j=1}^n t_j (\partial_j a_i^k) + a_i^k \right| \lesssim |A|^2 + (r|A| + r)/K$$

for all k , where $r := (\sum_{i=1}^N t_i^2)^{1/2}$ is the radial coordinate. We can rewrite this as

$$|\partial_r (ra_i^k)| \lesssim |A|^2 + (r|A| + r)/K$$

where $\partial_r = \sum_{j=1}^n \frac{t_j}{r} \partial_j$ is the radial vector field. From the definition of A we thus have

$$|\partial_r (rA)| \lesssim |A|^2 + (r|A| + r)/K.$$

On the other hand, we have $A = 0$ when $r = 0$, so we have the integral inequality

$$r|A(r\omega)| \lesssim \int_0^r |A(s\omega)|^2 + \frac{1}{K}(s|A(s\omega)| + s) ds$$

for all ω on the unit sphere. In particular we have

$$\frac{K|A(r\omega)|}{r} \lesssim 1 + \frac{1}{K} \left(\sup_{0 \leq s \leq r} \frac{K|A(s\omega)|}{s} + \left(\sup_{0 \leq s \leq r} \frac{K|A(s\omega)|}{s} \right)^2 \right)$$

for $r = O(1)$. If K is sufficiently large, we thus see from standard continuity arguments (recall that A is smooth and vanishes at the origin) that A is in fact defined for all r and that $\frac{K|A(r\omega)|}{r} \lesssim 1$ for all $r\omega \in B$. In other words we have

$$(28) \quad A(t) = O(r/K)$$

for all $t \in B$, and (22) and (23) follow.

Step 2. Control Λ and $\lambda_{I_{x_0}}$.

The next step is to understand the behaviour of the vector-valued function Λ . We begin by proving some regularity properties of Λ in the directions X_w .

Let $I \in \mathbf{I}$, $w \in W$, $x \in V$ be arbitrary. We observe the identity

$$(29) \quad \begin{aligned} X_w \det(X_{w_1}, \dots, X_{w_n}) &= \operatorname{div}(X_w) \det(X_{w_1}, \dots, X_{w_n}) \\ &\quad + \sum_{j=1}^n \det(X_{w_1}, \dots, [X_w, X_{w_j}], \dots, X_{w_n}); \end{aligned}$$

this formula is just the Lie derivative of $X_{w_1} \wedge \dots \wedge X_{w_n}$ with respect to the vector field X_w and can be deduced (for instance) by writing everything in coordinates and using the product rule together with the formula

$$([X, Y](x))_i = \sum_{j=1}^n X_j(x) \partial_j Y_i(x) - Y_j(x) \partial_j X_i(x).$$

From (29) and the observation (using the Jacobi identity) that $[X_w, X_{w_j}]$ is a linear combination of those $X_{w'}$ with $\deg(w') = \deg(w) + \deg(w_j)$, we see from (16) that

$$(30) \quad |(K\delta)^{\deg(w)+\deg(I)} X_w \lambda_I(x)| \lesssim C_w |\Lambda(x)|.$$

From this and (14) we obtain

$$(31) \quad |(K\delta)^{\deg(w)} X_w \Lambda(x)| \lesssim C_w |\Lambda(x)|.$$

By iterating (29), we thus obtain

$$(32) \quad |(K\delta)^{\deg(w_1)+\dots+\deg(w_k)+\deg(I)} X_{w_1} \cdots X_{w_k} \lambda_I(x)| \lesssim C_{w,k} |\Lambda(x)|$$

for any $k \geq 1$; note that any term of the form $\text{div}(X_w)$ is smooth and so will behave very nicely with respect to derivatives.

From (31) and (18) we obtain

$$\partial_r |\Lambda(\Phi(t))| \lesssim K^{-1} |\Lambda(\Phi(t))|$$

for all $t \in B$. Similarly, from (30) and (18) we have

$$|\partial_r (K\delta)^{\deg(I_{x_0})} \lambda_{I_{x_0}}(\Phi(t))| \lesssim K^{-1} |\Lambda(\Phi(t))|.$$

From these estimates, Gronwall's inequality, and (17) we thus obtain

$$(33) \quad |(K\delta)^{\deg(I_{x_0})} \lambda_{I_{x_0}}(x)| \sim |\Lambda(x)| \sim |\Lambda(x_0)|$$

for all $x \in \Phi(B)$, if K is sufficiently large. In particular we have

$$(34) \quad \lambda_{I_{x_0}}(x) \sim \lambda_{I_{x_0}}(x_0)$$

for all $x \in \Phi(B)$.

Step 3. Control Y_w .

Having controlled Y_{w_i} , and having controlled the determinants λ_I , we will now be able to control all the other Y_w , and in particular Y_1 and Y_2 . As a consequence we will be able to prove (21).

From Cramer's rule we see that for all $w \in W$ we may write

$$(35) \quad (K\delta)^{\deg(w)} X_w(x) = \sum_{j=1}^n c_{w,j}(x) (K\delta)^{\deg(w_j)} X_{w_j}(x)$$

for all $x \in \Phi(B)$, where the coefficients $c_{w,j}$ are linear combinations of ratios

$$(36) \quad \frac{(K\delta)^{\deg(I)} \lambda_I(x)}{(K\delta)^{\deg(I_{x_0})} \lambda_{I_{x_0}}(x)}$$

for various $I \in W^n$. In particular, from (16) we see that $c_{w,j} = O(1)$ (with the implicit constant depending on w).

Pulling back (35) under Φ we obtain

$$(37) \quad Y_w = \sum_{j=1}^n \tilde{c}_{w,j} Y_{w_j}$$

where $\tilde{c}_{w,j} := c_{w,j} \circ \Phi$.

Fix $M \geq 0$, $1 \leq j, k \leq n$, $w \in W$ and $t \in B$. To prove (21), it suffices by (37) to show the bounds

$$(38) \quad |\nabla_t^M \tilde{c}_{w,j}(t)| \lesssim C_{M,w}$$

and

$$(39) \quad |\nabla_t^M a_j^k(t)| \lesssim C_M.$$

The claims (38), (39) have already been proven for $M = 0$. Now suppose inductively that $M > 0$ and that (38), (39) have already been proven for all smaller

values of M . Recall that $c_{w,j}$ is a linear combination of ratios (36). By repeated applications of the quotient rule, (16), (33), and (32), we thus have that

$$(K\delta)^{\deg(w_{i_1})+\dots+\deg(w_{i_M})}|X_{w_{i_1}} \cdots X_{w_{i_M}} c_{w,j}(x)| \lesssim C_{M,w}$$

for all $w \in W$, $x \in \Phi(B)$ and $1 \leq i_1, \dots, i_M \leq n$. Pulling this back by Φ and throwing away some powers of K , we obtain

$$|Y_{w_{i_1}} \cdots Y_{w_{i_M}} \tilde{c}_{w,j}(t)| \lesssim C_{M,w}$$

for all $w \in W$, $t \in B$ and $1 \leq i_1, \dots, i_M \leq n$. Expanding out the derivative operators $Y_{w_{i_j}}$ using (22) and the induction hypothesis (38), (39) for smaller values of M , we obtain

$$|\partial_{i_1} \partial_{i_2} \cdots \partial_{i_M} \tilde{c}_{w,j}(t)| \lesssim C_{M,w} \left(1 + \frac{1}{K} |\nabla^M \tilde{c}_{w,j}|\right)$$

and (38) follows if K is sufficiently large depending on M .

Now we show (39). Let D be a constant coefficient operator of order M . We take the Lie bracket of D with (27). From the Euler identity

$$\sum_{j=1}^n Dt_j(\partial_j a_i^k) = MDa_i^k + \sum_{j=1}^n t_j(\partial_j Da_i^k)$$

we obtain

$$\begin{aligned} & \sum_{j=1}^n \sum_{k=1}^n t_j(\partial_j Da_i^k) \partial_k + (M+1) \sum_{k=1}^n (Da_i^k) \partial_k \\ &= - \sum_{j=1}^n \sum_{k=1}^n D(a_i^j a_j^k) \partial_k - \sum_{j=1}^n [D, t_j [Y_{w_i}, Y_{w_j}]]. \end{aligned}$$

We now estimate the terms on the right-hand side. From the inductive hypothesis (39), (28) and the Leibnitz rule we have

$$|D(a_i^j a_j^k)| \leq C_M(1 + r|\nabla^M A|).$$

We can write $[Y_{w_i}, Y_{w_j}]$ using the Jacobi identity as a linear combination of Y_w . From (37), the Leibnitz rule, the inductive hypotheses (38), (39) for smaller values of M , and the claim (38) for M that was just proven, we thus have

$$|[D, t_j [Y_{w_i}, Y_{w_j}]]| \leq C_M(1 + r|\nabla^M A|).$$

Combining these facts together, we obtain

$$\left| \sum_{k=1}^n r(\partial_r Da_i^k) \partial_k + (M+1) \sum_{k=1}^n (Da_i^k) \partial_k \right| \leq C_M(1 + r|\nabla^M A|).$$

Taking ∂_k coefficients and then letting i vary, we thus have

$$|r\partial_r DA + (M+1)DA| \leq C_M(1 + r|\nabla^M A|).$$

Multiplying by r^M , and letting D vary, this becomes

$$|\partial_r(r^{M+1}\nabla^M A)| \leq C_M(r^M + r^{M+1}|\nabla^M A|).$$

By Gronwall's inequality (and noting from the a priori smoothness of all quantities that $r^{M+1}\nabla^M A$ vanishes at the origin), we thus have

$$|r^{M+1}\nabla^M A| \leq C_M r^{M+1},$$

and (39) follows.

