ENDPOINT LEBESGUE ESTIMATES FOR WEIGHTED AVERAGES ON POLYNOMIAL CURVES

MICHAEL CHRIST, SPYRIDON DENDRINOS, BETSY STOVALL, AND BRIAN STREET

ABSTRACT. We establish optimal Lebesgue estimates for a class of generalized Radon transforms defined by averaging functions along polynomial-like curves. The presence of an essentially optimal weight allows us to prove uniform estimates, wherein the Lebesgue exponents are completely independent of the curves and the operator norms depend only on the polynomial degree. Moreover, our weighted estimates possess rather strong diffeomorphism invariance properties, allowing us to obtain uniform bounds for averages on curves satisfying natural nilpotency and nonoscillation hypotheses.

1. INTRODUCTION

Let (P_1, g_1) and (P_2, g_2) be two smooth Riemannian manifolds of dimension $n-1$, with $n \geq 2$. In [27], Tao–Wright established near-optimal Lebesgue estimates for local averaging operators of the form

$$Tf(x_2) := \int f(\gamma_{x_2}(t)) a(x_2, t) |\gamma'_{x_2}(t)|_{g_1} dt, \qquad f \in C^0(P_1), \qquad (1.1)$$

with a continuous and compactly supported, under the hypothesis that the map $(x_2, t) \mapsto \gamma_{x_2}(t) \in P_1$ is a smooth submersion on the support of a.

Our goal in this article is to sharpen the Tao–Wright theorem to obtain optimal Lebesgue space estimates, without the cutoff, under an additional polynomial-like hypothesis on the map γ. We replace the Riemannian arclength with a natural generalization of affine arclength measure; this enables us to prove estimates wherein the Lebesgue exponents are independent of the manifolds and curves involved (provided γ is polynomial-like), and operator norms for a fixed exponent pair and fixed polynomial degree are uniformly bounded. Our results are strongest at the Lebesgue endpoints, where the generalized affine arclength measure is essentially the largest measure for which these estimates can hold and, moreover, the resulting inequalities are invariant under a variety of coordinate changes.

By duality, bounding the operator T in (1.1) is equivalent to bounding the bilinear form

$$\mathcal{B}(f_1, f_2) := \int_M f_1(\gamma_{x_2}(t)) f_2(x_2) a(x_2, t) |\gamma'_{x_2}(t)|_{g_1} d\nu_2(x_2) dt, \qquad (1.2)$$

where $M := P_2 \times \mathbb{R}$. For the remainder of the article, we will focus on the problem of bounding such bilinear forms.

1.1. **The Euclidean case.** The Tao–Wright theorem, being local, may be equivalently stated in Euclidean coordinates. Though we will obtain more general results on manifolds (and also in Euclidean space) by applying diffeomorphism invariance

of our operator and basic results from Lie group theory, the Euclidean version is, in some sense, our main theorem.

Let $\pi_1, \pi_2 : \mathbb{R}^n \to \mathbb{R}^{n-1}$ be smooth mappings. Define vector fields

$$X_j := \star(d\pi_j^1 \wedge \cdots \wedge d\pi_j^{n-1}), \qquad (1.3)$$

where $\star$ denotes the map from $n-1$ forms to vector fields obtained by composing the Riemannian Hodge star with the natural identification of 1 forms with vector fields given by the Euclidean metric. The geometric significance of the X_j is that they are tangent to the fibers of the π_j, and their magnitude arises in the coarea formula:

$$|\Omega| = \int_{\pi_j(\Omega)} \int_{\pi_j^{-1}(y)} \chi_\Omega(t) |X_j(t)|^{-1} \, d\mathcal{H}^1(t) \, dy, \qquad \Omega \subseteq \{X_j \neq 0\}, \qquad (1.4)$$

where $\mathcal{H}^1$ denotes 1-dimensional Hausdorff measure.

We define a map $\Psi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ by

$$\Psi_x(t) := e^{t_n X_n} \circ \cdots \circ e^{t_1 X_1}(x), \qquad (1.5)$$

where we are using the cyclic notation $X_j := X_{j \bmod 2}$, $j = 3, \ldots, n$. Given a multiindex β, we define

$$b = b(\beta) := \Big(\sum_{j \text{ odd}} 1 + \beta_j, \ \sum_{j \text{ even}} 1 + \beta_j \Big) \qquad (1.6)$$

$$\rho_\beta(x) := \big| (\partial_t^\beta \det D_t \Psi_x)(0) \big|^{\frac{1}{b_1+b_2-1}} \qquad (1.7)$$

$$(p_1, p_2) = (p_1(b), p_2(b)) := \Big(\tfrac{b_1+b_2-1}{b_1}, \tfrac{b_1+b_2-1}{b_2} \Big). \qquad (1.8)$$

Our main theorem is the following.

Theorem 1.1. *Let $n \geq 3$, let N be a positive integer, and let β be a multiindex. Assume that the maps π_j and associated vector fields X_j, defined in (1.3) satisfy the following:*
(i) The X_j generate a nilpotent Lie algebra $\mathfrak{g}$ of step at most N, and for each $X \in \mathfrak{g}$, the map $(t, x) \mapsto e^{tX}(x)$ is a polynomial of degree at most N;
(ii) For each $j = 1, 2$ and a.e. $y \in \mathbb{R}^{n-1}$, $\pi_j^{-1}(\{y\})$ is contained in a single integral curve of X_j.
Then with ρ_β satisfying (1.7) and p_1, p_2 as in (1.8),

$$\Big| \int_{\mathbb{R}^n} f_1 \circ \pi_1(x) \, f_2 \circ \pi_2(x) \, \rho_\beta(x) \, dx \Big| \leq C_N \|f_1\|_{p_1} \|f_2\|_{p_2}, \qquad (1.9)$$

for some constant C_N depending only on the degree N.

No explicit nondegeneracy (i.e. finite type) hypothesis is needed, because the weight ρ_β is identically zero in the degenerate case.

The weights ρ_β were introduced in [25], wherein local, non-endpoint Lebesgue estimates were proved in the C^∞ case for a multilinear generalization. In Section 10, we give examples showing that the endpoint estimate (1.9) may fail in the multilinear case, and that it may also fail in the bilinear case when Hypothesis (i), Hypothesis (ii), or the dimensional restriction $n \geq 3$ is omitted.

Theorem 1.1 uniformizes, makes global, and sharpens to Lebesgue endpoints the Tao–Wright theorem for averages along curves, under our additional hypotheses. (As the Tao–Wright theorem is stated in terms of the spanning of elements from $\mathfrak{g}$, not the non-vanishing of ρ, the relationship between the results will take some

explanation, which will be given in Section 4.) Moreover, our result generalizes to the fully translation non-invariant case the results of [8, 10, 15, 21, 23], wherein endpoint Lebesgue estimates were established for convolution and restricted X-ray transforms along polynomial curves with affine arclength measure.

1.2. **Averages on curves in manifolds and other generalizations.** As our results are global and uniform, it is natural to ask whether they lead to global results in the more general setting described at the outset, wherein operators are defined for functions on manifolds. This is the content of our Theorem 9.3. Roughly speaking, this theorem allows one to compute the X_j and ρ_β in local coordinates and removes the polynomial hypothesis in (i) of Theorem 1.1. We leave the precise statement for later because it requires some additional terminology.

Another natural question is the extent to which one can relax hypothesis (i). In this article, we prove a local result (Proposition 9.1) for the mild generalization considered in [13], wherein it is only assumed that there exist vector fields tangent to the X_j generating a nilpotent Lie algebra. A number of counter-examples to other possible generalizations are given in Section 10.

1.3. **Background and sketch of proof.** We turn to an outline of the proof of Theorem 1.1, and a discussion of the context in the recent literature.

We begin with the proof on a single torsion scale $\{\rho_\beta \sim 2^m\}$. By uniformity, it suffices to consider the case when $m = 0$, and thus the restricted weak type version of (1.9) is equivalent to the generalized isoperimetric inequality

$$|\Omega| \lesssim |\pi_1(\Omega)|^{\frac{1}{p_1}} |\pi_2(\Omega)|^{\frac{1}{p_2}}, \qquad \Omega \subseteq \{\rho_\beta \sim 1\}. \tag{1.10}$$

With b and p as in (1.6) and (1.8), $b = \left(\frac{1/p_1}{1/p_1 + 1/p_2 - 1}, \frac{1/p_2}{1/p_1 + 1/p_2 - 1}\right)$, so (1.10) is, after a bit of arithmetic, equivalent to the lower bound

$$\alpha_1^{b_1} \alpha_2^{b_2} \lesssim |\Omega|, \qquad \alpha_j := \frac{|\Omega|}{|\pi_j(\Omega)|}. \tag{1.11}$$

To establish (1.11), Tao–Wright [27], and later Gressman [13], used a version of the iterative approach from [3]. Roughly speaking, for a typical point $x_0 \in \Omega$, the measure of the set of times t such that $e^{tX_j}(x_0) \in \Omega$ is α_j. Iteratively flowing along the vector fields X_1, X_2 gives a smooth map, Ψ_{x_0} (recall (1.5)), from a measurable subset $F \subseteq \mathbb{R}^n$ into Ω. The containment $\Psi_{x_0}(F) \subseteq \Omega$ must then be translated into a lower bound on the volume of Ω.

Tao–Wright deduce from linear independence of a fixed n-tuple $Y_1, \ldots, Y_n \in \mathfrak{g}$ (the Lie algebra generated by X_1, X_2) a lower bound on some fixed derivative ∂^β of the Jacobian determinant $\det D\Psi_{x_0}$. For typical points $t \in \mathbb{R}^n$, we have a lower bound $|\det D\Psi_{x_0}(t)| \gtrsim |t^\beta||\partial^\beta \det D\Psi_{x_0}(0)|$, and this we should be able to use in estimating the volume of Ω:

$$|\Omega| \geq |\Psi_{x_0}(F)| \text{ '}\gtrsim\text{' } \int_F |\det D\Psi_{x_0}(t)| \, dt \text{ '}\gtrsim\text{' } |F||\partial^\beta \det D\Psi_{x_0}(0)| \max_{t \in F} |t^\beta| \gtrsim \alpha^b.$$

Unfortunately, the failure of Ψ_{x_0} to be polynomial in the Tao–Wright case and the fact that F is not simply a product of intervals means that this deduction is not so straightforward; in particular, the inequalities surrounded by quotes in the preceding inequality are false in the general case. More precisely, if Ψ_{x_0} is merely C^∞, we cannot uniformly bound the number of preimages in F of a typical point in Ω (so the first inequality may fail), and even for polynomial Ψ_{x_0}, if F is not an

axis parallel rectangle, then the inequality $|\det D\Psi_{x_0}(t)| \gtrsim |t^\beta|$ may fail for most $t \in F$.

In the nonendpoint case of [27], it is enough to prove (1.11) with a slightly larger power of α on the left; this facilitates an approximation of F by a small, axis parallel rectangle centered at 0, and (using the approximation of F) an approximation of Ψ_{x_0} by a polynomial. These approximations are sufficiently strong that Ψ_{x_0} is nearly finite-to-one on F (see also [5]) and $\det D\Psi_{x_0}$ grows essentially as fast on F as its derivative predicts, giving (1.11). In [13], wherein the Lie algebra $\mathfrak{g}$ is assumed to be nilpotent, the map Ψ_{x_0} is lifted to a polynomial map in a higher dimensional space, abrogating the need for the polynomial approximation. This leaves the challenge of producing a suitable approximation of F as a product of intervals, and Gressman takes a different approach from Tao–Wright, which avoids the secondary endpoint loss.

In Section 2, we reprove Gressman's single scale restricted weak type inequality. A crucial step is an alternate approach to approximating one-dimensional sets by intervals. This alternative approach gives us somewhat better lower bounds for the integrals of polynomials on these sets, and these improved bounds will be useful later on.