Step 4. Proof of (24).

From the chain rule and the definition of the Y_w we have the identity

$$\det(D\Phi) = \frac{\det(K^{-1}(K\delta)^{\deg(w_1)}X_{w_1}, \dots, K^{-1}(K\delta)^{\deg(w_n)}X_{w_n})}{\det(Y_{w_1}, \dots, Y_{w_n})}.$$

From (23), (34), (17) we thus have

$$\det(D\Phi)(t) \sim K^{-n}|\Lambda(x_0)| \quad \text{for all } t \in B.$$

To prove (24), it thus suffices to show that Φ is essentially one-to-one on B . More precisely, we will be able to cover the large ball B by $O(1)$ small balls of radius ~ 1 such that Φ is one-to-one on each ball.

To prove this, suppose for contradiction that there existed t, t' in B with $t \neq t'$, $|t - t'| \ll 1$ and $\Phi(t) = \Phi(t')$. From (22), (39), (23) and the inverse function theorem we see that there exists a non-zero linear combination $Y = \sum_{i=1}^N \alpha_i Y_{w_i}$ with $\alpha_i = O(1)$ such that $t' = \exp(Y)t$. This implies that $\Phi(t') = \exp(X)\Phi(t)$, where $X := \sum_{i=1}^N \alpha_i K^{-1}(K\delta)^{\deg(w_i)}X_{w_i}$. However from (33) we see that X is a smooth non-vanishing vector field, with a C^1 norm of $O(\delta_1 + \delta_2)$ (for instance). It is easy to see that for such vector fields, $\exp(X)$ cannot have any fixed points; indeed for such fields we differentiate the identity

$$\frac{d}{dt} \exp(tX)(x) = X(\exp(tX)(x))$$

in time to obtain

$$\frac{d^2}{dt^2} \exp(tX)(x) = O\left((\delta_1 + \delta_2) \left| \frac{d}{dt} \exp(tX)(x) \right| \right)$$

and from Gronwall's inequality we thus have

$$\frac{d}{dt} \exp(tX)(x) = X(x) + O((\delta_1 + \delta_2)|t||X(x)|)$$

for $|t| \leq 1$, and the claim follows if $\delta_1 + \delta_2$ is sufficiently small. Thus we obtain the desired contradiction. \square

We remark that one can use Proposition 4.1 to show that

$$|B(x_0; \delta_1, \delta_2)| \sim |\Lambda(x_0)|$$

by using (24), (21) and (iterated versions of) (5); we omit the details. In particular we can prove a rigorous version of (6) (with I restricted to \mathbf{I}).

5. NECESSITY OF THE NEWTON POLYTOPE

We now have enough machinery to prove the easy direction of Theorem 1.10. More precisely, we assume $1 \leq p_1 < p'_2 \leq \infty$ are such that the associated point (c_1, c_2) defined by (7) is outside of the Newton polytope P .

The (quarter-infinite) polytope P only has a finite number of vertices, all of which are of the form $\deg(I)$ for some collection of n -tuples $I \in W^n$. By choosing ε sufficiently small, one can assume that all of these n -tuples are in \mathbf{I} .

Since (c_1, c_2) is outside of P , we may find a half-plane of the form

$$\{(x_1, x_2) : a_1x_1 + a_2x_2 \geq 1\}$$

which contains P but does not contain c_1, c_2 . Since P is quarter-infinite and contained in the quadrant $\{(x_1, x_2) : x_1, x_2 \geq 1\}$, we may take $0 < a_1, a_2 < 1$; by taking ε sufficiently small, we may assume that $\varepsilon \leq a_1, a_2$ and that $a_1c_1 + a_2c_2 \leq 1 - \varepsilon$.

We let $0 < \delta_0 \ll 1$ be a small parameter, and set

$$\delta := (\delta_1, \delta_2) := (\delta_0^{a_1}, \delta_0^{a_2}).$$

Observe from construction that

$$\delta^{(c_1, c_2)} \geq \delta_0^{-\varepsilon} \delta^{(x_1, x_2)}$$

for all $(x_1, x_2) \in P$. In particular we have

$$(40) \quad \delta^{(c_1, c_2)} \gtrsim C_K^{-1} \delta_0^{-\varepsilon} |\Lambda(0)|.$$

We now apply the machinery of the previous section, with $x_0 := 0$ and δ set as above. We set

$$\Omega := \Phi(B_{1/K}),$$

where $B_{1/K}$ is the ball of radius $1/K$ centered at 0. From (24) we have

$$(41) \quad |\Omega| \sim C_K^{-1} |\Lambda(0)|.$$

On the other hand, from (21) we see that

$$e^{tY_j} B_{1/K} \subset B$$

for all $j = 1, 2$ and $|t| \leq 1/K$. Thus

$$e^{t\delta_j X_j} \Omega \subset \Phi(B)$$

for all $j = 1, 2$ and $|t| \leq 1/K$. In particular, from (24) we have

$$\left| \bigcup_{t: |t| \leq 1/K} e^{t\delta_j X_j} \Omega \right| \lesssim C_K |\Lambda(0)|.$$

Since X_j lies in the kernel of $d\pi_j$, is bounded away from zero, and π_j is a submersion, we thus have

$$\frac{1}{K} |\delta_j \pi_j(\Omega)| \lesssim C_K |\Lambda(0)|.$$

Inserting this bound and (41) into (4), we obtain

$$|\Lambda(0)| \lesssim C_K (|\Lambda(0)|/\delta_1)^{1/p_1} (|\Lambda(0)|/\delta_2)^{1/p_2}$$

which after some algebra and (7) becomes

$$|\Lambda(0)| \gtrsim C_K \delta^{(c_1, c_2)}.$$

But this contradicts (40) if we choose δ_0 sufficiently small. This concludes the proof of necessity for (c_1, c_2) to lie in the Newton polytope P .

6. SUFFICIENCY OF THE INTERIOR OF THE POLYTOPE: NOTATION AND PRELIMINARY REDUCTIONS

We now begin the proof of the more difficult direction of Theorem 1.10. For this direction we assume $1 \leq p_1 < p_2 \leq \infty$ is such that the exponent pair (c_1, c_2) defined by (7) lies in the interior of the polytope P , and our task is to prove (4) (the strong-type estimates then following by Marcinkiewicz interpolation).

We may assume that the set V is contained within a $1/K$ -neighborhood of the origin. In particular, if a is a Lipschitz function on V such that $|a(0)| \sim 1$, then $|a(x)| \sim 1$ for all $x \in V$ (if K was chosen sufficiently large).

Fix p_1, p_2, c_1, c_2 , and let $\Omega \subset V$ be a fixed set of positive measure. We define the quantities $\alpha_j = \alpha_j(\Omega)$ for $j = 1, 2$ by

$$(42) \quad \alpha_j := \frac{|\Omega|}{|\pi_j(\Omega)|};$$

these are the analogues of the δ_j in the previous section. Note that $0 < \alpha_j \lesssim 1/K$ (since V has diameter $O(1/K)$). By the same algebra used in the previous section, we can rewrite (4) as

$$(43) \quad |\Omega| \gtrsim \alpha_1^{c_1} \alpha_2^{c_2}.$$

We adopt the periodic notation that $\alpha_{j+2} = \alpha_j$ for all j . By symmetry we may assume that

$$\alpha_1 \geq \alpha_2.$$

Intuitively, the proof of (43) runs as follows⁵. From (42), we expect that for “most” points x_0 in Ω , one can flow using e^{tX_1} for some random $|t| \lesssim 1$, and return to Ω with probability $\sim \alpha_1$. Similarly for e^{tX_2} and α_2 . In particular, if one considers a generic expression of the form

$$(44) \quad e^{t_n X_n} \dots e^{t_1 X_1} x_0,$$

then such a point should have a probability at least $\gtrsim \alpha_n \dots \alpha_1$ of lying in Ω . If this expression as a function of $t = (t_1, \dots, t_n)$ had good injectivity properties (in that its Jacobian was bounded away from zero), this would then give a significant improvement to (43), namely

$$|\Omega| \gtrsim \alpha_n \dots \alpha_1 = \alpha_1^{\lfloor (n+1)/2 \rfloor} \alpha_2^{\lfloor n/2 \rfloor}.$$

This however is too good to be true, as the results of Section 5 already indicate. The problem is that the Jacobian of the map

$$(t_1, \dots, t_n) \rightarrow e^{t_n X_n} \dots e^{t_1 X_1} x_0$$

can degenerate to zero on a complicated set (although the Hörmander condition does ensure that this Jacobian does not vanish entirely).

In the work of Christ [6], the operator of convolution with the curve (t, t^2, \dots, t^{n-1}) was considered. This curve has special algebraic properties (for instance, the Jacobian turns out to be a Vandermonde determinant), and one could obtain a large set of times (t_1, \dots, t_n) where the Jacobian was reasonably large⁶ by combinatorics. However in our more general context the Jacobian has no special algebraic structure and the zero set is about as badly behaved as a generic algebraic variety. While algebraic varieties do have some structure which can be exploited (for instance, their intersections with other algebraic sets have a bounded number of connected

⁵Our argument follows the iteration ideas of Christ [6] (which in turn were inspired by the work of Bourgain, Wolff, and Schlag on Kakeya-type problems) adapted to the general vector field setting. Of course, the idea of iterating an operator supported mostly on one-dimensional sets to increase the dimension of the support is hardly new (see, e.g., [4], [9], [5], [12]); our main innovation is the utilization of the concept of width introduced in the previous section.

⁶Strictly speaking, the argument in [6] requires an iteration by $2n - 2$ steps instead of n in order to locate a sub- n -tuple of times which can avoid the zero set of the Jacobian. This additional complication is avoided in our approach by restricting our times to a central set of small width; if the times are allowed to range over long distances (much greater than their natural width), then one can give counterexamples to the above iteration scheme working if one only uses n iterations (Christ, personal communication).

components), it is not clear how a general set of times t_1, \dots, t_n can avoid this zero set.

To get around the possible algebraic complexity, we go back and do a preliminary pruning of the set of times t_1, \dots, t_n which we use in the above argument. In particular, by using a simple stopping time argument, we can restrict each time t_i to a central set of fixed width (the central nature comes because the set of times is essentially a difference set), although it will cost us an epsilon in the exponents to do so. Once one has restricted the time sets in this manner, then (44) is restricted to one of the two-parameter Carnot-Carathéodory balls constructed earlier. By using a rescaling adapted to the vector fields (specifically, we use the map Φ constructed in Section 4), we can assume that these widths are close to 1, at which point one can iterate Lemma 7.3 below to give satisfactory lower bounds on the Jacobian. (This trick was inspired by the “two ends reduction” of Wolff [49].) The condition that (c_1, c_2) lie in the interior of P then comes naturally from the Jacobian of the rescaling transformation (i.e., from (24)), by computations similar to those in Section 5.

To make the above argument rigorous, we shall need some additional notation. Let $A \lesssim B$ denote the estimate

$$A \leq C_{K,N,\varepsilon} \alpha_2^{-C_N \varepsilon} B,$$

and write $A \approx B$ if $A \lesssim B$ and $B \lesssim A$. If E and F are sets, we say that E is a *refinement* of F , or $E \prec F$, if $E \subset F$ and $|E| \approx |F|$.

To prove (43), it will suffice to show that

$$(45) \quad |\Omega| \gtrsim \alpha_1^{c_1} \alpha_2^{c_2 + C/N}.$$

Indeed, if (45) held, then by making N large and *then* making ε small, we obtain (43) with c_1, c_2 replaced by small perturbations of c_1, c_2 . But since (c_1, c_2) is an arbitrary *interior* point of P , and $\alpha_1, \alpha_2 \lesssim 1$, we thus obtain (43) for all (c_1, c_2) in the interior of P as desired.

It remains to show (45). This will occupy the remaining sections of the paper, after a digression on widths.

7. AN INTERLUDE ON WIDTHS

We now introduce the main innovation of this paper, that of a central one-dimensional set with a fixed width.

Definition 7.1. Let S be a subset of $[-C, C]$ with non-zero measure. We say that S is *central with width w* for some $w > 0$ if

$$|x| \lesssim w \quad \text{for all } x \in S$$

and

$$(46) \quad |I \cap S| \lesssim (|I|/w)^\varepsilon |S|$$

for all intervals I .

Thus central sets with width w are spread out somewhat evenly within an interval $[-Cw, Cw]$ which is centered at the origin. It is essential that this interval is *central* (i.e., centered at the origin) and that there is absolutely no portion of S outside this interval; otherwise the iteration argument we use will not give a set contained in a small Carnot-Carathéodory ball.

Note that central sets have diameter comparable to their width. In particular:

Corollary 7.2. *If one central set of width w is a subset of another central set of width w' , then $w \lesssim w'$.*

A central set S with width w is supported on an interval of length $\sim w$ and is not concentrated on any smaller interval. This non-concentration gives such sets good properties when it comes to obtaining lower bounds of integrals of these sets. The idea of using such a non-concentration condition was inspired by the work of Wolff [49] on the Kakeya problem, in which he utilized a very similar “two-ends” condition to achieve a similar effect.

In a later proposition (Lemma 8.2) we shall construct some central sets with some width w . In the remainder of this section, we show how the width property is useful.

Lemma 7.3. *Let $P(t)$ be a polynomial of one real variable of degree $d = O(1)$ and $\|P\|_\infty \lesssim 1$ on an interval $[-C, C]$. Let S be a central set in $[-C, C]$ of some width $0 < w \ll 1$. Suppose that we have the lower bound*

$$(47) \quad \left| \left(\frac{d}{dt} \right)^j P(0) \right| \gtrsim w^m$$

for some $j \geq 0$ and $m \geq 0$. Then, if w is sufficiently small, we can find a subset $S' \subset S$ of measure

$$|S'| \approx |S|$$

such that

$$(48) \quad |P(t)| \gtrsim w^{C(m,d)} \quad \text{for all } t \in S'.$$

Observe how the hypothesis (47) only requires a lower bound at a single point, while the conclusion (48) yields a lower bound on a large fraction of the set S . Such a bound would not be possible without the non-concentration assumption (46) (unless one was willing to lose a power of $|S|$ in (48), which would almost certainly lead to much worse exponents⁷ p_1, p_2 for the positive results in Theorem 1.10).

Because of the unspecified power of w in (48) we will only be able to usefully apply this lemma when the set S has width close to 1. However we will be able to use the map Φ developed in the previous section to rescale sets of small width to sets of width ≈ 1 to resolve this problem.

Proof. By considering the homogeneous functional

$$\|P\| := \max_{k \leq d} \inf_{t \in [-C, C]} |P^{(k)}(t)|$$

on the finite dimensional space of polynomials of degree d , we see that (47) implies

$$(49) \quad |P^{(k)}(t)| \gtrsim w^m \quad \text{on } [-C, C]$$

for some $k \leq d$. We may assume $k \geq 1$. Consider the set of real numbers t such that $P(t) = O(w^{m+2d})$. This is contained in the union of at most $O(d)$ intervals, each interval has length at most $O(w^2)$ by (49). By (46), each such interval thus contains at most $O(w^\varepsilon)$ of the set S . If we then set S' to be the portion of S

⁷Admittedly, if one was very careful, then one might still be able to obtain near-optimal results; see, e.g., [6]. However the general theory of lower bounds is far from optimal at this stage; see [3] for some recent progress.

which is outside these intervals, we have $|S'| \approx |S|$ (if w is sufficiently small) and $|P(t)| \gtrsim w^{m+2d}$ on S' as desired. \square

8. MAIN ARGUMENT

We now give the rigorous proof of (45). Let $\mathbf{I}_0 \subset \mathbf{I}$ denote those elements of \mathbf{I} for which $\lambda_I(0) \neq 0$. Since \mathbf{I}_0 is finite and non-empty, we may assume by continuity that $\lambda_I(x) \sim 1$ for all $x \in V$. In particular we have

$$(50) \quad |\Lambda(x)| \gtrsim \sum_{I \in \mathbf{I}_0} \delta^{\deg(I)}$$

for all $x \in V$ and $0 < \delta_1, \delta_2 \ll 1$.