An advantage of the positive, iterative approach to bounding generalized Radon transforms has been its flexibility, particularly relative to the much more limited exponent range that seems to be amenable to Fourier transform methods. A disadvantage of this approach is that it seems best suited to proving restricted weak type, not strong type estimates. Let us examine the strong type estimate on torsion scale 1. By positivity of our bilinear form, it suffices to prove

$$\sum_{j,k} 2^{j+k} \int_{\{\rho_\beta \sim 1\}} \chi_{E_1^j} \circ \pi_1(x) \chi_{E_2^k} \circ \pi_2(x)\, dx \lesssim \Big(\sum_j 2^{jp_1} |E_1^j|\Big)^{\frac{1}{p_1}} \Big(\sum_k 2^{kp_2} |E_2^k|\Big)^{\frac{1}{p_2}},$$

for measurable sets $E_1^j, E_2^k \subseteq \mathbb{R}^{n-1}$, $j, k \in \mathbb{Z}$. Thus a scenario in which we might expect the strong type inequality to fail is when there is some large set $\mathcal{J}$ and some set $\mathcal{K}$ such that the $2^j \chi_{E_1^j}$, $j \in \mathcal{J}$, evenly share the L^{p_1} norm of f_1, the $2^k \chi_{E_2^k}$, $k \in \mathcal{K}$ evenly share the L^{p_2} norm of f_2, and the restricted weak type inequality is essentially an equality

$$\int_{\{\rho_\beta \sim 1\}} \chi_{E_1^j} \circ \pi_1(x) \chi_{E_2^k} \circ \pi_2(x)\, dx \sim |E_1^j|^{\frac{1}{p_1}} |E_2^k|^{\frac{1}{p_2}}, \tag{1.12}$$

for each $(j, k) \in \mathcal{J} \times \mathcal{K}$.

In [4] a technique was developed for proving strong type inequalities by defeating such enemies, and this approach was used to reprove Littman's bound [16] for convolution with affine surface measure on the paraboloid. This approach was later used [8, 10, 11, 15, 21, 23] to prove optimal Lebesgue estimates for translation invariant and semi-invariant averages on various classes of curves with affine arclength measure. Key to these arguments was what was called a 'trilinear' estimate in [4], which we now describe. We lose if one E_2^k interacts strongly, in the sense of (1.12) with many sets E_1^j of widely disparate sizes. Suppose that E_2^k interacts strongly with two sets $E_1^{j_i}$, $i = 1, 2$. Letting

$$\Omega_i := \pi_1^{-1}(E_1^{j_i}) \cap \pi_2^{-1}(E_2^k) \cap \{\rho_\beta \sim 1\},$$

our hypothesis (1.12) and the restricted weak type inequality imply that $\pi_2(\Omega_i)$ must have large intersection with E_2^k for $i = 1, 2$; let us suppose that $E_2^k = \pi_2(\Omega_1) = \pi_2(\Omega_2)$. Assuming that every π_2 fiber is contained in a single X_2 integral curve, for a typical $x_0 \in \Omega_i$, the set of times t such that $e^{tX_2}(x_0) \in \Omega_{i'}$ must have measure about $\alpha_2^{i'} := \frac{|\Omega_{i'}|}{|E_2^k|}$; thus we have $\Psi_{x_0}(F_i) \subseteq \Omega_i$ for measurable sets F_i, which are not well-approximated by products of intervals centered at 0. In all of the above mentioned articles [4, 8, 10, 11, 15, 21, 23], rather strong pointwise bounds on the Jacobian determinant $\det D\Psi_{x_0}$ were then used to derive mutually incompatible inequalities relating the volumes of the three sets, $E_1^{j_1}, E_1^{j_2}, E_2^k$ (whence the descriptor 'trilinear'). In generalizing this approach, we encounter a number of difficulties. First, we lack explicit lower bounds on the Jacobian determinant. We can try to recover these using our estimate $1 \sim |\partial^\beta \det D\Psi_{x_0}(0)|$, but this is difficult to employ on the sets F_i, since it is impossible to approximate these sets using products of intervals centered at 0. Finally, in the translation invariant case, it is natural to decompose the bilinear form in time,

$$\mathcal{B}(f_1, f_2) = \sum_j \int_{\mathbb{R}^{n-1}} \int_{t \in I_j} f_1(x - \gamma(t)) f_2(x) \, \rho_\beta(t) \, dt \, dx,$$

and, thanks to the geometric inequality of [9], there is a natural choice of intervals I_j that makes the trilinear enemies defeatable. It is not clear to the authors that an analogue of this decomposition in the general polynomial-like case is feasible.

Our solution is to dispense entirely with the pointwise approach. In Section 5, we prove that if the set Ω nearly saturates the restricted weak type inequality (1.10), then Ω can be very well approximated by Carnot–Carathéodory balls. Thus, if E_1 and E_2 interact strongly, then E_1 and E_2 can be well-approximated by projections (via π_1, π_2) of Carnot–Carathéodory balls. The proof of this inverse result relies on the improved polynomial approximation mentioned above, as well as new information, proved in Section 3, on the structure of Carnot–Carathéodory balls generated by nilpotent families of vector fields. In Section 6, we prove that it is not possible for a large number of Carnot–Carathéodory balls with widely disparate parameters to have essentially the same projection; thus one set E_2^k cannot interact strongly with many E_1^j, and so the strong type bounds on a single torsion scale hold. In Section 7, we sum up the torsion scales. In the non-endpoint case considered in [25], this was simply a matter of summing a geometric series, but here we must control the interaction between torsion scales. The crux of our argument is that many Carnot–Carathéodory balls at different torsion scales cannot have essentially the same projection.

Section 8 gives relevant background on nilpotent Lie groups which will be used in deducing from Theorem 1.1 more general results, including the above-mentioned result on manifolds. The results of this section are essentially routine deductions from known results in the theory of nilpotent Lie groups, but the authors could not find elsewhere the precise formulations needed here. In Section 9, we prove extensions of our result to the nilpotent case and other generalizations. In Section 10, we give counter-examples to a few "natural" generalizations of our main theorem, discuss its optimality at Lebesgue endpoints, and recall the impossibility of an optimal weight away from Lebesgue endpoints. The appendix, Section 11, contains various useful lemmas on polynomials of one and several variables. Some of these results are new and may be useful elsewhere.

Notation. Constants are allowed to depend on N and may change from line-to-line. Constants may depend on those that come logically before. Thus constants in conclusions depend on those arising in proofs (or in lemmas used in the proofs), which in turn depend on N and the constants in hypotheses. Further subscripts will be used to denote other parameters on which constants depend. Capital letters (usually C) will typically be used to denote large constants and lower case letters (usually c) to denote small ones.

We will use the now-standard $\lesssim, \gtrsim, \sim$ and the non-standard $\lessapprox, \gtrapprox, \approx$. We describe their use using two nonnegative quantities A and B. When found in the hypothesis of a statement, $A \lesssim B$ means that the conclusion holds whenever $A \leq CB$ for any C (with constants in the conclusion allowed to depend on C). In the conclusion, $A \lesssim B$ means that $A \leq CB$ for some C. Later on, we will introduce a small parameter $0 < \varepsilon \leq 1$, and many quantities depend on ε in some way as well. We will use $A \lessapprox B$ to mean that $A \leq C\varepsilon^{-C}B$ for C quantified in the same way as the implicit constant in the $\lesssim$ notation. (In Section 7, this notation will depend instead on a small parameter $\delta > 0$.) Finally $A \sim B$ means $A \lesssim B$ and $B \lesssim A$, and $A \approx B$ means $A \lessapprox B$ and $B \lessapprox A$. We will occasionally subscript these symbols to indicate their dependence on parameters other than N.

2. The restricted weak type inequality on a single scale

This section is devoted to a proof, or, more accurately, a reproof, of the restricted weak type inequality on the region where $\rho_\beta \sim 1$. The following result is due to Gressman in [13]. (Uniformity is not explicitly claimed in [13], but the arguments therein may easily be adapted.)

Proposition 2.1. [13] *For each pair $E_1, E_2 \subseteq \mathbb{R}^{n-1}$ of measurable sets,*

$$|\{\rho_\beta \sim 1\} \cap \pi_1^{-1}(E_1) \cap \pi_2^{-1}(E_2)| \lesssim |E_1|^{1/p_1}|E_2|^{1/p_2} \tag{2.1}$$

holds uniformly, with definitions and hypotheses as in Theorem 1.1.

We give a complete proof of the preceding, using partially alternative methods from those in [13], because our approach will facilitate a resolution, in Section 5, of a related inverse problem, namely, to characterize those pairs (E_1, E_2) for which the inequality in (2.1) is reversed. Our proof of Proposition 2.1 is based on the following proposition.

Proposition 2.2. *Let $S \subseteq \mathbb{R}$ be a measurable set. For each N, there exists an interval $J = J(N, S)$ with $|J \cap S| \gtrsim |S|$ such that for any polynomial P of degree*

at most N,

$$\int_S |P|\, dt \gtrsim \sum_{j=0}^{N} \|P^{(j)}\|_{L^\infty(J)} \left(\tfrac{|J|}{|S|}\right)^{(1-\varepsilon)j} |S|^{j+1}. \tag{2.2}$$

The key improvement of this lemma over the analogous result in [13] is the gain $\left(\tfrac{|J|}{|S|}\right)^{(1-\varepsilon)j}$ in the higher order terms. This gain will allow us to transfer control over $\int_S |P|$ into control over the length of J.

Proof of Proposition 2.2. If S has infinite measure, the left hand side of (2.2) is infinite whenever it is nonzero. Thus we may assume that S has finite measure. Replacing S by a bounded subset with comparable measure, we may assume that $S \subseteq I$ for some finite interval I. Now we turn to a better approximation.

Lemma 2.3. *Given $c > 0$, there exist intervals $J, K \subseteq I$ with the following properties.*

  i. $|J| \sim |K| \sim \mathrm{dist}(K, J)$
  ii. $|S \cap J| \gtrsim |S|$
  iii. $|S \cap K| \gtrsim \left(\tfrac{|S|}{|K|}\right)^c |S|.$

Proof. Let $c' > 0$ be a small constant, to be determined.

Starting from $i = 0$ and $I_0 = I$, we use the following stopping time procedure.

Let $m_i := \lceil \log_{\frac{4}{3}}(\tfrac{|I_i|}{|S|}) \rceil$. Divide $I_i = I_i^1 \cup I_i^2 \cup I_i^3 \cup I_i^4$ into four non-overlapping intervals of equal length, arranged in order of increasing index.

If

$$|S \cap I_i^j| > c' 2^{-cm_i} |S \cap I_i|,$$

for $j = 1$ and $j = 4$, then stop. Set $J = I_i^j$, where j is chosen to maximize $|S \cap I_i^j|$ and set $K = I_i^k$, where $k \in \{1, 4\}$ is not adjacent to j. Then we are done, provided $|S \cap I_i| \gtrsim |S|$.

If (say) $|S \cap I_i^1| \le c' 2^{-cm_i} |S \cap I_i|$ (the case where $|S \cap I_i^4| \le c' 2^{-cm_i} |S \cap I_i|$ being handled analogously), discard I_i^1 and repeat the procedure on $I_{i+1} := I_i^2 \cup I_i^3 \cup I_i^4$. Note that $m_{i+1} = m_i - 1$.

On the one hand, $|I_i| = 2(\tfrac{3}{4})^i |I_0|$ tends to zero as $i \to \infty$, while on the other hand,

$$|I_i| \ge |S \cap I_i| \ge \Big[\prod_{j=0}^{i-1}(1 - c' 2^{-m_j c})\Big]|S| \gtrsim |S|,$$

where the last inequality is valid for c' sufficiently small. Thus the process terminates after finitely many steps. $\qquad\square$

We apply Lemma 2.3 iteratively, N times, to obtain a sequence of pairs of bounded intervals $K_1, J_1 \subseteq I$, $K_{i+1}, J_{i+1} \subseteq J_i$, $1 \le i \le N - 1$, satisfying

$$|K_i| = |J_i| = \mathrm{dist}(K_i, J_i)$$
$$|S \cap J_i| \gtrsim |S|$$
$$|S \cap K_i| \gtrsim \left(\tfrac{|S|}{|K_i|}\right)^c |S|.$$

Let $m_i := \log_2\left(\tfrac{|K_i|}{|S|}\right)$. We observe that $m_1 \ge m_2 \ge \cdots \ge m_N$.

It remains to prove that if P is any degree N polynomial,

$$\int_S |P| \gtrsim \sum_{j=0}^{N} \|P^{(j)}\|_{L^\infty(J_N)} 2^{j(1-c)m_N} |S|^{j+1}. \tag{2.3}$$

We will repeatedly use, without comment, the equivalence of all norms on the finite dimensional vector space of polynomials of degree at most N. (Examples of norms that we use are $\|P\|_{L^\infty([0,1])}$, $\sum_j |P^{(j)}(\zeta_0)|$ for a fixed $\zeta_0 \in \{|\zeta| < 1\}$, $\|P\|_{L^1([0,1])}$, $\|P\|_{L^\infty([\frac{2}{3},1])}$, $\|P\|_{L^\infty(\{|z|<1\})}$, etc.) By scaling and translation, we can map $[0,1]$ onto any closed interval, and the norms transform accordingly.