Choose $\varepsilon > 0$ small enough so that \mathbf{I}_0 contains all the vertices of P . Since (c_1, c_2) lies in the polytope P and $0 < \alpha_1, \alpha_2 \lesssim 1$, we see from construction of P that

$$\alpha_1^{c_1} \alpha_2^{c_2} = \alpha^{(c_1, c_2)} \lesssim \sum_{I \in \mathbf{I}_0} \alpha^{\deg(I)}.$$

Thus to prove (45), it will suffice to show

$$(51) \quad |\Omega| \gtrsim \alpha_2^{C/N} \sum_{I \in \mathbf{I}_0} \alpha^{\deg(I)}.$$

From (42) we expect, heuristically, that for generic points in $x \in \Omega$ and $t = O(1)$, the point $e^{tX_j}(x)$ has a probability $\sim \alpha_j$ of remaining in Ω . To make this precise, we will need to pass from Ω to various refinements of Ω , which we now construct.

We first make a technical reduction to reduce the diameter of Ω . By the pigeon-hole principle (covering V by balls of radius $O(\alpha_2^{2/N})$) one can find a subset $\tilde{\Omega} \subset \Omega$ which has diameter $O(\alpha_2^{2/N})$ and is such that

$$(52) \quad |\tilde{\Omega}| \gtrsim \alpha_2^{C/N} |\Omega|.$$

Henceforth we fix this $\tilde{\Omega}$.

Definition 8.1. Let j be an integer. A j -sheaf Ω' of width w_j is defined to be a refinement Ω' of $\tilde{\Omega}$ such that, for all $x \in \Omega'$, the set

$$(53) \quad \{ |t| \ll 1 : e^{tX_j}(x) \in \Omega' \}$$

is a central set of width w_j and measure $\gtrsim \alpha_2^{C/N} \alpha_j$.

This definition and the following lemma is very reminiscent of the restricted weak-type reductions used in, e.g., [6]. Our main innovation here is that the sets of times are central sets with fixed width; this will be crucial in placing (a large subset of) Ω inside one of the Carnot-Carathéodory balls used earlier.

Lemma 8.2. For any refinement Ω' of $\tilde{\Omega}$ and any integer j , there exists a refinement $\langle \Omega' \rangle_j$ of Ω' such that $\langle \Omega' \rangle_j$ is a j -sheaf of some width $\alpha_2^{C/N} \alpha_j \lesssim w_j \lesssim \alpha_2^{2/N}$.

Proof. Fix j, Ω' . It will be convenient to change coordinates so that Σ is locally like \mathbf{R}^n and X_j is just the constant vector field e_n . In this case we can reparameterize M_j as \mathbf{R}^{n-1} , so that $\pi_j : \mathbf{R}^n \rightarrow \mathbf{R}^{n-1}$ just becomes the Euclidean projection

$$\pi_j(x_1, \dots, x_n) = (x_1, \dots, x_{n-1}).$$

We can of course let 0_Σ be the usual origin $(0, \dots, 0)$.

Consider the function $(\pi_j)_*\chi_{\Omega'}$, the push-forward of $\chi_{\Omega'}$ by π_j . This function has L^1 norm $|\Omega'|$ and is supported on a set of measure

$$\leq |\pi_j(\Omega)| = |\Omega|/\alpha_j \lesssim \alpha_2^{-C/N} |\Omega'|/\alpha_j.$$

Because of this, we can find a set E in \mathbf{R}^{n-1} such that $(\pi_j)_*\chi_{\Omega'} \gtrsim \alpha_2^{C/N} \alpha_j$ on E and

$$\int_E (\pi_j)_*\chi_{\Omega'} \approx |\Omega'|.$$

At this point one could attempt to set $\langle \Omega' \rangle_j$ equal to $\Omega' \cap \pi_j^{-1}(E)$, but we would not get the crucial property that (53) is central with a fixed width. To obtain this additional property, we must refine further.

Let x be any point in $E \subset \mathbf{R}^{n-1}$, and consider the set

$$S(x) := \{t : (x, t) \in \Omega'\}.$$

From the diameter restriction on Ω_0 we see that this is a subset of $[-C\alpha_2^{2/N}, C\alpha_2^{2/N}]$ with measure $\gtrsim \alpha_2^{C/N} \alpha_j$.

Let $I(x)$ be the dyadic interval of minimal length such that

$$|I(x) \cap S(x)| \geq \frac{1}{4} |I(x)|^\varepsilon |S(x)|;$$

if there are many intervals of this length, we choose $I(x)$ arbitrarily. Let $w(x)$ denote the length of $I(x)$.

This interval is well defined and must have length $\alpha_2^{C/N} \alpha_j \lesssim w(x) \lesssim \alpha_2^{2/N}$. Set $S'(x) := I(x) \cap S(x)$. By construction we see that

$$(54) \quad |S'(x)| \approx |S(x)| \gtrsim \alpha_2^{C/N} \alpha_j$$

and that

$$(55) \quad |I \cap S'(x)| \lesssim (|I|/w(x))^\varepsilon |S'(x)|$$

for all intervals I . (Note that any interval I can be covered by $O(1)$ dyadic intervals of comparable length, and each of the dyadic intervals are either contained in $I(x)$ or are disjoint from it.)

Since $w(x)$ can only take $O(\log(1/\alpha_2)) \approx 1$ many values, there exists an $\alpha_2^{C/N} \alpha_j \lesssim w_j \lesssim \alpha_2^{2/N}$ such that

$$\int_{x \in E: w(x)=w_j} |S'(x)| \approx \int_E |S'(x)|,$$

which by the previous estimates implies that

$$\int_{x \in E: w(x)=w_j} |S'(x)| \approx \int_E |S(x)| = \int_E (\pi_j)_*\chi_{\Omega'} \approx |\Omega'| \approx |\tilde{\Omega}|.$$

We then choose such a w_j and define

$$\langle \Omega' \rangle_j := \{(x, t) : x \in E, w(x) = w_j, t \in S'(x)\}.$$

We now verify that $\langle \Omega' \rangle_j$ is a j -sheaf of width w_j . Choose $(x_0, t_0) \in \langle \Omega' \rangle_j$. By construction, t_0 lies in $S'(x_0)$, which is itself contained in an interval $I(x_0)$ of length w_j . Since the set of times (53) corresponding to (x_0, t_0) is just $S'(x_0) - t_0$, we see from (54), (55) that (53) is indeed a central set with width w_j and measure

$\gtrsim \alpha_2^{C/N} \alpha_j$. (The all-important centrality condition follows since $I(x_0)$ has length w_j , so that $S'(x_0) - t_0$ is contained in the interval $[-w_j, w_j]$.) \square

The above lemma allows one to refine a set Ω so that the π_1 fibers (for instance) have good properties. It is tempting to iterate this lemma so that the π_2 fibers are also good, but unfortunately the latter refinement can destroy some of the properties of the former. One could then try to iterate the lemma repeatedly, but the widths w_1, w_2 may change in doing so. Fortunately we can salvage matters by using a pigeonholing argument (dating back at least to [51]; we use the formulation in [35]; somewhat similar ideas appear in [6]), if we are willing to lose a power of $\alpha_2^{1/N}$:

Corollary 8.3. *There exist dyadic numbers $\alpha_2^{C/N} \alpha_j \lesssim \mu_j \lesssim \alpha_2^{2/N}$ for $j = 1, 2$ and an increasing sequence*

$$\Omega_0 \prec \Omega_1 \cdots \prec \Omega_n \prec \tilde{\Omega}$$

such that for all $0 \leq j \leq n$, the set Ω_j is a j -sheaf whose width w_j satisfies $\alpha_2^{C/N} \mu_j \lesssim w_j \lesssim \mu_j$. Here we have adopted the convention that the subscripts of μ_1, μ_2 are extended periodically.

Proof. We first define a much longer decreasing sequence of refinements

$$\tilde{\Omega} = \Omega^{(0)} \succ \Omega^{(1)} \cdots \succ \Omega^{(N^2)}$$

by setting $\Omega^{(j)} = \langle \Omega^{(j-1)} \rangle_j$ for all $1 \leq j \leq N^2$. (Note that we are extending the construction in Lemma 8.2 periodically in j in the usual fashion.) By construction we see that each $\Omega^{(j)}$ is an j -sheaf of width $w^{(j)}$ for some $\alpha_2^{C/N} \alpha_j \lesssim w^{(j)} \lesssim \alpha_2^{2/N}$. (Note that we are allowed to apply the lemma N^2 times because our implicit constants are allowed to depend on N .)

Since $\Omega^{(j+2)} \subset \Omega^{(j)}$, we see from Corollary 7.2 that $w^{(j+2)} \lesssim w^{(j)}$ for all $1 \leq j \leq N^2 - 2$. Thus if we set N to be large enough, then by the pigeonhole principle one can find an even number $n \leq j_0 \leq N^2$ and numbers μ_1, μ_2 (extended periodically) such that $\alpha_2^{C/N} \alpha_j \lesssim \mu_j \lesssim \alpha_2^{2/N}$ and $\alpha_2^{C/N} \mu_j \lesssim w^{(j)} \lesssim \mu_j$ for all $j_0 - n \leq j \leq j_0$. The claim then follows by setting $\Omega_j = \Omega^{(j_0-j)}$ and $w_j = w^{(j_0-j)}$ (note that $j_0 - j$ and j always have the same parity). \square

Let μ_j, Ω_j be as above. Since Ω_0 is a refinement of $\tilde{\Omega}$, it is non-empty. We fix a point $x_0 \in \Omega_0$. Define δ_1, δ_2 by

$$\delta_j := \alpha_2^{-1/N} \mu_j$$

and extend this periodically in the usual manner, $\delta_{j+2} = \delta_j$. Then we have

$$\alpha_2^{C/N} \delta_j \lesssim w_j \lesssim \alpha_2^{1/N} \delta_j \quad \text{for all } j = 1, \dots, n,$$

$$0 < \alpha_i / \delta_i \lesssim \alpha_2^{-C/N},$$

and

$$\alpha_2^{1+C/N} \lesssim \delta_1, \delta_2 \lesssim \alpha_2^{1/N}.$$

In particular we have

$$\delta_1^{1/\varepsilon} \lesssim \delta_2 \lesssim \delta_1^\varepsilon.$$

For all $0 \leq i \leq n$ define the map Φ_i from a ball of radius $O(1)$ and center 0 in \mathbf{R}^i to Σ by

$$\Phi_i(t_1, \dots, t_i) := e^{t_i \delta_i X_i} \cdots e^{t_1 \delta_1 X_1}(x_0),$$

and let $T_i \subset \mathbf{R}^i$ denote the set

$$T_i := \{t \in \mathbf{R}^i : \Phi_j(t_1, \dots, t_j) \in \Omega_{j-1} \text{ for all } 1 \leq j \leq i\},$$

with the convention that $T_0 = \mathbf{R}^0 = \{(\)\}$, the singleton set consisting of the unique 0-tuple. Observe that for each $0 \leq i < n$ and $t \in T_i$ the set

$$\tau_i(t) := \{t_{i+1} \in \mathbf{R} : (t, t_{i+1}) \in T_{i+1}\}$$

is a central set of measure $\gtrsim \alpha_i/\delta_i$ and width \tilde{w}_i for some $\alpha_2^{C/N} \lesssim \tilde{w}_i \lesssim \alpha_2^{1/N}$; indeed, the sets $\tau_i(t)$ are just the sets (53) rescaled by δ_i .

Since the t_i are bounded (and this is the key place where we use the centrality of the set (53)), we see from (21) that

$$\Phi_j(T_j) \subset \Phi(B) \quad \text{for all } 1 \leq j \leq n$$

where Φ , B are as in Section 4; here we assume the radius of B to be sufficiently large.

Now by construction, the set $\Phi_n(T_n)$ lies in Ω . By (24) we thus have

$$|\Omega| \geq |\Phi_n(T_n)| \approx |\Lambda(x_0)| |\Phi^{-1} \circ \Phi_n(T_n)|.$$

However, we see that

$$\Phi^{-1} \circ \Phi_j = \Psi_j$$

where Ψ_j maps a ball in \mathbf{R}^j to B and is defined by

$$\Psi_j(t_1, \dots, t_j) := e^{t_j Y_j} \dots e^{t_1 Y_1}(0)$$

and the Y_i are the vector fields from Section 4. Thus we have

$$|\Omega| \gtrsim |\Lambda(x_0)| |\Psi_n(T_n)|.$$

In the next section we shall show the bound

$$(56) \quad |\Psi_n(T_n)| \gtrsim \alpha_2^{C/N} \frac{\alpha_1}{\delta_1} \dots \frac{\alpha_n}{\delta_n}.$$

Inserting this into the previous estimate and using the fact that α_j and δ_j are periodic, we can write this as

$$|\Omega| \gtrsim \alpha_2^{C/N} |\Lambda(x_0)| \left(\frac{\alpha_1}{\delta_1}\right)^{\lfloor (n+1)/2 \rfloor} \left(\frac{\alpha_2}{\delta_2}\right)^{\lfloor n/2 \rfloor}.$$

Now for all $I \in \mathbf{I}_0$, we observe that $\deg(I)_1, \deg(I)_2 \geq n-1$. This is because all the vector fields X_w in I must be distinct (otherwise $\lambda_I \equiv 0$), and apart from the vector fields X_1, X_2 , every other vector field X_w has $\deg(w)_1, \deg(w)_2 \geq 1$. Since⁸

$$n-1 \geq \lfloor (n+1)/2 \rfloor, \lfloor n/2 \rfloor$$

and $\alpha_j/\delta_j \lesssim 1$, we thus have

$$|\Omega| \gtrsim \alpha_2^{C/N} |\Lambda(x_0)| \left(\frac{\alpha_1}{\delta_1}\right)^{\deg(I)_1} \left(\frac{\alpha_2}{\delta_2}\right)^{\deg(I)_2}$$

for all $I \in \mathbf{I}_0$. Combining this with (50), we obtain (51) as desired.

It remains to show (56). This will occupy the entirety of the next section.

⁸Notice the “slack” in the argument here. What this is saying is that when one has a two-ends condition on the set of times (53), then the lower bounds on $|\Omega|$ improve substantially. This is consistent with the experience with the Keakeya problem in, e.g., [49].

9. CONCLUSION OF THE ARGUMENT

We now prove (56). To simplify the notation, we write $\beta_j := \alpha_j/\delta_j$. Our situation is as follows. We have a point $x_0 \in V$ which is within $O(1/K)$ of the origin 0. We have sets $T_i \in \mathbf{R}^i$ for $i = 0, \dots, n$, which have the structure

$$T_{i+1} = \{(t, t_{i+1}) \in \mathbf{R}^{i+1} : t \in T_i; t_{i+1} \in \tau_i(t)\}$$

for $i = 0, \dots, n - 1$, with $T_0 = \mathbf{R}^0 = \{()\}$ and for each $t \in T_i$, the set $\tau_i(t)$ is a central set of width

$$\alpha_2^{C/N} \lesssim \tilde{w}_i \lesssim \alpha_2^{1/N}$$

and $\tau_i(t)$ has measure

$$(57) \quad |\tau_i(t)| \gtrsim \alpha_2^{C/N} \beta_i.$$

From our upper bound on \tilde{w}_i we see that

$$T_i \subseteq [-r, r]^i$$

for all $0 \leq i \leq n$ and some $r \sim \alpha_2^{1/N}$, which we fix.

From (57) and Fubini's theorem we have

$$|T_{i+1}| \gtrsim \alpha_2^{C/N} \beta_i |T_i|$$

and hence

$$|T_n| \gtrsim \alpha_2^{C/N} \beta_1 \cdots \beta_n.$$

On the other hand, we wish to prove that

$$(58) \quad |\Psi_n(T_n)| \gtrsim \alpha_2^{C/N} \beta_1 \cdots \beta_n.$$

It will therefore suffice to find a subset T'_n of T_n with

$$(59) \quad |T'_n| \gtrsim \alpha_2^{C/N} \beta_1 \cdots \beta_n$$

such that we have the Jacobian bound

$$(60) \quad |\det(D\Psi_n)(t)| \gtrsim \alpha_2^{C/N} \quad \text{for all } t \in T'_n.$$

Indeed, if we can find such a T'_n , then we cover T'_n by balls of radius $\alpha_2^{C_1/N}$ for some large constant C_1 . If C_1 is large enough, then Ψ_n is one-to-one on every one of the balls that intersects T'_n , thanks to (60), the smoothness of Ψ_n (recall from Section 4 that the Ψ_n are smooth uniformly in x_0, δ_1, δ_2), and the inverse function theorem. By the pigeonhole principle one of these balls must contain at least $\alpha_2^{CC_1/N}$ of the set T'_n , and the claim (58) then follows from (60), (59) and the change-of-variables formula.

It remains to construct a set (59). This will be achieved by the width property of the $\tau_i(t)$ and several applications of Lemma 7.3, but first we need to control some derivative of $\det(D\Psi_n(t))$. We shall do so using the Hörmander condition assumption, by invoking the following quantitative analogue of Frobenius's theorem (which can also be derived from the arguments in [8]):

Lemma 9.1. *There exists a multi-index $\beta = (\beta_1, \dots, \beta_n)$ with $|\beta| \leq C(I_0, n)$ such that*

$$|\partial_t^\beta \det(D\Psi_n)|_{t=0} \gtrsim C_K^{-1}.$$

Proof. We shall prove a more general statement. For any $1 \leq i \leq n$, let J_i denote the i -form

$$J_i(t) := \partial_{t_1} \Psi_i(t) \wedge \cdots \wedge \partial_{t_i} \Psi_i(t)$$

defined for all t in a ball in \mathbf{R}^i . We shall show that for each i there exists a multi-index β^i with $|\beta^i| \leq C(I_0, n)$ and a small number $0 < \theta_i \ll 1$ depending on K such that

$$(61) \quad |\partial_t^{\beta^i} J_i|_{t=0}| \gtrsim \theta_i.$$

The lemma will then follow by applying (61) with $i = n$.