Multiplying P by a constant if needed, we may write $P(t) = \prod_{j=1}^{N}(z - \zeta_j)$, where the ζ_j are the complex zeros, counted according to multiplicity.

First, suppose that $\text{dist}(\zeta_j, J_N) \geq \frac{1}{100}|J_N|$ for all j. Then $|P(t)| \sim |P(t_0)|$ throughout J_N, so

$$\int_S |P| \geq \int_{S\cap J_N} |P| \gtrsim \|P\|_{L^\infty(J_N)}|S| \sim \sum_{j=0}^{N} |P^{(j)}(t_0)||J_N|^j |S|,$$

which dominates the right side of (2.3).

Now suppose that $\text{dist}(\zeta_1, J_N) < \frac{1}{100}|J_N|$. We have that

$$\|P\|_{L^\infty(J_N)} \sim \sum_{j=0}^{N} |P^{(j)}(\zeta_1)||J_N|^j = |J_N| \sum_{j=0}^{N-1} |(P')^{(j)}(\zeta_1)||J_N|^j$$

$$\sim |J_N| \|P'\|_{L^\infty(J_N)} \sim \sum_{j=1}^{N} \|P^{(j)}\|_{L^\infty(J_N)}|J_N|^j.$$

By construction, for each $j \geq 2$, $\text{dist}(\zeta_j, K_i) < \frac{1}{100}|K_i|$ can hold for at most one value of i. Thus there exists $1 \leq i \leq N$ such that $\text{dist}(\zeta_j, K_i) \geq \frac{1}{100}|K_i|$ for all j, so $|P(t)| \sim |P(t_i)|$, for any $t, t_i \in K_i$. Therefore

$$\int_S |P| \gtrsim \int_{S\cap K_i} |P| \sim \|P\|_{L^\infty(K_i)}|S \cap K_i| \sim \|P\|_{L^\infty(J_i)}|S \cap K_i|$$

$$\gtrsim \sum_{j=1}^{N} \|P^{(j)}\|_{L^\infty(J_i)}|J_i|^j |S \cap K_i| \gtrsim \sum_{j=1}^{N} \|P^{(j)}\|_{L^\infty(J_N)} 2^{(j-c)m_i}|S|^{j+1}, \tag{2.4}$$

which is again larger than the right side of (2.3). $\square$

Proof of Proposition 2.1. We may assume that E_1, E_2 are open sets. We take the now-standard approach of iteratively refining the set

$$\pi_1^{-1}(E_1) \cap \pi_2^{-1}(E_2) \cap \{\rho_\beta \sim 1\}.$$

Since $X_j \neq 0$ a.e. on $\{\rho_\beta \sim 1\}$, π_j is a submersion a.e. on $\{\rho_\beta \sim 1\}$. By the implicit function theorem and hypothesis (ii) of Theorem 1.1, points x, x' at which π_j is a submersion that lie on distinct X_j integral curves cannot have $\pi_j(x) = \pi_j(x')$. Thus there exists an open set $\Omega \subseteq \mathbb{R}^n$ with

$$\Omega \subseteq \pi_1^{-1}(E_1) \cap \pi_2^{-1}(E_2) \cap \{\rho_\beta \sim 1\}, \qquad |\Omega| \sim |\pi_1^{-1}(E_1) \cap \pi_2^{-1}(E_2) \cap \{\rho_\beta \sim 1\}|,$$

such that for each $y \in \mathbb{R}^{n-1}$, $\pi_j^{-1}(\{y\}) \cap \Omega$ is contained in a single integral curve of X_j.

Define
$$\alpha_j := \frac{|\Omega|}{|E_j|}, \qquad j = 1, 2.$$

Our goal is to prove that
$$|\Omega| \gtrsim \alpha_1^{b_1} \alpha_2^{b_2},$$

where $b = b(p)$ is as in (1.6); after some arithmetic, this implies (2.1). We may thus assume that Ω is a bounded set.

We may write the coarea formula as
$$|\Omega'| = \int_{\pi_j(\Omega')} \int \chi_{\Omega'}(e^{tX_j}(\sigma_j(y))) \, dt \, dy, \qquad \Omega' \subseteq \Omega,$$

and we use this formula to refine iteratively, starting with $j = n$ and $\Omega_n := \Omega$. For $x \in \Omega_j$, we define
$$S_j(x) := \{t : e^{tX_j}(x) \in \Omega_j\}.$$

Let $\sigma_j : \pi_j(\Omega_j) \to \Omega_j$ be a measurable section of π_j, with further properties to be determined later. If we define $t_j(x) \in \mathbb{R}$ by the formula $x =: e^{t_j(x)X_j}(\sigma_j(\pi_j(x)))$, we see that $S_j(x) + t_j(x) = S_j(\sigma_j(\pi_j(x)))$; in particular, both sides depend only on $\pi_j(x)$. We further define
$$J_j(x) := J(N, S_j(x) + t_j(x)) - t_j(x),$$

where $J(N, S)$ is the interval whose existence was guaranteed in Proposition 2.2. We choose this somewhat cumbersome definition so that $J_j(x) + t_j(x)$ depends only on $\pi_j(x)$ and $|S_j(x) \cap J_j(x)| \gtrsim |S_j(x)|$. Finally, we set
$$\Omega_{j-1} := \{x \in \Omega_j : |S_j(x)| \geq C_j^{-1}\alpha_j, \, 0 \in J_j(x)\}, \tag{2.5}$$

with C_j sufficiently large. Note that $0 \in J_j(x)$ if and only if $x \in \{e^{tX_j}(x) : t \in S_j(x) \cap J_j(x)\}$.

We claim that $|\Omega_{j-1}| \sim |\Omega_j|$. Indeed,
$$|\Omega_j| = \int_{\pi_j(\Omega_j)} |S_j(\sigma_j(y))| \, dy \sim \int_{\pi_j(\Omega_j^g)} |S_j(\sigma_j(y))| \, dy$$
$$\sim \int_{\pi_j(\Omega_j^g)} |S_j(\sigma_j(y)) \cap J_j(\sigma_j(y))| \, dy = |\Omega_{j-1}|,$$

where $\Omega_j^g := \{x \in \Omega_j : |S_j(x)| > C_j^{-1}\alpha_j\}$ (same constant as in (2.5)), and the second '$\sim$' uses Proposition 2.2.

We claim that that each Ω_j is open (possibly after a minor refinement). Since Ω is open, it suffices to prove that Ω_{j-1} is open whenever Ω_j is open. By deleting a set of measure much smaller than $|\Omega_j|$, we may assume that
$$\Omega_j = \bigcup_{\alpha \in \mathcal{A}} \{e^{tX_j}(\sigma_j(y)) : y \in B_\alpha, t \in S_\alpha\},$$

where the B_α are disjoint open subsets of $\mathbb{R}^{n-1}$, the S_α are open subsets of $\mathbb{R}$, and $(t, y) \mapsto e^{tX_j}(\sigma_j(y))$ is a diffeomorphism on $B_\alpha \times S_\alpha$. (We make no hypotheses on $\#\mathcal{A}$.) Then $S_j(e^{tX_j}(\sigma_j(y))) = S_\alpha + t$ and $J_j(e^{tX_j}(\sigma_j(y))) = J(N, S_\alpha) + t$, for each $(y, t) \in B_\alpha \times S_\alpha$. By construction, there exists a subset $\mathcal{A}' \subseteq \mathcal{A}$ such that we may write
$$\Omega_{j-1} = \bigcup_{\alpha \in \mathcal{A}'} \{e^{tX_j}(\sigma_j(y)) : y \in B_\alpha, t \in S_\alpha \cap J(N, S_\alpha)\},$$

a union of open sets.

Let $x_0 \in \Omega_0$, and for $t \in \mathbb{R}^n$, define

$$\Psi_{x_0}(t) := e^{t_n X_n} \circ \cdots \circ e^{t_1 X_1}(x_0). \tag{2.6}$$

Define $\mathcal{F}_1 := S_1(x_0)$, and for each $j = 2, \ldots, n$,

$$\mathcal{F}_j := \{(t', t_j) \in \mathbb{R}^j : t' \in \mathcal{F}_{j-1}, \ t_j \in S_j(\Psi_{x_0}(t', 0))\}.$$

Thus for $t \in \mathcal{F}_j$, $\Psi_{x_0}(t, 0) \in \Omega_j$, so $0 \in J_j(\Psi_{x_0}(t, 0))$.

In particular, $\Psi_{x_0}(\mathcal{F}_n) \subseteq \Omega$, so by Lemma 11.7,

$$|\Omega| \geq |\Psi_{x_0}(\mathcal{F}_n)| \gtrsim \int_{\mathcal{F}_n} |\det D\Psi_{x_0}(t)| \, dt.$$

Since $0 \in J_j(\Psi_{x_0}(t', 0))$ for each $t' \in \mathcal{F}_j$, we compute

$$\begin{aligned}
\int_{\mathcal{F}_n} |\det D\Psi_{x_0}(t)| \, dt &= \int_{\mathcal{F}_{n-1}} \int_{S_n(\Psi_{x_0}(t', 0))} |\det D\Psi_{x_0}(t', t_n)| \, dt_n \, dt' \\
&\gtrsim \alpha_n^{\beta_n+1} \int_{\mathcal{F}_{n-1}} |\partial_{t_n}^{\beta_n} \det D\Psi_{x_0}(t', 0)| \, dt' \\
&\gtrsim \alpha_n^{\beta_n+1} \cdots \alpha_1^{\beta_1+1} |\partial_t^{\beta} \det D\Psi_{x_0}(0)| \sim \alpha_1^{b_1} \alpha_2^{b_2}.
\end{aligned} \tag{2.7}$$

After a little arithmetic, we see that (2.1) is equivalent to $|\Omega| \gtrsim \alpha_1^{b_1} \alpha_2^{b_2}$, so the proposition is proved. $\qquad\square$

We have not yet used the gain in Proposition 2.2; we will take advantage of that in Section 5 when we prove a structure theorem for pairs of sets for which the restricted weak type inequality (2.1) is nearly reversed. Before we state this structure theorem, it will be useful to understand better the geometry of the image under Ψ_{x_0} of axis parallel rectangles.

3. Carnot–Carathéodory balls associated to polynomial flows

In the previous section, we proved uniform restricted weak type inequalities at a single scale. To improve these to strong type inequalities, we need more, namely, an understanding of those sets for which the inequality (2.1) is nearly optimal. In this section, we lay the groundwork for that characterization by establishing a few lemmas on Carnot–Carathéodory balls associated to nilpotent vector fields with polynomial flows. Results along similar lines have appeared elsewhere, [6, 17, 26, 27] in particular, but we need more uniformity and a few genuinely new lemmas, and, moreover, our polynomial and nilpotency hypotheses allow for simpler proofs than are available in the general case.

We begin by reviewing our hypotheses and defining some new notation. We have vector fields $X_1, X_2 \in \mathcal{X}(\mathbb{R}^n)$ that are assumed to generate a Lie subalgebra $\mathfrak{g} \subseteq \mathcal{X}(\mathbb{R}^n)$ that is nilpotent of step at most N, and such that for each $X \in \mathfrak{g}$, the exponential map $(t, x) \mapsto e^{tX}(x)$ is a polynomial of degree at most N in t and in x.

Lemma 3.1. *The elements of $\mathfrak{g}$ are divergence-free.*

Proof. Let $X \in \mathfrak{g}$. Both $\det De^{tX}(x)$ and its multiplicative inverse, which may be written $\det(De^{-tX})(e^{tX}(x))$, are polynomials, so both must be constant in t and x. Evaluating at $t = 0$, we see that these determinants must equal 1, so the flow of X is volume-preserving, i.e. X is divergence-free. $\qquad\square$

A *word* is a finite sequence of 1's and 2's, and associated to each word w is a vector field X_w, where $X_{(i)} := X_i$, $i = 1, 2$, and $X_{(i,w)} := [X_i, X_w]$. We let $\mathcal{W}$ denote the set of all words w with $X_w \not\equiv 0$. For $I \in \mathcal{W}^n$, we define $\lambda_I := \det(X_{w_1}, \ldots, X_{w_n})$, and we define $\Lambda := (\lambda_I)_{I \in \mathcal{W}^n}$. We denote by $|\Lambda|$ the sup-norm.

As in the proof of Proposition 2.2, we will repeatedly, and without comment, use the fact that all norms on the finite dimensional vector space of polynomials of degree at most (e.g.) N are equivalent.

Throughout this section, c denotes a sufficiently small constant depending on N.