The idea of the proof can be sketched as follows. Informally, the statement (61) asserts that the vector fields X_1, \dots, X_i are non-degenerate, and so the iterated map Ψ_i fills out a large subset of a j -dimensional manifold. If X_{i+1} is transverse to this manifold, then one can induct and obtain (61) for $i + 1$ and similarly if X_{i+1} is tangent to the manifold to some (bounded) finite order. The only remaining case is if X_{i+1} is tangent to the manifold to some extremely high order. But this will imply that the iterated Lie brackets of X_{i+1} and X_i (for instance) are also tangent to the manifold, and this will eventually contradict the Hörmander condition assumption.

We now return to the rigorous proof of (61). For $i = 1$ this is clear just by taking $\beta = 0$ and using the non-vanishing of Y_1 . Now suppose that $i > 1$, and the claim has already been proven for $i - 1$.

Let $0 < \theta_i \ll 1$ be a small number depending on K , θ_{i-1} to be chosen later, let $A = A(I_0, n)$ be a large integer to be chosen later, and let $\epsilon = \epsilon(A)$ be a small number to be chosen later.

By induction hypothesis and the smoothness of J_{i-1} , we can find a ball $B_{i-1} \subset \mathbf{R}^{i-1}$ of radius $\gtrsim C_{\epsilon, \theta_{i-1}}^{-1}$ and at a distance $O(\epsilon \theta_{i-1})$ from the origin such that

$$(62) \quad |J_{i-1}(t)| \gtrsim \epsilon^C \theta_{i-1}^C \quad \text{for all } t \in B_{i-1}$$

(cf. the arguments in Lemma 7.3); the constant C can depend on β_{i-1} . By shrinking B_{i-1} if necessary we may then assume that Ψ_{i-1} is one-to-one on B_{i-1} . The set $\Psi_{i-1}(B_{i-1})$ is thus an open subset of a smooth $i - 1$ -dimensional manifold in B . Observe that the $i - 1$ -form $J_{i-1}(t)$ is tangent to $\Psi_{i-1}(B_{i-1})$ at $\Psi_{i-1}(t)$ for every $t \in B_{i-1}$.

Suppose for contradiction that we have

$$|\partial_t^\beta J_i|_{t=0}| \lesssim \theta_i$$

for all multi-indices β with $|\beta| \leq A$. Then by Taylor expansion we have

$$|J_i(t_1, \dots, t_{i-1}, 0)| \lesssim C_A \theta_i + C_A \epsilon^A \theta_{i-1}^A$$

whenever $t_1, \dots, t_{i-1} = O(\epsilon \theta_{i-1})$, and more generally that

$$|\partial_t^\beta J_i(t_1, \dots, t_{i-1}, 0)| \lesssim C_A \theta_i + C_A (\epsilon \theta_{i-1})^{A-|\beta|}$$

whenever $|\beta| < A$. On the other hand, by definition we have

$$\begin{aligned} J_i(t_1, \dots, t_{i-1}, 0) &= J_{i-1}(t_1, \dots, t_{i-1}) \wedge \partial_{t_i} \Psi_i(t_1, \dots, t_{i-1}, t_i)|_{t_i=0} \\ &= J_{i-1}(t_1, \dots, t_{i-1}) \wedge Y_i(\Psi_{i-1}(t_1, \dots, t_{i-1})). \end{aligned}$$

Because of this and (62), we see that for every $t \in B_{i-1}$ we can write

$$Y_i(\Psi_{i-1}(t)) = Y^{tangent}(\Psi_{i-1}(t)) + O(C_{A, \theta_{i-1}, \epsilon} \theta_i + C_A (\epsilon \theta_{i-1})^{A-C})$$

where $Y^{tangent}(\Psi_{i-1}(t))$ is tangent to $\Psi_{i-1}(B_{i-1})$ at $\Psi_{i-1}(t)$. Since $\Psi_{i-1}(B_{i-1})$ is a smooth manifold, we may thus write

$$Y_i = Y^{tangent} + Y^{error}$$

on a $C_{\theta_{i-1}}^{-1}$ -neighborhood of $\Psi_{i-1}(B_{i-1})$, where $Y^{tangent}$ is a smooth vector field which when restricted to $\Psi_{i-1}(B_{i-1})$ is tangent to $\Psi_{i-1}(B_{i-1})$, and Y^{error} is another smooth vector field such that

$$|\nabla^\beta Y^{error}| \lesssim C_{A,\theta_{i-1},\epsilon} \theta_i + C_A (\epsilon \theta_{i-1})^{A-|\beta|-C}$$

for all multi-indices $\beta = (\beta_1, \dots, \beta_n)$ with $|\beta| \ll A$.

From the identity

$$Y_{i-1}(\Psi_{i-1}(t)) = \partial_{t_{i-1}} \Psi_{i-1}(t)$$

we see that Y_{i-1} is also tangent to $\Psi_{i-1}(B_{i-1})$ when restricted to $\Psi_{i-1}(B_{i-1})$. Thus any Lie bracket combination of Y_{i-1} and $Y^{tangent}$ must also be tangent to $\Psi_{i-1}(B_{i-1})$ when restricted to $\Psi_{i-1}(B_{i-1})$.

Now let w be any word with $\deg(w)_1 + \deg(w)_2 \ll A$. The vector field Y_w is $K^{1-\deg(w)_1-\deg(w)_2}$ times a Lie bracket combination of Y_1 and Y_2 . One of these vector fields is Y_{i-1} and the other is Y_i , which can be split into $Y^{tangent}$ and Y^{error} . From the above discussion we thus see that Y_w when restricted to $\Psi_{i-1}(B_{i-1})$ is equal to a vector field tangent to $\Psi_{i-1}(B_{i-1})$, plus an error term of magnitude at most

$$C_{K,A,\theta_{i-1},\epsilon} \theta_i + C_{K,A} \epsilon \theta_{i-1}.$$

Applying this to Y_{w_1}, \dots, Y_{w_n} and taking wedge products, we obtain

$$|Y_{w_1}(\Psi_{i-1}(t)) \wedge \dots \wedge Y_{w_n}(\Psi_{i-1}(t))| \lesssim C_{K,A,\theta_{i-1},\epsilon} \theta_i + C_{K,A} \epsilon \theta_{i-1}.$$

If we pick ϵ sufficiently small depending on A and K and θ_i sufficiently small depending on $A, K, \theta_{i-1}, \epsilon$, we see that this bound contradicts (23), and we are done. \square

We can now quickly construct a set T'_n with the properties (59), (60), which will finish the proof of (56) and thus of Theorem 1.10.

By a Taylor expansion we may write

$$\det(D\Psi_n)(t) = P(t) + O(\alpha_2^N)$$

on $t \in [-r, r]^n$, where P is a polynomial of degree $O(N^2) = O(1)$, $\|P\|_\infty \lesssim 1$ and $|\partial_t^\alpha P(0)| \gtrsim C_K^{-1}$ (the index $\alpha = (\alpha_1, \dots, \alpha_n)$ with bounded coefficients is as in Lemma 9.1). It therefore suffices to construct T'_n satisfying (59) so that (60) holds with $P(t)$ in place of $\det(D\Psi_n)(t)$.

We apply Lemma 7.3 with S equal to $\tau_1(())$. This gives us a subset $\tau'_1(())$ of $\tau_1(())$ with measure

$$|\tau'_1(())| \gtrsim \alpha_2^{C/N} \beta_1$$

such that

$$|\partial_{t_2}^{\alpha_2} \dots \partial_{t_n}^{\alpha_n} P(t_1, 0, \dots, 0)| \gtrsim \alpha_2^{C/N}$$

for all $t_1 \in \tau'_1(())$.

Fix $t_1 \in \tau'_1(())$. We now apply Lemma 7.3 with S equal to $\tau_2(t_1)$. This gives us a subset $\tau'_2(t_1)$ of $\tau_2(t_1)$ with measure

$$|\tau'_2(t_1)| \gtrsim \alpha_2^{C/N} \beta_2$$

such that

$$|\partial_{t_3}^{\alpha_3} \dots \partial_{t_n}^{\alpha_n} P(t_1, t_2, 0, \dots, 0)| \gtrsim \alpha_2^{C/N}$$

for all $t_2 \in \tau_2'(t_1)$.