Lemma 3.2. *Assume that $|\lambda_I(0)| \geq \delta|\Lambda(0)|$, for some $\delta > 0$. Then for any $w \in \mathcal{W}$,*

$$|\lambda_I(e^{tX_w}(0))| \sim |\lambda_I(0)|, \qquad |\lambda_I(e^{tX_w}(0))| \gtrsim \delta|\Lambda(e^{tX_w}(0))|,$$

for all $|t| < c\delta$.

Proof. By Lemma 3.1, X_w is divergence-free. By the formula for the Lie derivative of a determinant, for any $I' = (w'_1, \ldots, w'_n) \in \mathcal{W}^n$,

$$X_w \lambda_{I'} = \sum_{i=1}^{n} \lambda_{I'_i},$$

where I'_i is obtained from I' by replacing the i-th entry with $[X_w, X_{w'_i}]$. Thus for each k,

$$|\tfrac{d^k}{dt^k} \Lambda(e^{tX_w}(0))| \lesssim |\Lambda(0)| \lesssim \delta^{-1}|\lambda_I(0)|. \tag{3.1}$$

As $t \mapsto \Lambda \circ e^{tX_w}(0)$ is a polynomial of bounded degree, the first inequality in (3.1) implies that $|\Lambda(e^{tX_w}(0))| \sim |\Lambda(0)|$ for $|t| < c$. Moreover, (3.1) implies that $|\tfrac{d}{dt} \lambda_I(e^{tX_w}(0))| \lesssim \delta^{-1}|\lambda_I(0)|$, so $|\lambda_I(e^{tX_w}(0))| \sim |\lambda_I(0)|$ for $|t| < c\delta$. The conclusion of the lemma follows. $\qquad\square$

For $I = (w_1, \ldots, w_n) \in \mathcal{W}^n$, we define a map

$$\Phi^I_{x_0}(t_1, \ldots, t_n) := e^{t_n X_{w_n}} \circ \cdots \circ e^{t_1 X_{w_1}}(x_0).$$

Lemma 3.3. *Let $I \in \mathcal{W}^n$, and assume that $|\lambda_I(0)| \geq \delta|\Lambda(0)|$. Then for all $|t| < c\delta$,*

$$|\det D\Phi^I_0(t)| \sim |\lambda_I \circ \Phi^I_0(t)| \sim |\lambda_I(0)|, \tag{3.2}$$

and $|\Lambda \circ \Phi^I_0(t)| \sim |\Lambda(0)|$.

Proof. By Lemma 3.2 and a simple induction, we have only to show that $|\det D\Phi^I_0(t)| \sim |\lambda_I(0)|$, for all $|t| < c\delta$. Since the flow of each X_w is volume-preserving, we may directly compute

$$\det D\Phi^I_0(t) = \det(X_{w_1}(0), \phi^*_{t_1 X_{w_1}} X_{w_2}(0), \ldots, \phi^*_{t_1 X_{w_1}} \cdots \phi^*_{t_{n-1} X_{w_{n-1}}} X_{w_n}(0)),$$

where

$$\phi^*_X Y(x) := De^{-X}(e^X(x))Y(e^X(x)).$$

Since

$$\tfrac{d}{dt} \phi^*_{tX} Y = \phi^*_{tX}[X, Y], \tag{3.3}$$

this gives

$$|\partial_t^\beta \det D\Phi^I_0(0)| \lesssim |\Lambda(0)| \leq \delta^{-1}|\lambda_I(0)| = \delta^{-1}|\det D\Phi^I_0(0)|,$$

for all multiindices β. This gives us the desired bound on $|\det D\Phi^I_0(t)|$, for $|t| < c\delta$. $\qquad\square$

Lemma 3.4. *Assume that $|\lambda_I(0)| \geq \delta|\Lambda(0)|$. Then Φ_0^I is one-to-one on $\{|t| < c\delta\}$, and for each $w \in \mathcal{W}$, the pullback $Y_w := (\Phi_0^I)^* X_w$ satisfies $|Y_w(t)| \lesssim \delta^{-1}$ on $\{|t| < c\delta\}$.*

Proof. We write $D\Phi_0(t) = A(t, \Phi_0(t))$, where A is the matrix-valued function given by

$$A(t, x) := (\phi^*_{-t_n X_{w_n}} \cdots \phi^*_{-t_2 X_{w_2}} X_{w_1}(x), \cdots, \phi^*_{-t_n X_{w_n}} X_{w_{n-1}}(x), X_{w_n}(x)).$$

By the nilpotency hypothesis and (3.3), each column of A is polynomial in t, and thus may be computed by differentiating and evaluating at $t = 0$. Using the Jacobi identity, iterated Lie brackets of the X_{w_i} may be expressed as iterated Lie brackets of the X_i, and so

$$\phi^*_{-t_n X_{w_n}} \cdots \phi^*_{-t_{i+1} X_{w_{i+1}}} X_{w_i} = X_{w_i} + \sum_{w \in \mathcal{W}} p_{w_i, w}(t) X_w, \tag{3.4}$$

where each $p_{w_i, w}$ is a polynomial in $(t_{i+1}, \ldots, t_n)$, with bounded coefficients and $p_{w_i, w}(0) = 0$. By Cramer's rule, for each w,

$$\lambda_I X_w = \sum_{i=1}^{n} \lambda_{I(w,i)} X_{w_i}, \tag{3.5}$$

where $I(w, i)$ is obtained from I by replacing X_{w_i} with X_w. Combining (3.4) and (3.5), we may write

$$A = (X_{w_1}, \ldots, X_{w_n})(\mathbb{I}_n + \lambda_I^{-1} P),$$

where $\mathbb{I}_n$ is the identity matrix and P is a matrix-valued polynomial whose entries are linear combinations of the products $p_{w_i, w} \lambda_{I(w,i)}$. Since $p_{w_i, w}$ has bounded coefficients and vanishes at zero, $|p_{w_i, w}(t)| \lesssim \delta$ on $\{|t| < c\delta\}$, and so by Lemma 3.3,

$$|P \circ \Phi_0^I(t)| \lesssim |\lambda_I \circ \Phi_0^I(t)| \sim |\lambda_I(0)|, \tag{3.6}$$

on $\{|t| < c\delta\}$.

Recalling the definition of Y_w in the statement of the lemma, $Y_{w_i}(0) = \frac{\partial}{\partial t_i}$, $1 \leq i \leq n$. Let

$$\widetilde{Y}_w := \lambda_I(0)^{-1}(\det D\Phi_0^I) Y_w. \tag{3.7}$$

By Cramer's rule, $\widetilde{Y}_w$ is a polynomial; we also have $\widetilde{Y}_w(0) = Y_w(0)$. We expand

$$Y_w(t) = A(t, \Phi_0(t))^{-1} X_w \circ \Phi_0^I(t)$$

$$= (\mathbb{I}_n + \lambda_I^{-1} \Phi_0^I(t) P \circ \Phi_0^I(t))^{-1} (X_{w_1} \circ \Phi_0^I(t), \ldots, X_{w_n} \circ \Phi_0^I(t))^{-1} X_w \circ \Phi_0^I(t),$$

which directly implies

$$Y_{w_i}(t) = (\mathbb{I}_n + \lambda_I^{-1} \circ \Phi_0^I(t) P \circ \Phi_0^I(t))^{-1} e_i. \tag{3.8}$$

By (3.8) and inequality (3.6),

$$|Y_{w_i} - \tfrac{\partial}{\partial t_i}| \lesssim 1 \tag{3.9}$$

on $\{|t| < c\delta\}$. By Cramer's rule,

$$Y_w = \sum_{i=1}^{n} \frac{\lambda_{I(i)} \circ \Phi_0^I}{\lambda_I \circ \Phi_0^I} Y_{w_i},$$

while Lemma 3.4 bounds the coefficients; combined with inequality (3.9), we obtain $|Y_w| \lesssim \delta^{-1}$ on $\{|t| < c\delta\}$.

By inequality (3.6), $|\frac{\det D\Phi_0^I}{\lambda_I(0)} - 1| \lesssim 1$. Therefore,

$$|\widetilde{Y}_{w_i} - \tfrac{\partial}{\partial t_i}| \lesssim 1 \qquad (3.10)$$

on $\{|t| < c\delta\}$. The vector field $\widetilde{Y}_{w_i}$ is a polynomial that satisfies $\widetilde{Y}_{w_i}(0) = \frac{\partial}{\partial t_i}$, while (3.10) (and the equivalence of norms) implies bounds on the coefficients of $\widetilde{Y}_{w_i}$; taken together, these imply the stronger estimate

$$|\widetilde{Y}_{w_i}(t) - \tfrac{\partial}{\partial t_i}| \lesssim \delta^{-1}|t| \qquad (3.11)$$

on $\{|t| < c\delta\}$. Similarly, $|\frac{\det D\Phi_0^I(t)}{\lambda_I(0)} - 1| \lesssim \delta^{-1}|t|$, whence, from the definition (3.7) of $\widetilde{Y}_{w_i}$ and (3.11),

$$|Y_{w_i}(t) - \tfrac{\partial}{\partial t_i}| \lesssim \delta^{-1}|t|,$$

on $\{|t| < c\delta\}$. Therefore

$$|D_s e^{s_n Y_{w_n}} \circ \cdots \circ e^{s_1 Y_{w_1}}(0) - \mathbb{I}_n| \lesssim \delta^{-1}|t|, \qquad |t| < c\delta,$$

which, by the contraction mapping proof of the Inverse Function Theorem, implies that

$$(s_1, \ldots, s_n) \mapsto e^{s_n Y_{w_n}} \circ \cdots \circ e^{s_1 Y_{w_1}}(0)$$

is one-to-one on $\{|t| < c\delta\}$. Finally, by naturality of exponentiation, Φ_0^I must also be one-to-one on this region. $\qquad\square$

Lemma 3.5. *Let $x_j \in \mathbb{R}^n$, $j = 1, 2$, and assume that $I_j \in \mathcal{W}^n$ are such that $|\lambda_{I_j}(x_j)| \geq \delta|\Lambda(x_j)|$, $j = 1, 2$. Let $0 < \rho < c\delta$. If $\bigcap_{j=1}^2 \Phi_{x_j}^{I_j}(\{|t| < c\delta\rho\}) \neq \emptyset$, then $\Phi_{x_1}^{I_1}(\{|t| < c\delta\rho\}) \subseteq \Phi_{x_2}^{I_2}(\{|t| < \rho\})$.*

Proof. By assumption, each element of $\Phi_{x_1}^{I_1}(\{|t| < c\delta\rho\})$ can be written in the form

$$e^{t_{3n} X_{w_{3n}}} \circ \cdots \circ e^{t_1 X_{w_1}}(x_2),$$

with $w_j \in \mathcal{W}$, $j = 1, \ldots, 3n$, and $|t| < 3c\delta\rho$. Setting $Y_w := (\Phi_{x_2}^{I_2})^* X_w$, Lemma 3.4 (together with the Mean Value Theorem) implies that

$$|e^{t_{3n} Y_{w_{3n}}} \circ \cdots \circ e^{t_1 Y_{w_1}}(0)| < \rho,$$

whenever $|t| < 3c\delta\rho$, and so the containment claimed in the lemma follows by applying $\Phi_{x_2}^{I_2}$ to both sides. $\qquad\square$

We recall that $\Psi_{x_0} = \Phi_{x_0}^{(1,2,1,2,\ldots)}$, and we define $\widetilde{\Psi}_{x_0} := \Phi_{x_0}^{(2,1,2,1,\ldots)}$. For $\beta \in \mathbb{Z}_{\geq 0}^n$ a multiindex, we define

$$J^\beta(x_0) := \partial^\beta \det D\Psi_{x_0}(0), \qquad \widetilde{J}^\beta(x_0) := \partial^\beta \det D\widetilde{\Psi}_{x_0}(0). \qquad (3.12)$$

Lemma 3.6.

$$|\Lambda(0)| \sim \sum_\beta |J^\beta(0)| + |\widetilde{J}^\beta(0)|. \qquad (3.13)$$

Proof. The argument that follows is due to Tao–Wright, [27]; we reproduce it to keep better track of constants to preserve the uniformity that we need.

Direct computation shows that the J^β and $\widetilde{J}^\beta$ are linear combinations of determinants λ_I, and it immediately follows that the left side of (3.13) bounds the right.

To bound the left side, it suffices to prove that there exists $|t| \lesssim 1$ such that

$$|\Lambda(0)| \lesssim |\det D\Psi_0(t)| + |\det D\widetilde{\Psi}_0(t)|,$$

which is equivalent (via naturality of exponentiation and Lemma 3.4) to finding a point $|s| \lesssim 1$ such that

$$1 \lesssim |\det D_s e^{s_n Y_n} \circ \cdots \circ e^{s_1 Y_1}(0)| + |\det D_s e^{s_n Y_{n+1}} \circ \cdots \circ e^{s_1 Y_2}(0)|,$$

where the vector fields Y_i are those defined in Lemma 3.4, the n-tuple I having been chosen to maximize $\lambda_I(0)$.

By Lemma 3.4, $\|Y_w\|_{C^N(\{|t|<c\})} \lesssim 1$, for all $w \in \mathcal{W}$. By induction, this implies that $|Y_w(0)| \lesssim (|Y_1(0)| + |Y_2(0)|)$. Since $|Y_{w_i}(0)| = 1$, $|Y_1(0)| + |Y_2(0)| \sim 1$. Thus (3.14) holds for $k = 1$, $s = 0$. Without loss of generality, we may assume that $|Y_1(0)| \sim 1$. Now we proceed inductively, proving that for each $1 \leq k \leq n$, there exists a point $|(s_1, \ldots, s_{k+1})| < c$ such that

$$1 \sim |\partial_1 e^{s_k Y_k} \circ \cdots \circ e^{s_1 Y_1}(0) \wedge \cdots \wedge \partial_k e^{s_k Y_k} \circ \cdots \circ e^{s_1 Y_1}(0)|; \qquad (3.14)$$

the case $k = 1$, $s = 0$ having already been proved. Assume that (3.14) holds for some $k < n$, $|s| = |s^0| < c$. Then $(s_1, \ldots, s_k) \mapsto e^{s_k Y_k} \circ \cdots \circ e^{s_1 Y_1}(0)$, $|s - s^0| < c'$ parametrizes a k-dimensional manifold M, and the vector fields

$$Z_i(e^{s_k Y_k} \circ \cdots \circ e^{s_1 Y_1}(0)) := \partial_{s_i} e^{s_k Y_k} \circ \cdots \circ e^{s_1 Y_1}(0), \qquad i = 1, \ldots, k,$$

form a basis for the tangent space of M at each point.

Let us suppose that the analogue of (3.14) for $k+1$ fails. Then for all $|s-s^0| < c'$, we may decompose Y_{k+1} as

$$Y_{k+1}(e^{s_k Y_k} \circ \cdots \circ e^{s_1 Y_1}(0)) = \sum_{i=1}^{k} a_i(s) Z_i(e^{s_k Y_k} \circ \cdots \circ e^{s_1 Y_1}(0)) + Y^\perp(e^{s_k Y_k} \circ \cdots \circ e^{s_1 Y_1}(0)),$$

$$(3.15)$$

with $\|a_i\|_{C^N(\{|s-s^0|<1\})} \lesssim 1$, and $|\partial_s^\alpha Y^\perp| = |Z^\alpha Y^\perp| < c''$, for c'' as small as we like and all $|\alpha| < N$; otherwise, by equivalence of norms, the analogue of (3.14) for $k + 1$ would hold. By construction, $Z_k = Y_k$, thus by induction and (3.15),

$$|Z_1 \wedge \cdots \wedge Z_k \wedge Y_w(e^{s_k^0 Y_k} \circ \cdots \circ e^{s_1^0 Y_1}(0))| < c'',$$

for any word w with $\deg_i(w) > 0$, where $i \equiv k + 1 \mod 2$. By (3.14) and boundedness of the Y_w,

$$|\det(Y_{w_1}, \ldots, Y_{w_n})(e^{s_k^0 Y_k} \circ \cdots \circ e^{s_1^0 Y_1}(0))| < c'',$$

for an (possibly different but) arbitrarily small constant c''. Thus

$$|\lambda_I(e^{s_k^0 X_k} \circ \cdots \circ e^{s_1^0 X_1}(0))| < c''|\Lambda(0)|,$$

which, by Lemma 3.2, contradicts our assumption that $|\lambda_I(0)| \sim |\Lambda(0)|$. $\square$

We say that a k-tuple $(w_1, \ldots, w_k) \in \mathcal{W}$ is *minimal* if $w_1, w_2 \in \{(1), (2)\}$, and for $i \geq 3$, $w_i = (j, w_l)$ for some $j = 1, 2$ and $l < i$. It will be important later that a minimal n-tuple must contain the indices (1), (2), and $(1, 2)$.

Lemma 3.7. *Under the assumption that $|\Lambda(0)| \lesssim \sum_\beta |J^\beta(0)|$, there exists a minimal n-tuple $I^0 \in \mathcal{W}^n$ such that for all $\varepsilon > 0$,*

$$\left|\{x \in \Psi_0(\{|t| < 1\}) : |\lambda_{I^0}(x)| \gtrsim |\Lambda(x)|\}\right| \geq (1 - \varepsilon)|\Psi_0(\{|t| < 1\})|. \qquad (3.16)$$

We recall that the implicit constants in the conclusion can depend on the implicit constants in the hypothesis.

The proof of Lemma 3.7 will utilize the following simple fact.

Lemma 3.8. *Let P be a polynomial of degree at most N on $\mathbb{R}^n$. Then for each $\varepsilon > 0$,*

$$|\{t \in \mathbb{I}^n : |P(t)| < \varepsilon \|P\|_{C^0(\mathbb{I}^n)}\}| \lesssim \varepsilon^{1/C}, \qquad \mathbb{I}^n := [-1,1]^n.$$

Proof of Lemma 3.8. Assume that $n = 1$, $N \geq 1$, and $P(t) = \prod_{i=1}^N (t - w_i)$. We may assume $|w_i| \leq 2$ for all i, as for other i, $|t - w_i| \sim |w_i|$ on $\mathbb{I}$. Thus $A := \|P\|_{C^0(\mathbb{I})} \sim 1$. The set $\{-\varepsilon^2 A^2 < P\overline{P} < \varepsilon^2 A^2\}$ is a union of at most $2N$ bad intervals I on which $|P| < \varepsilon A$. For any interval $I \subseteq \mathbb{I}$, $\|P\|_{C^0(I)} \geq |P^{(N)}||I|^N \gtrsim |I|^N$, so a bad interval has length $|I| \lesssim \varepsilon^{1/N}$.

Now assume that the lemma has been proved for dimensions at most n. Given P, a polynomial of degree at most N on $\mathbb{R}^{n+1}$, set

$$Q(t') := \int_{-1}^1 |P(t', t_{n+1})|^2 \, dt_{n+1}, \qquad t' \in \mathbb{R}^n,$$

a polynomial of degree at most $2N$ on $\mathbb{R}^n$. If $|P(t)| < \varepsilon \|P\|_{C^0(\mathbb{I}^{n+1})}$, then $|Q(t')| < \varepsilon \|P\|^2_{C^0(\mathbb{I}^{n+1})}$, or $|P(t', t_n)|^2 < \varepsilon |Q(t')|$. However, by equivalence of norms, $|Q(t')| \sim \|P(t', \cdot)\|^2_{C^0(\mathbb{I})}$, so $\|Q\|_{C^0(\mathbb{I}^n)} \sim \|P\|^2_{C^0(\mathbb{I}^{n+1})}$, and the conclusion follows from the 1 and n-dimensional cases. $\qquad\square$

Proof of Lemma 3.7. Fix an n-tuple $I = (w_1, \ldots, w_n) \in \mathcal{W}^n$ such that $|\widetilde{\lambda}_I(0)| \gtrsim |\widetilde{\Lambda}(0)|$, where $\widetilde{\lambda}_I$ and $\widetilde{\Lambda}$ are defined using vector fields $\widetilde{X}_w$ generated by the $\widetilde{X}_i := KX_i$, $i = 1, 2$. (The constant K allows us to apply the technical lemmas above on large balls.) By Lemma 3.4 (see also the proof of Lemma 3.5), for K sufficiently large,

$$\Psi_0(\{|t| < 1\}) \subseteq \Phi_0^I(\{|t| < C\}). \tag{3.17}$$

With the vector fields Y_w defined as in Lemma 3.4 (using the X_i, not the KX_i), Lemmas 3.2 and 3.3 imply

$$|\det(Y_{w_1}, \ldots, Y_{w_n})| \sim 1, \quad \text{throughout} \quad \{|t| < C\}, \tag{3.18}$$

provided K is sufficiently large.

We will prove that there exists a minimal n-tuple $I^0 = (w_1^0, \ldots, w_n^0)$ such that

$$\|\det(Y_{w_1^0}, \ldots, Y_{w_n^0})\|_{C^0(\{|t| < C\})} \sim 1. \tag{3.19}$$

Before proving (3.19), we show that it implies inequality (3.16).

By Lemma 3.2 $|\Lambda(x)| \sim |\Lambda(0)|$ for all $x \in \Phi_0^I(\{|t| < C\})$. Thus, unwinding the definition (from Lemma 3.4) of the Y_w, (3.19) implies that

$$\|\lambda_{I^0} \circ \Phi_0^I\|_{C^0\{|t| < C\}} \sim |\Lambda(0)| \sim \|\Lambda \circ \Psi_0\|_{C^0\{|t| < 1\}}. \tag{3.20}$$

By (3.20) and (3.17),

$$\{x \in \Psi_0(\{|t| < 1\}) : |\lambda_{I^0}(x)| < \delta|\Lambda(x)|\} \subseteq \{x \in \Phi_0^I(\{|t| < C\}) : |\lambda_{I^0}(x)| \lesssim \delta|\Lambda(0)|\},$$

for $\delta > 0$. By the change of variables formula and (3.20), Lemmas 3.3 and 3.8, our hypothesis and the equivalence of norms, and finally the change of variables formula and Lemma 11.7,

$$|\{x \in \Phi_0^I(\{|t| < C\}) : |\lambda_{I^0}(x)| \lesssim \delta|\Lambda(0)|\}|$$

$$\leq \|\det D\Phi_0^I\|_{C^0(\{|t| < C\})}|\{|t| < C : |\lambda_{I^0} \circ \Phi_0^I(t)| \lesssim \delta\|\lambda_{I^0} \circ \Phi_0^I\|_{C^0(\{|t| < C\})}\}|$$

$$\lesssim \delta^{1/C}|\Lambda(0)| \lesssim \delta^{1/C}\|\det D\Psi_0\|_{L^1(\{|t| < 1\})} \lesssim \delta^{1/C}|\Psi_0(\{|t| < 1\})|.$$

Setting $\delta = c'\varepsilon^{C'}$ with c' and C' sufficiently large depending on C and the implicit constant in the hypothesis of the lemma yields (3.16).

It remains to prove (3.19). We will prove inductively that for each $1 \le k \le n$, there exists a minimal k-tuple $(w_1^0, \ldots, w_k^0)$ such that $\|Y_{w_1^0} \wedge \cdots \wedge Y_{w_k^0}\|_{C^0(\{|t|<1\})} \sim 1$. Boundedness of the Y's and our hypothesis imply that $|Y_1(0)| \sim 1$. For the induction step, it will be useful to have two constants, $c, \delta_N > 0$, depending only on N. We will choose c sufficiently small that the deductions below are valid, and then choose δ_N sufficiently small (depending on c and various implicit constants) to derive a contradiction if the induction step fails.

Suppose that for some $k < n$, we have found a minimal k-tuple $(w_1^0, \ldots, w_k^0)$, with $w_1^0 = (1)$, and some $|t^0| < 1$ such that

$$|Y_{w_1^0}(t^0) \wedge \cdots \wedge Y_{w_k^0}(t^0)| \sim 1. \tag{3.21}$$

Set $\mathcal{W}_k^0 := \{w_1^0, \ldots, w_k^0\}$.