We continue in this manner, defining sets $\tau_i'(t_{i-1})$ recursively. Note that since our constants C are allowed to depend on n , we will not have trouble iterating the above scheme n times. If we recursively define $T'_0 = \{(\)\}$ and

$$T'_{i+1} := \{(t, t_{i+1}) \in \mathbf{R}^{i+1} : t \in T'_i; t_{i+1} \in \tau_i'(t)\},$$

we thus see that T'_n obeys (59) and (60) as desired. This completes the proof of Theorem 1.10.

10. REMARKS

- The estimate (51) can be rewritten using (6), and the fact that N was arbitrary, in the rather appealing form

$$(63) \quad |\Omega| \gtrsim |B(0; \alpha_1, \alpha_2)|.$$

In other words, if one fixes the thicknesses $\alpha_j = |\Omega|/|\pi_j(\Omega)|$ of a set Ω in a small neighborhood of 0, then the volume of the set Ω is essentially minimized when Ω is a two-parameter ball centered at the origin. One may thus think of our main result as a kind of two-parameter isoperimetric inequality in the spirit of [26].

- It would be nice if the \gtrsim in (63) could be replaced by a \gtrsim , as this would give restricted-type boundedness on the entire boundary of the Newton polytope P (bringing these results in line with those in, e.g., [6]). If one wishes to adapt the above argument to do this, there seem to be two major obstacles. The first is that the argument requires more and more regularity on $(\Sigma, M_1, M_2, \pi_1, \pi_2, 0_\Sigma)$ as one approaches the boundary of P . This particular difficulty can be avoided by restricting one's attention to model cases, such as when Σ is a nilpotent Lie group and X_1, X_2 are left-invariant vector fields. The other major difficulty is that one cannot refine the set of times (53) to a central set of a fixed width without losing an ε in the exponents. We do not know how to get around this loss.
- One can be more ambitious still and try to obtain strong-type boundedness at the endpoints of P . This is already difficult to do (at least if one only uses geometric combinatorics as in this paper) in such model cases as convolution with a compactly supported measure on the parabola $\{(t, t^2)\}$ in \mathbf{R}^2 . In this model case the endpoint mapping is $L^{3/2} \rightarrow L^3$, but a naive combinatorial argument only gives $L^{3/2,1} \rightarrow L^{3,\infty}$ (however see [30]; also see [6] for more discussion on this). Of course, one can obtain the strong-type endpoint in this case if one is willing to use such tools as complex interpolation and L^2 smoothing estimates (see, e.g., [25]). It may also be possible to do so by pure geometric combinatorics, but one probably has to control the extent to which various two-parameter balls of varying radii can overlap each other. In particular, a two-parameter covering lemma of some sort may be needed.
- It seems likely that one could extend these results (with the aid of the results in [8] or [12]) to obtain good $(L^{p_1}, L^{p'_2})$ mapping properties for fractional

integral operators such as

$$T_\alpha f(x) := \int f(\gamma(x, t)) \frac{dt}{|t|^\alpha}$$

for $0 < \alpha < 1$, where for each x , the map $t \mapsto \gamma(x, t)$ is a smooth parameterization of the curve γ_x of the type discussed in the introduction, and where the Hörmander condition is obeyed. In fact, if $\gamma(x, 0) \equiv x$ and $\alpha = 1 - \frac{1}{k}$ with k odd, then one easily sees (from a simple change of variables) that T_α falls within the class of operators described by (2) where the corresponding vector fields X_1 and X_2 are now linearly dependent at the origin.

- The methods here should probably be able to give an alternate way to obtain the L^p boundedness of maximal operators of the form

$$Mf(x) := \sup_{k \geq 0} 2^k \int f(\gamma(x, t)) a(2^k t) dt;$$

in particular, it is now quite possible that one could prove boundedness of such operators using purely geometric methods (avoiding the Fourier transform, as is used in, e.g., [8]). This is of course not the only type of maximal function one could study; more generally one could consider operators of the form

$$Mf(x) := \sup_r \int f(\gamma(x, t, r)) a(t) dt$$

for some parameter r ; model examples here include the circular maximal function and the Keakeya maximal function (see [52] for a survey of these operators). It seems difficult however to adapt our techniques to these operators except in very low dimensions; one would have to linearize the parameter r as $r(x)$, and this begins to destroy the differentiability properties of the iterated flow map Φ_n when $n > 3$. On the other hand, Christ and Erdogan have recently used geometric combinatorics ideas of a similar flavor to those in [6] and in this paper to obtain sharp mixed-norm estimates for certain classes of x-ray transforms; see [7]. However, the general question of understanding maximal operators of this type is still far beyond our current technology.

- It is also tempting to try to use these techniques to prove L^p smoothing estimates (i.e., mapping L^p to L^α_p rather than to L^q). However this problem is much more difficult due to the presence of cancellation and does not have a nice geometric interpretation such as (63). Indeed this problem is quite difficult even for model cases such as convolution with the curve (t, t^2, t^3) , as it is related to the local smoothing conjecture (see [34]). On the other hand, Wolff [53] has recently combined geometric combinatorics techniques with Fourier methods to obtain some progress on these types of problems. However, with respect to curves in the plane, Seeger [47], [46] has obtained sharp (up to endpoints) L^p to L^α_p estimates.
- Another possible generalization would be to higher-dimensional averages, or to asymmetric averages in which the manifolds M_1 and M_2 have different dimensions. In our language, this would mean that π_1 and π_2 now have corank k_1 and k_2 which are possibly greater than 1, and we would replace the single vector field X_j by a family of commuting vector fields

$X_j^1, \dots, X_j^{k_j}$. While it is possible to use this machinery (perhaps combined with the techniques in [8]) to get some non-trivial (L^p, L^q) mapping result, it seems difficult to obtain sharp results, because there seems to be no satisfactory analogue of the notion of width in more than one dimension. See Seeger [47] where non-trivial (L^p, L^q) as well as (L^p, L^p_α) results are obtained when Σ is a hypersurface in $M_1 \times M_2$ (i.e., $\dim \Sigma = k_1 + k_2 + 1$).

- It is crucial in our arguments that the sets $\tau_i(t)$ of times are central, so that the times t_i are always close to 0. This allows us to restrict the set $\phi_n(T_n)$ to lie inside a small two-parameter Carnot-Carathéodory ball, whose geometry can be well controlled by the machinery of Section 4. If we allowed $\tau_i(t)$ to wander far away from the origin, then it would in fact be impossible to obtain the estimate

$$(64) \quad |\Phi_n(T_n)| \gtrsim \alpha_1^{c_1} \alpha_2^{c_2}$$

for the desired values of c_1, c_2 , even in the model case when the $\tau_i(t)$ are all intervals of length α_i and we are considering convolution with (t, \dots, t^{n-1}) for some $n > 5$ (see [6] for some further discussion of this issue; this was also independently observed by Greenblatt). The point is that we lose control of the geometry if one flows too far along one or more vector fields. On the other hand, failure of the estimate (64) does not imply that failure of the lower bound on $|\Omega|$, because $\Phi_n(T_n)$ may only occupy a small portion of Ω . Our particular selection method for T_n (using the machinery of j -sheaves) is thus essential to ensure that $\Phi_n(T_n)$ does not degenerate to only a small fraction of Ω .

ACKNOWLEDGMENTS

We would like to thank Mike Christ, Allan Greenleaf, Andreas Seeger, and the referee for several useful discussions and comments. The authors also thank Mike Greenblatt for pointing out an error in the first version of this manuscript.

This work was conducted at UNSW. The first author is a Clay Prize Fellow and is supported by a grant from the Packard Foundation. The second author is supported in part by an ARC grant.

REFERENCES

- [1] J. Bourgain, *Averages in the plane over convex curves and maximal operators*, J. Analyse Math. **47** (1986), 69–85. MR **88f**:42036
- [2] J. Bourgain, *Besicovitch-type maximal operators and applications to Fourier analysis*, Geom. and Funct. Anal. **1** (1991), 147–187. MR **92g**:42010
- [3] A. Carbery, M. Christ, J. Wright, *Multidimensional van der Corput and Sublevel set estimates*, J. Amer. Math. Soc. **12** (1999), 981–1015. MR **2000h**:42010
- [4] M. Christ, *Weak type (1,1) bounds for rough operators*, Annals of Math. **128** (1988), 19–42. MR **89m**:42013
- [5] M. Christ, *Hilbert transforms along curves I: Nilpotent groups*, Annals of Math. **122** (1985), 575–596. MR **87f**:42039a
- [6] M. Christ, *Convolution, Curvature, and Combinatorics. A Case Study*, Internat. Math. Research Notices **19** (1998), 1033–1048. MR **2000a**:42026
- [7] M. Christ, M. B. Erdogan, *Mixed-norm estimates for a restricted x-ray transform*, preprint.
- [8] M. Christ, A. Nagel, E.M. Stein, S. Wainger, *Singular Radon transforms: Analysis and geometry*, Ann. of Math. (2) **150** (1999), 489–577. MR **2000j**:42023
- [9] M. Christ and J. L. Rubio de Francia, *Weak-type (1,1) bounds for rough operators II*, Invent. Math. **93** (1988): 225–237. MR **90d**:42021