By (3.18), we may extend $Y_{w_1^0}, \ldots, Y_{w_k^0}$ to a frame on $\{|t - t^0| < c\}$ by adding vector fields Y_{w_i}. Thus (after possibly reordering the w_i) failure of the inductive step implies that for each

$$w \in \mathcal{W}_k^1 := \begin{cases} \{(1), (2)\}, & \text{if } k = 1, \\ \mathcal{W}_k^0 \cup \{(i, w) : i \in \{1, 2\}, w \in \mathcal{W}_k^0\}, & k > 1, \end{cases}$$

$|Y_{w_1^0} \wedge \cdots \wedge Y_{w_k^0} \wedge Y_w(t)| < \delta_N$ for all t such that $|t - t^0| < c$. Therefore we can write

$$Y_w(t) = \sum_{i=1}^{k} a_w^i(t) Y_{w_i^0}(t) + \sum_{j=k+1}^{n} a_w^j(t) Y_{w_j}(t), \quad w \in \mathcal{W}_k^1, \tag{3.22}$$

where

$$\|a_w^i\|_{C^N(\{|t - t^0| < c\})} \lesssim \begin{cases} 1, & 1 \le i \le k \\ \delta_N, & k + 1 \le j \le n. \end{cases} \tag{3.23}$$

Taking the Lie bracket of Y_i, $i = 1, 2$ (or just Y_1, when $k = 1$), with some Y_w, $w \in \mathcal{W}_k^1 \setminus \mathcal{W}_k^0$,

$$[Y_i, Y_w] = \sum_{i=1}^{k} Y_i(a_i) Y_{w_i^0} + \sum_{i=1}^{k} a_i^0 [Y_i, Y_{w_i^0}] + \sum_{j=k+1}^{n} Y_i(a_j) Y_{w_j} + \sum_{j=k+1}^{n} a_j [Y_i, Y_{w_j}],$$

and we see that (3.22-3.23) hold for

$$w \in \mathcal{W}_k^2 := \mathcal{W}_k^1 \cup \{(i, w) : i \in \{1, 2\}, w \in \mathcal{W}_k^1\}.$$

By induction, (3.22-3.23) are valid for each Y_w, $w \in \mathcal{W}$, so

$$|\det(Y_{w_1}(t^0), \ldots, Y_{w_n}(t^0))| < \delta_N$$

(because the Y_{w_i} must all lie near the span of $Y_{w_1^0}, \ldots, Y_{w_k^0}$), a contradiction to (3.18). □

For $I = (w_1, \ldots, w_n) \in \mathcal{W}^n$ and $\sigma \in S_n$ a permutation, we set $I_\sigma := (w_{\sigma(1)}, \ldots, w_{\sigma(n)})$.

Lemma 3.9. *Assume that* $|\Lambda(0)| \lesssim \sum_\beta |J^\beta(0)|$. *There exist* c', C' *such that for all sufficiently small* c *and large* C, *the following holds. There exists a minimal n-tuple* $I \in \mathcal{W}^n$, *which is allowed to depend on the* X_j, *such that for all* $\varepsilon > 0$, *there exists a collection* $\mathcal{A} \subseteq \Psi_0(\{|t| < 1\})$, *of cardinality* $\#\mathcal{A} \lesssim_{c,C} 1$, *such that*
(i) $|\Psi_0(\{|t| < 1\}) \cap \bigcup_{x \in \mathcal{A}} \bigcap_{\sigma \in S_n} \Phi_x^{I_\sigma}(\{|t| < c\varepsilon^C\})| \ge (1 - \varepsilon)|\Psi_0(\{|t| < 1\})|,$

and, moreover, for all $x \in \mathcal{A}$, $\sigma, \sigma' \in S_n$, and $y \in \Phi_x^{I_\sigma}(\{|t| < c\varepsilon^C\})$,

(ii) $\Phi_y^{I_{\sigma'}}$ is one-to-one on $\{|t| < c'\varepsilon^{C'}\}$, with Jacobian determinant

$$|\det D\Phi_y^{I_{\sigma'}}(t)| \sim |\lambda_I(y)| \sim |\lambda_I(x)| \gtrsim |\Lambda(x)| \sim |\Lambda(\Phi_y^{I_{\sigma'}}(t))|,$$

(iii) $\Phi_x^{I_\sigma}(\{|t| < c\varepsilon^C\}) \subseteq \Phi_y^{I_{\sigma'}}(\{|t| < c'\varepsilon^{C'}\})$.

Proof. By Lemma 3.7, there exists $\delta \gtrsim 1$ and a minimal $I \in \mathcal{W}^n$ such that if

$$G := \{x \in \Psi_0(\{|t| < 1\}) : |\lambda_I(x)| \geq \delta|\Lambda(x)|\},$$

then $|G| \geq (1 - \varepsilon)|\Psi_0(\{|t| < 1\})|$. We may assume: that c' is sufficiently small, that $\varepsilon^{C'} < c'\delta$, and that $c\varepsilon^C < c'^3\delta^2$. Conclusions (ii) and (iii) of the lemma for any choice of such balls are direct applications of Lemmas 3.3, 3.4, and 3.5.

It remains to cover G by a controllable number of balls of the form

$$B_x(\rho) := \bigcap_{\sigma \in S_n} \Phi_x^{I_\sigma}(\{|t| < \rho\}), \qquad x \in G,$$

in the special case $\rho = c\varepsilon^C$. We will use the generalized version of the Vitali Covering Lemma in [20], for which we need to verify the doubling and engulfing properties. By Lemma 3.5, for all $0 < \rho < c'\delta$, $\sigma \in S_n$, and $x \in G$,

$$\Phi_x^{I_\sigma}(\{|t| < \rho\}) \supseteq B_x(\rho) \supseteq \Phi_x^{I_\sigma}(\{|t| < c'\delta\rho\}). \tag{3.24}$$

Hence by Lemma 3.3, $|B_x(\rho)| \approx |\lambda_I(x)|\rho^n \approx |\Lambda(0)|\rho^n$. Therefore the balls are indeed doubling. The engulfing property also follows from Lemma 3.5, since $B_{x_1}(c'\delta\rho) \cap B_{x_2}(c'\delta\rho) \neq \emptyset$ implies that $B_{x_1}(c'\delta\rho) \subseteq B_{x_2}(\rho)$.

If we choose $\mathcal{A} \subseteq G$ so that $\{B_x(c^2\varepsilon^C)\}_{x \in \mathcal{A}}$ is a maximal disjoint set, then $\bigcup_{x \in \mathcal{A}} B_x(c^2\varepsilon^C) \subseteq \Psi_0(\{|t| < 2\})$ and $G \subseteq \bigcup_{x \in \mathcal{A}} B_x(c\varepsilon^C)$. Applying (3.24) and Lemma 3.6,

$$\#\mathcal{A}|\Lambda(0)|(c^2\varepsilon^C)^n \gtrsim |\Psi_0(\{|t| < 2\})| \lesssim |\Lambda(0)|.$$

$\square$

4. Connection with the work of Tao–Wright

In this section, we translate Theorem 1.1 into results more closely connected with the main theorem of [27]. We are also able to prove variants of Theorem 1.1 with weights that are, in principle, easier to compute.

The results of [27] are stated in terms of the Newton polytope associated to the vector fields X_1, X_2. To define it (and two other, closely related, polytopes), we need some additional notation. The degree of a word $w \in \mathcal{W}$ is defined to be the element $\deg w \in \mathbb{Z}_{\geq 0}^2$ whose i-th entry is the number of i's in w. The degree of a k-tuple $I \in \mathcal{W}^k$ is the sum of the degrees of the entries of I. We denote by ch the operation of taking the convex hull of a set. For $E \subseteq \mathbb{R}^n$, we define

$$\mathcal{P}_E^\cup := \text{ch} \bigcup_{I \in \mathcal{W}^n : \lambda_I \not\equiv 0 \text{ on } E} \deg I + [0, \infty)^2. \tag{4.1}$$

Given x_0, we may define

$$\mathcal{P}_{x_0} := \mathcal{P}_{\{x_0\}}^\cup. \tag{4.2}$$

Finally, given a set $E \subseteq \mathbb{R}^n$, we may define

$$\mathcal{P}_E^\cap := \bigcap_{x_0 \in E} \mathcal{P}_{x_0}. \tag{4.3}$$

In [27], Tao–Wright considered bounds of the form

$$| \int f_1 \circ \pi_1(x) f_2 \circ \pi_2(x) \, a(x) \, dx | \leq C_{a,\pi_1,\pi_2} \|f_1\|_{p_1} \|f_2\|_{p_2}, \qquad (4.4)$$

with a a continuous function with compact support, π_1, π_2 smooth submersions (no polynomial nor nilpotency hypothesis), and $p_1, p_2 \in [1, \infty]$. Such bounds are easily seen to be true if $p_1^{-1} + p_2^{-1} \leq 1$. In the case $p_1^{-1} + p_2^{-1} > 1$, we define

$$b(p) = (b_1, b_2) := \Big(\frac{p_1^{-1}}{p_1^{-1} + p_2^{-1} - 1}, \frac{p_1^{-1}}{p_1^{-1} + p_2^{-1} - 1} \Big). \qquad (4.5)$$

Tao–Wright proved that (4.4) fails if $(b_1, b_2) \notin \mathcal{P}^{\cap}_{\{a \neq 0\}}$ and holds if $(b_1, b_2) \in \text{int} \, \mathcal{P}^{\cap}_{\text{supp} \, a}$.

These results leave open two natural questions: what is the role played by the behavior of a near its zero set, and what happens on the boundaries of these polytopes. This article answers these questions in some special cases. To understand how, we first recall the connection between the polytopes defined above and the weights ρ_β.

Given a multiindex $\beta \in \mathbb{Z}^n_{\geq 0}$, we define

$$\widetilde{b}(\beta) := \Big(\sum_{j \, \text{even}} 1 + \beta_j, \sum_{j \, \text{odd}} 1 + \beta_j \Big),$$

and recall the definition (1.6) of $b(\beta)$ and (3.12) of J_β and $\widetilde{J}_\beta$. Proposition 2.3 of [25] implies that

$$\mathcal{P}_{x_0} = \text{ch} \left[\Big(\bigcup_{\beta: J_\beta(x_0) \neq 0} b(\beta) + [0, \infty)^2 \Big) \cup \Big(\bigcup_{\beta: \widetilde{J}_\beta(x_0) \neq 0} \widetilde{b}(\beta) + [0, \infty)^2 \Big) \right], \qquad (4.6)$$

and further that for b an extreme point of $\mathcal{P}_{x_0}$,

$$\sum_{I: \deg I = b} |\lambda_I(x_0)| \sim_b \sum_{\beta: b(\beta) = b} |J_\beta(x_0)| + \sum_{\beta: \widetilde{b}(\beta) = \widetilde{b}} |\widetilde{J}_\beta(x_0)|. \qquad (4.7)$$

The comparison (4.7) is thus valid everywhere on E for b an extreme point of $\mathcal{P}^{\cup}_E$, since both sides of (4.7) are zero when b is not an extreme point of $\mathcal{P}_{x_0}$. Combining these results with Theorem 1.1 and Proposition 2.2 of [25], we obtain the following sharp result.

Theorem 4.1. *Assume that hypotheses (i) and (ii) of Theorem 1.1 are in effect, and that $b := b(p)$ is an extreme point of $\mathcal{P}^{\cup}_{\mathbb{R}^n}$. Then*

$$\sup_{f_1, f_2: \|f_1\|_{p_1} = \|f_2\|_{p_2} = 1} | \int_{\mathbb{R}^n} \prod_{j=1}^{2} f_j \circ \pi_j(x) \, a(x) \, dx | \sim \big\| \tfrac{a}{w_b} \big\|_{C^0(\{x: w_b(x) \neq 0\})}, \qquad (4.8)$$

where w_b is the weight defined by

$$w_b := \sum_{I: \deg I = b} |\lambda_I|^{\frac{1}{b_1 + b_2 - 1}}. \qquad (4.9)$$

Proof. The '$\lesssim$' direction directly follows from Theorem 1.1, (4.7), and the comments after (4.7). The '$\gtrsim$' direction is a direct application of Proposition 2.2 of [25] (a different notation for polytopes was used in that article). $\square$

Uniform upper bounds are also possible under slightly weaker hypotheses on b.

Theorem 4.2. *Under the hypotheses (i) and (ii) of Theorem 1.1, if $b := b(p)$ is a minimal element of $\mathcal{P}_{\mathbb{R}^n}^{\cup}$ under the coordinate-wise partial order on $\mathbb{R}^2$, then*

$$| \int_{\mathbb{R}^n} f_1 \circ \pi_1(x) \, f_2 \circ \pi_2(x) \, w_b(x) \, dx | \lesssim \|f_1\|_{p_1} \|f_2\|_{p_2}.$$

Proof. Set

$$J_b^{\Sigma} := \sum_{\beta:b(\beta)=b} |J_\beta| + \sum_{\beta:\widetilde{b}(\beta)=b} |\widetilde{J}_\beta|.$$

By Lemma 3.6, for all $\alpha \in (0,\infty)^2$ and $x_0 \in \mathbb{R}^n$, and $\deg I = b$,

$$\alpha^b |\lambda_I(x_0)| \lesssim \sum_{b'} \alpha^{b'} J_{b'}^{\Sigma}(x_0) = \sum_{b' \in \mathcal{P}_{\mathbb{R}^n}^{\cup}} \alpha^{b'} J_{b'}^{\Sigma}(x_0). \qquad (4.10)$$

By our assumption on b and the definition of $\mathcal{P}_{\mathbb{R}^n}^{\cup}$, there exists $\nu \in (0,\infty)^2$ such that $b \cdot \nu \leq b' \cdot \nu$ for all $b' \in \mathcal{P}_{\mathbb{R}^n}^{\cup}$. Replacing $\alpha = (\alpha_1, \alpha_2)$ with $(\delta^{\nu_1}\alpha_1, \delta^{\nu_2}\alpha_2)$ in (4.10) and sending $\delta \searrow 0$, we see that

$$\alpha^b |\lambda_I(x_0)| \lesssim \sum_{b' \in \mathbb{F}} \alpha^{b'} J_{b'}^{\Sigma}(x_0), \qquad (4.11)$$

where $\mathbb{F} := \{b' \in \mathcal{P}_{\mathbb{R}^n}^{\cup} : b' \cdot \nu = b \cdot \nu\}$. The face $\mathbb{F}$ is a line segment (possibly a singleton),

$$\mathbb{F} = \{b^0 + t\omega : 0 \leq t \leq 1\},$$

for some vector ω perpendicular to ν. Setting $\alpha^{\omega} := \delta$, (4.11) is equivalent to

$$\delta^{\theta_0} |\lambda_I(x_0)| \lesssim \sum_i \delta^{\theta_i} J_{b^{\theta_i}}^{\Sigma}(x_0), \qquad \delta > 0,$$

where $b := b^0 + \theta_0 \omega$ and $\mathbb{F} \cap \mathbb{N}^2 = \{b^0 + \theta^i \omega : 1 \leq i \leq m_n\}$. By Lemma 11.1,

$$|\lambda_I(x_0)| \lesssim J_b^{\Sigma}(x_0) + \sum_{\theta_i < \theta_0 < \theta_j} (J_{b^{\theta_i}}^{\Sigma}(x_0))^{\frac{\theta_j - \theta_0}{\theta_j - \theta_i}} (J_{b^{\theta_j}}^{\Sigma}(x_0))^{\frac{\theta_0 - \theta_i}{\theta_j - \theta_i}} =: (J^{\theta_0})(x_0),$$

for all $x_0 \in \mathbb{R}^n$. Finally, by Theorem 1.1, complex interpolation, and the triangle inequality,

$$| \int_{\mathbb{R}^n} f_1 \circ \pi_1(x) \, f_2 \circ \pi_2(x) \, |J_0^{\theta}(x)|^{\frac{1}{b_1 + b_2 - 1}} \, dx | \lesssim \|f_1\|_{p_1} \|f_2\|_{p_2}.$$

$\square$

Finally, we give the endpoint version of the main result of [27].

Theorem 4.3. *Let a be a continuous function with compact support, and assume that π_1, π_2 obey the hypotheses of Theorem 1.1 and, in addition, that the π_j are submersions throughout $\operatorname{supp} a$. If $b(p) \in \mathcal{P}_{\operatorname{supp} a}^{\cap}$, then*

$$| \int_{\mathbb{R}^n} f_1 \circ \pi_1(x) \, f_2 \circ \pi_2(x) \, a(x) \, dx | \lesssim_{a,\pi_1,\pi_2} \|f_1\|_{p_1} \|f_2\|_{p_2}. \qquad (4.12)$$

Proof. Let $x_0 \in \operatorname{supp} a$. By (4.6) and our hypothesis, there exist b^i, $i = 0, 1$, and $0 \leq \theta \leq 1$ such that $b(p) \succeq b^{\theta} := (1-\theta)b^0 + \theta b^1$ and $J_{b^i}^{\Sigma}(x_0) \sim_{a,\pi_1,\pi_2} 1$.

By continuity, $J_{b^i}^{\Sigma}(x) \sim_{a,\pi_1,\pi_2} 1$ for x in some neighborhood U of x_0. By Theorem 1.1,

$$| \int_U f_1 \circ \pi_1(x) \, f_2 \circ \pi_2(x) \, dx | \lesssim_{a,\pi_1,\pi_2} \prod_{j=1}^2 \|f_j\|_{L^{q_j}(\pi_j(U))}, \qquad q \in [1,\infty]^2, \quad (4.13)$$

holds with $q = p^i$ computed from the b^i using (1.8). By interpolation, (4.13) also holds with $q = p^\theta$ computed from b^θ using (1.8). An elementary computation shows that

$$(p_1^{-1}, p_2^{-1}) = (1 - \nu)(q^{-1}, 1 - q^{-1}) + \nu((p_1^\theta)^{-1}, (p_2^\theta)^{-1}).$$

Our hypothesis that the π_j are submersions on supp a and Hölder's inequality imply that (4.13) holds whenever $q_1^{-1} + q_2^{-1} \leq 1$, and hence by interpolation, (4.13) holds at p. Inequality (4.12) follows by using a partition of unity. □

5. Quasiextremal pairs for the restricted weak type inequality

The purpose of this section is to prove that pairs E_1, E_2 that nearly saturate inequality (2.1) are well approximated as a bounded union of "balls" parametrized by maps of the form Φ_x^I, with I a (reordering of a) minimal n-tuple of words. Results of this type had been previously obtained in [4, 22] for other operators and in [2] for a particular instance of the class considered here.

We begin with some further notation.

Notation. We recall the maps

$$\Phi_{x_0}^I(t) := e^{t_n X_{w_n}} \circ \cdots \circ e^{t_1 X_{w_1}}(x_0), \qquad I = (w_1, \ldots, w_n) \in \mathcal{W}^n,$$

$\Psi_{x_0} := \Phi_{x_0}^{(1,2,1,2,\ldots)}$, $\widetilde{\Psi}_{x_0}(t) := \Phi_{x_0}^{(2,1,2,1,\ldots)}$ from the previous section. For $\alpha \in (0, \infty)^2$, we define parallelepipeds

$$Q_\alpha^I := \{t \in \mathbb{R}^n : |t_i| < \alpha^{\deg w_i}, \ 1 \leq i \leq n\}, \qquad I \in \mathcal{W}^n,$$

$Q_\alpha := Q_\alpha^{(1,2,1,2,\ldots)}$, $\widetilde{Q}_\alpha := Q_\alpha^{(2,1,2,1,\ldots)}$. These give rise to families of balls,

$$B^I(x_0; \alpha) := \Phi_{x_0}^I(Q_\alpha^I), \qquad B^n(x_0; \alpha) := \Psi_{x_0}(Q_\alpha) \cup \widetilde{\Psi}_{x_0}(\widetilde{Q}_\alpha).$$

For $I = (w_1, \ldots, w_n)$ an n-tuple of words and $\sigma \in S_n$ a permutation, we recall that $I_\sigma := (w_{\sigma(1)}, \ldots, w_{\sigma(n)})$.

Proposition 5.1. *Let c, c', C, C' be as described in Lemma 3.9. Let E_1, E_2 be open sets, and let $\varepsilon > 0$. Define*

$$\Omega := \{\rho_\beta \sim 1\} \cap \pi_1^{-1}(E_1) \cap \pi_2^{-1}(E_2), \qquad \alpha_j := \frac{|\Omega|}{|E_j|}, \ j = 1, 2.$$

If

$$|\Omega| \geq \varepsilon |E_1|^{\frac{1}{p_1}} |E_2|^{\frac{1}{p_2}}, \tag{5.1}$$

there exist a set $\mathcal{A} \subseteq \Omega$ of cardinality $\#\mathcal{A} \lesssim_{c,C} 1$ and a minimal n-tuple $I \in \mathcal{W}^n$ such that
(i)

$$\left|\Omega \cap \bigcup_{x \in \mathcal{A}} \bigcap_{\sigma \in S_n} B^{I_\sigma}(x; c\varepsilon^C \alpha)\right| \gtrsim |\Omega|,$$

(ii) For every $x \in \mathcal{A}$, $\sigma, \sigma' \in S_n$, and $y \in B^{I_\sigma}(x; c\varepsilon^C \alpha)$, $\Phi_y^{I_{\sigma'}}$ is one-to-one with Jacobian determinant

$$|\alpha^{\deg I} \det D\Phi_y^{I_{\sigma'}}| \sim \alpha^{\deg I} |\lambda_I(x)| \approx \alpha^b \approx |\Omega|,$$

on $Q_{c'\varepsilon^{C'} \alpha}^{I_{\sigma'}}$, and, moreover, $B^{I_\sigma}(x, c\varepsilon^C \alpha) \subseteq B^{I_{\sigma'}}(y; c'\varepsilon^{C'} \alpha)$.

By applying Lemma 3.9 with $C'\varepsilon^{-C'} \alpha_1 X_1, C'\varepsilon^{-C'} \alpha_2 X_2$ in place of X_1, X_2, to prove Proposition 5.1, it suffices to prove the following.

Lemma 5.2. *Under the hypotheses of Proposition 5.1, there exist a set $\mathcal{A}$ of cardinality $\#\mathcal{A} \lessapprox 1$ such that*

(i) $|\Omega \cap \bigcup_{x \in \mathcal{A}} B^n(x; C'\varepsilon^{-C'}\alpha)| \gtrsim |\Omega|$

(ii) For every $x \in \mathcal{A}$ and $y \in B^n(x; C'\varepsilon^{-C'}\alpha)$, $\sum_I \alpha^{\deg I}|\lambda_I(y)| \approx \alpha^b$.

Proof of Lemma 5.2. Inequality (5.1) implies, after some arithmetic, that

$$|\Omega| \lessapprox \alpha^b. \tag{5.2}$$

Conversely, the conclusion of Proposition 2.1 is equivalent to $|\Omega| \gtrsim \alpha^b$. We will prove this lemma by essentially repeating the proof of Proposition 2.1, while keeping in mind the constraint (5.2). In the proof, we will extensively use the notations from the proof of Proposition 2.1.

In the proof of Proposition 2.1, we only needed to refine the set Ω n times, but here it will be useful to refine further. Letting $x_0 \in \Omega_{-1} \subseteq \Omega_0$,

$$\widetilde{\Psi}_{x_0}(t) \in \Omega_{n-1}, \text{ if } t_j \in S_{j-1}(\widetilde{\Psi}_{x_0}(t_1, \dots, t_{j-1}, 0)), \ j = 1, \dots, n$$

$$\Psi_{x_0}(t) \in \Omega_n, \text{ if } t_j \in S_j(\Psi_{x_0}(t_1, \dots, t_{j-1}, 0)), \ j = 1, \dots, n.$$

Thus exactly the arguments leading up to (2.7) imply that

$$|\Omega| \gtrsim \sum_{\beta'} \alpha^{b(\beta')}|\partial^{\beta'} \det D\Psi_{x_0}(0)| + \alpha^{\widetilde{b}(\beta')}|\partial^{\beta'} \det D\widetilde{\Psi}_{x_0}(0)|.$$

As was observed in (2.7), the right side above is at least α^b, and by (5.2), it is at most $C'\varepsilon^{-C'}\alpha^b$. Let $\widetilde{\Omega} := \Omega_{-1}$. We have just seen that

$$\sum_{\beta'} \alpha^{b(\beta')}|\partial^{\beta'} \det D\Psi_{x_0}(0)| + \alpha^{\widetilde{b}(\beta')}|\partial^{\beta'} \det D\widetilde{\Psi}_{x_0}(0)| \approx \alpha^b, \qquad x_0 \in \widetilde{\Omega},$$

so by Lemma 3.6,

$$\sum_I \alpha^{\deg I}|\lambda_I(x_0)| \approx \alpha^b, \qquad x_0 \in \widetilde{\Omega}. \tag{5.3}$$

Moreover, by the proof of Proposition 2.1, $|\widetilde{\Omega}| \sim |\Omega| \approx \alpha^b$. Thus the proof of our lemma will be complete if we can cover a large portion of $\widetilde{\Omega}$ using a set $\mathcal{A} \subseteq \widetilde{\Omega}$.

To simplify the notation, we will give the remainder of the argument under the assumption that (5.3) holds on Ω; the general case follows from the same proof, since (5.1) holds with Ω replaced by $\widetilde{\Omega}$. Our next task is to obtain better control over the sets $\mathcal{F}_j$, $S_j(\cdot)$ arising in the proof of Proposition 2.1. We begin by bounding the measure of these sets.