- [10] I.M. Gelfand, M.I. Graev, *Line complexes in the space \mathbf{C}^n* , Func. Ann. Appl. **2** (1968), 219-229. MR **38**:6522
- [11] I.M. Gelfand, M.I. Graev, Z.J. Shapiro, *Differential forms and integral geometry*, Func. Ann. Appl. **3** (1969), 101-114. MR **39**:6232
- [12] M. Greenblatt, *A method for proving L^p -boundedness of singular Radon transforms in codimension 1*, Duke Math. J. **108** (2001), 363-393. MR **2002d**:42012
- [13] A. Greenleaf, A. Seeger, *Fourier integral operators with fold singularities*, J. Reine Angew. Math. **455** (1994), 35-56. MR **95h**:58130
- [14] A. Greenleaf, A. Seeger, *Fourier integral operators with cusp singularities*, Amer. J. Math. **120** (1998), 1077-1119. MR **99g**:58120
- [15] A. Greenleaf, A. Seeger, *Oscillatory and Fourier integral operators with degenerate canonical relations*, preprint.
- [16] A. Greenleaf, A. Seeger, S. Wainger, *Estimates for generalized Radon transforms in three and four dimensions*, Analysis, geometry, number theory: the mathematics of Leon Ehrenpreis (Philadelphia, PA, 1998), 243-254, Contemp. Math., 251. MR **2001j**:58047
- [17] A. Greenleaf, A. Seeger, S. Wainger, *On X-ray transforms for rigid line complexes and integrals over curves in \mathbf{R}^4* , Proc. Amer. Math. Soc. **127** (1999), 3533-3545. MR **2001a**:44002
- [18] A. Greenleaf, G. Uhlmann, *Nonlocal inversion formulas for the X-ray transform*, Duke Math. J. **58** (1989), 205-240. MR **91b**:58251
- [19] A. Greenleaf, G. Uhlmann *Composition of some singular Fourier integral operators and estimates for the X-ray transform, I*, Ann. Inst. Fourier (Grenoble) **40** (1990), 443-466. MR **91k**:58126
- [20] A. Greenleaf, G. Uhlmann *Composition of some singular Fourier integral operators and estimates for the X-ray transform, II*, Duke Math. J. **64** (1991), 413-419. MR **93b**:58146
- [21] V. Guillemin, S. Sternberg, *Geometric asymptotics*, Amer. Math. Soc. Surveys 14, Providence, RI, 1977. MR **58**:24404
- [22] V. Guillemin, *Cosmology in $(2 + 1)$ -dimensions, cyclic models and deformations of $M_{2,1}$* , Ann. of Math. Studies 121, Princeton Univ. Press, Princeton, 1989. MR **91k**:58140
- [23] S. Helgason, *A duality in integral geometry on symmetric spaces*, in Proc. U.S.-Japan Seminar in Differential Geometry, Kyoto 1965 (1966), 37-56. MR **37**:4765
- [24] L. Hormander, *Fourier integral operators*, Acta Math. **127** (1971), 79-183. MR **52**:9299
- [25] W. Littman, *$L^p - L^q$ estimates for singular integral operators*, Proc. Symp. Pure Math. **23**, American Mathematical Society, Providence RI, 1973. MR **50**:10909
- [26] L.H. Loomis, H. Whitney, *An inequality related to the isoperimetric inequality*, Bull. Amer. Math. Soc. **55** (1949), 961-962. MR **11**:166d
- [27] A. Nagel, E. Stein, S. Wainger, *Balls and metrics defined by vector fields I. Basic properties*, Acta. Math. **155** (1985), 103-147. MR **86k**:46049
- [28] J. D. McMichael, *Damping oscillatory integrals with polynomial phase*, Math. Scand. **73** (1993), 215-228. MR **95f**:42020
- [29] D. Oberlin, *Convolution estimates for some measures on curves*, Proc. Amer. Math. Soc. **99** (1987), 56-60. MR **88f**:42033
- [30] D. Oberlin, *Multilinear proofs for two theorems on circular averages*, Colloq. Math. **63** (1992), 187-190. MR **93m**:42005
- [31] D. Oberlin, *A convolution estimates for a measure on a curve in \mathbf{R}^4* , Proc. Amer. Math. Soc. **125** (1997), 1355-1361. MR **97g**:42009
- [32] D. Oberlin, *A convolution estimates for a measure on a curve in \mathbf{R}^4 II*, Proc. Amer. Math. Soc. **127** (1999), 217-221. MR **2000f**:42007
- [33] D. Oberlin, *An estimate for a restricted x-ray transform*, Canad. Math. Bull. **43** (2000), 472-476. MR **2001k**:44004
- [34] D. Oberlin, H. Smith, C. Sogge, *Averages over curves with torsion*, Math Res. Lett. **5** (1998), 535-539. MR **99j**:42021
- [35] I. Laba, T. Tao, *An x-ray estimate in \mathbf{R}^n* , to appear, Revista Iberoamericana.
- [36] D. Phong, *Regularity of Fourier integral operators*, Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994), 862-874, Birkhuser, Basel, 1995. MR **97e**:58216
- [37] D. Phong, *Singular integrals and Fourier integral operators*, Essays on Fourier analysis in honor of Elias M. Stein (Princeton, NJ, 1991), 286-320, Princeton Math. Ser., 42, Princeton Univ. Press, Princeton, NJ, 1995. MR **95k**:42026

- [38] D. Phong, E.M. Stein, *Hilbert transforms, singular integrals, and Radon transforms I*, Acta Math. **157** (1986), 99–157. MR **88i**:42028a
- [39] D. Phong, E.M. Stein, *Radon transforms and torsion*, IMRN **1991**, 49–60. MR **93g**:58144
- [40] D. Phong, E.M. Stein, *On a stopping process for oscillatory integrals*, J. Geom. Anal. **4** (1994), 105–120. MR **95k**:42025
- [41] D. Phong, E.M. Stein, *Models of degenerate Fourier integral operators and Radon transforms*, Ann. of Math. (2) **140** (1994), 703–722. MR **96c**:35206
- [42] D. Phong, E.M. Stein, *The Newton polyhedron and oscillatory integral operators*, Acta Math. **179** (1997), 105–152. MR **98j**:42009
- [43] F. Ricci, E. M. Stein, *Harmonic analysis on nilpotent groups and singular integrals II: Singular kernels supported on submanifolds*, J. Funct. Anal. **78**:56–84. MR **89g**:42030
- [44] W. Schlag, *A generalization of Bourgain’s circular maximal theorem*, J. Amer. Math. Soc. **10** (1997), 103–122. MR **97c**:42035
- [45] S. Secco, *L^p improving properties of measures supported on curves on the Heisenberg group*, Studia Math. **132** (1999), 179–201. MR **2000c**:43010
- [46] A. Seeger, *Degenerate Fourier integral operators in the plane*, Duke Math. J. **71** (1993), 685–745. MR **94h**:35292
- [47] A. Seeger, *Radon transforms and finite type conditions*, J. Amer. Math. Soc. **11** (1998), 869–897. MR **99f**:58202
- [48] E. M. Stein, *Oscillatory integrals related to Radon-like transforms*, Proceedings of the Conference in Honor of Jean-Pierre Kahane (Orsay, 1993). J. Fourier Anal. Appl. 1995, Special Issue, 535–551. MR **96i**:42013
- [49] T. H. Wolff, *An improved bound for Keakeya type maximal functions*, Revista Mat. Iberoamericana. **11** (1995), 651–674. MR **96m**:42034
- [50] T. H. Wolff, *A Keakeya type problem for circles*, Amer. J. Math. **119** (1997), 985–1026. MR **98m**:42027
- [51] T. H. Wolff, *A mixed norm estimate for the x-ray transform*, Revista Mat. Iberoamericana. **14** (1998), 561–600. MR **2000j**:44006
- [52] T. Wolff, *Recent work connected with the Keakeya problem*, Prospects in mathematics (Princeton, NJ, 1996), 129–162, Amer. Math. Soc., Providence, RI, 1999. MR **2000d**:42010
- [53] T. Wolff, *Local smoothing estimates in L^p for large p* , Geom. Funct. Anal. **10** (2000), 1237–1288. MR **2001k**:42030

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, LOS ANGELES, LOS ANGELES, CALIFORNIA 90095-1555

E-mail address: tao@math.ucla.edu

SCHOOL OF MATHEMATICS, UNIVERSITY OF EDINBURGH, JCMB, KING’S BUILDINGS, MAYFIELD ROAD, EDINBURGH EH9 3JZ, SCOTLAND

E-mail address: wright@maths.ed.ac.uk