If $|S_n(x)| \geq C'\varepsilon^{-C'}\alpha_n$ for all x in some subset $\Omega' \subseteq \Omega$ with $|\Omega'| \gtrsim |\Omega|$, we could have refined so that $\Omega_{n-1} \subseteq \Omega'$, yielding

$$|\Omega| \gtrsim \alpha_n^{\beta_n}(C'\varepsilon^{-C'}\alpha_n) \int_{\mathcal{F}_{n-1}} |\partial_n^{\beta_n} \det D_t\Psi_{x_0}(t', 0)| \, dt'$$

$$\geq C'\varepsilon^{-C'}\alpha_1^{b_1}\alpha_2^{b_2},$$

a contradiction to (5.2) for C' sufficiently large. Thus we may assume that $|S_n(x)| \lessapprox \alpha_n$ on at least half of Ω, and we may refine so that $|S_n(x)| \lessapprox \alpha_n$ throughout Ω_{n-1}. Similarly, we may refine so that $|S_{n-1}(x)| \lessapprox \alpha_{n-1}$ for each $x \in \Omega_{n-2}$. Thus, by adjusting the refinement procedure at each step, we may assume that for each $1 \leq j \leq n-1$ and each $t \in \mathcal{F}_{j-1}$,

$$|S_j(\Psi_{x_0}(t, 0))| = |\{t_j \in \mathbb{R} : (t, t_j) \in \mathcal{F}_j\}| \lessapprox \alpha_j. \tag{5.4}$$

We have not yet used the gain coming from Proposition 2.2. We will do so now to control the diameter of our parameter set. The key observation is that we may assume that $\sum_{j\,\text{odd}} \beta_j$ and $\sum_{j\,\text{even}} \beta_j$ are both positive. Indeed, this positivity is trivial for $n \geq 4$, because if $t_j = 0$ for any $1 < j < n$, then $\det D\Psi_{x_0}(t) = 0$. Thus the only way our claim can fail is if $n = 3$ and $\beta = (0, k, 0)$, but in this case,

$$\partial^\beta \det D\widetilde{\Psi}_{x_0}(0) = \partial_2 \partial_1^{k-1} \det D\widetilde{\Psi}_{x_0}(0),$$

and we can simply interchange the roles of the indices 1 and 2 throughout the argument.

Let j be the maximal odd index with $\beta_j > 0$. Suppose that on at least half of Ω_j, $|J_j(x)| \geq C'\varepsilon^{-C'}|S_j(x)|$. Then by adjusting our refinement procedure, we may assume that $x \in \Omega_{j-1}$ implies that $|J_j(x)| \geq C'\varepsilon^{-C'}|S_j(x)|$; we note that this implies $|J_j(x)| \geq C'\varepsilon^{-C'}\alpha_j$. In view of (5.4),

$$|\Omega| \gtrsim \alpha_n^{\beta_n+1} \cdots \alpha_{j+1}^{\beta_{j+1}+1} \int_{\mathcal{F}_{j-1}} |\partial_n^{\beta_n} \cdots \partial_j^{\beta_j} \det D\Psi_{x_0}(t', 0)|$$

$$\times \Big(\frac{|J_j(\Psi_{x_0}(t',0))|}{|S_j(\Psi_{x_0}(t',0))|} \Big)^{(1-\delta)j} |S_j(\Psi_{x_0}(t', 0))|^{j+1} \, dt'$$

$$\geq C'\varepsilon^{-C'j} \alpha_1^{b_1} \alpha_2^{b_2}.$$

For C' sufficiently large, this gives a contradiction. Thus on at least half of Ω_j, $|J_j(x)| \lessapprox \alpha_j = \alpha_1$, so we may refine so that for each $x \in \Omega_{j-1}$, $|J_j(x)| \lessapprox \alpha_1$. Repeating this argument for the maximal even index j' with $\beta_{j'} > 0$, we may ensure that for each $x \in \Omega_{j'-1}$, $|J_{j'}(x)| \lessapprox \alpha_2$. Finally, replacing Ω_n with $\Omega_{\min\{j,j'\}-1}$ and then refining, we can ensure that for $x_0 \in \Omega_0$, $1 \leq j \leq n$, and $t \in \mathcal{F}_{j-1}$,

$$|J_j(\Psi_{x_0}(t,0))| = |J(N, \{t_j \in \mathbb{R} : (t, t_j) \in \mathcal{F}_j\})| \lessapprox \alpha_j. \tag{5.5}$$

Refining further, we obtain a set $\Omega_{-n} \subseteq \Omega_0$, with $|\Omega_{-n}| \gtrsim |\Omega|$, such that for each $x_0 \in \Omega_{-n}$, there exists a parameter set

$$\mathcal{F}_{x_0} \subseteq [-C'\varepsilon^{-C'}\alpha_1, C'\varepsilon^{-C'}\alpha_1] \times [-C'\varepsilon^{-C'}\alpha_2, C'\varepsilon^{-C'}\alpha_2] \times \cdots$$

such that

$$\Psi_{x_0}(\mathcal{F}_n) \subseteq \Omega_0 \cap B(x_0; C'\varepsilon^{-C'}\alpha),$$

$$|\Psi_{x_0}(\mathcal{F}_n)| \gtrapprox |B^n(x_0; C'\varepsilon^{-C'}\alpha)|. \tag{5.6}$$

We fix a point $x_0 \in \Omega_{-n}$ and a parameter set $\mathcal{F}_{x_0}$ as above. We add x_0 to $\mathcal{A}$. If (i) holds, we are done. Otherwise, we apply the preceding to

$$\Omega \setminus \bigcup_{x \in \mathcal{A}} B^n(x; C'\varepsilon^{-C'}\alpha),$$

and find another point to add to $\mathcal{A}$. By (5.6) and $|\Omega| \lessapprox \alpha^b$, this process stops while $\#\mathcal{A} \lessapprox 1$.

This completes the proof of Lemma 5.2, and thus of Proposition 5.1 as well. $\square$

6. Strong-type bounds on a single scale

This section is devoted to a proof of the following.

Proposition 6.1.

$$\Big| \int_{\{\rho_\beta \sim 1\}} f_1 \circ \pi_1 \, f_2 \circ \pi_2 \, dx \Big| \lesssim \|f_1\|_{p_1} \|f_2\|_{p_2}.$$

Proof of Proposition 6.1. It suffices to prove the proposition in the special case

$$f_i = \sum_k 2^k \chi_{E_i^k}, \qquad \|f_i\|_{p_i} \sim 1, \qquad i = 1, 2,$$

with the E_1^k pairwise disjoint, and likewise, the E_2^k. Thus we want to bound

$$\sum_{j,k} 2^{j+k} |\Omega^{j,k}|, \qquad \Omega^{j,k} := \{\rho_\beta \sim 1\} \cap \pi_1^{-1}(E_1^j) \cap \pi_2^{-1}(E_2^k).$$

We know from Proposition 2.1 that

$$|\Omega^{j,k}| \lesssim |E_1^j|^{1/p_1} |E_2^k|^{1/p_2}.$$

For $0 < \varepsilon \lesssim 1$, we define

$$\mathcal{L}(\varepsilon) := \{(j,k) : \tfrac{1}{2}\varepsilon |E_1^j|^{1/p_1} |E_2^k|^{1/p_2} \le |\Omega^{j,k}| \le 2\varepsilon |E_1^j|^{1/p_1} |E_2^k|^{1/p_2}\}.$$

We additionally define for $0 < \eta_1, \eta_2 \le 1$,

$$\mathcal{L}(\varepsilon, \eta_1, \eta_2) := \{(j,k) \in \mathcal{L}(\varepsilon) : 2^{jp_1}|E_1^j| \sim \eta_1, \ 2^{kp_2}|E_2^k| \sim \eta_2\}.$$

Let $\varepsilon, \eta_1, \eta_2 \lesssim 1$ and let $(j,k) \in \mathcal{L}(\varepsilon, \eta_1, \eta_2)$. Set

$$\alpha^{j,k} = (\alpha_1^{j,k}, \alpha_2^{j,k}) := \big(\tfrac{|\Omega^{j,k}|}{|E_1^j|}, \tfrac{|\Omega^{j,k}|}{|E_2^k|} \big).$$

Proposition 5.1 guarantees the existence of a minimal $I \in \mathcal{W}^n$ and a finite set $\mathcal{A}^{j,k} \subseteq \Omega^{j,k}$ such that (i) and (ii) of that proposition (appropriately superscripted) hold. (Since there are a bounded number of minimal n-tuples, we may assume in proving the proposition that all of these minimal n-tuples are the same.) Set

$$\widetilde{\Omega}^{j,k} := \Omega^{j,k} \cap \bigcup_{x \in \mathcal{A}^{j,k}} \bigcap_{\sigma \in S_n} B^{I_\sigma}(x, c\varepsilon^C \alpha^{j,k}). \tag{6.1}$$

Our main task in this section is to prove the following lemma.

Lemma 6.2. *Fix $\varepsilon, \eta_1, \eta_2 \lesssim 1$ and set $\mathcal{L} := \mathcal{L}(\varepsilon, \eta_1, \eta_2)$. Then*

$$\sum_{k : (j,k) \in \mathcal{L}} |\pi_1(\widetilde{\Omega}^{j,k})| \lesssim (\log \varepsilon^{-1}) |E_1^j|, \qquad j \in \mathbb{Z} \tag{6.2}$$

$$\sum_{j : (j,k) \in \mathcal{L}} |\pi_2(\widetilde{\Omega}^{j,k})| \lesssim (\log \varepsilon^{-1}) |E_2^k|, \qquad k \in \mathbb{Z}. \tag{6.3}$$

We assume Lemma 6.2 for now and complete the proof of Proposition 6.1. It suffices to show that for each $\varepsilon, \eta_1, \eta_2$, if $\mathcal{L} := \mathcal{L}(\varepsilon, \eta_1, \eta_2)$, then

$$\sum_{(j,k) \in \mathcal{L}} 2^{j+k} |\Omega^{j,k}| \lesssim \varepsilon^{a_0} \eta_1^{a_1} \eta_2^{a_2}, \tag{6.4}$$

with each a_i positive. Indeed, once we have proved the preceding inequality, we can just sum on dyadic values of $\varepsilon, \eta_1, \eta_2$.

We turn to the proof of (6.4). It is a triviality that $\#\mathcal{L}(\varepsilon, \eta_1, \eta_2) \lesssim \eta_1^{-1} \eta_2^{-1}$, so

$$\sum_{(j,k) \in \mathcal{L}} 2^{j+k} |\Omega^{j,k}| \sim \varepsilon \sum_{(j,k) \in \mathcal{L}} 2^{j+k} |E_1|^{1/p_1} |E_2|^{1/p_2}$$
$$\sim \varepsilon (\#\mathcal{L}) \eta_1^{1/p_1} \eta_2^{1/p_2} \lesssim \varepsilon \eta_1^{-1/p_1'} \eta_2^{-1/p_2'}. \tag{6.5}$$

Define

$$q_i := (p_1^{-1} + p_2^{-1}) p_i, \qquad i = 1, 2,$$

then since
$$p_1^{-1} + p_2^{-1} = \frac{b_1 + b_2}{b_1 + b_2 - 1} > 1,$$

we have $q_i > p_i$, $i = 1, 2$, and $q_1 = q_2'$. Applying Lemma 6.2,

$$\sum_{(j,k)\in\mathcal{L}} 2^{j+k}|\Omega^{j,k}| \sim \sum_{(j,k)\in\mathcal{L}} 2^{j+k}|\widetilde{\Omega}^{j,k}| \lesssim \sum_{(j,k)\in\mathcal{L}} 2^{j+k}|\pi_1(\widetilde{\Omega}^{j,k})|^{1/p_1}|\pi_2(\widetilde{\Omega}^{j,k})|^{1/p_2}$$

$$\lesssim \Big(\sum_{(j,k)\in\mathcal{L}} 2^{jq_1}|\pi_1(\widetilde{\Omega}^{j,k})|^{q_1/p_1} \Big)^{1/q_1} \Big(\sum_{(j,k)\in\mathcal{L}} 2^{kq_2}|\pi_2(\widetilde{\Omega}^{j,k})|^{q_2/p_2} \Big)^{1/q_2}$$

$$\lesssim \eta_1^{1/p_1 - 1/q_1} \eta_2^{1/p_2 - 1/q_2}$$
$$\times \Big(\sum_{(j,k)\in\mathcal{L}} 2^{jp_1}|\pi_1(\widetilde{\Omega}^{j,k})| \Big)^{1/q_1} \Big(\sum_{(j,k)\in\mathcal{L}} 2^{kp_2}|\pi_2(\widetilde{\Omega}^{j,k})| \Big)^{1/q_2}$$

$$\lesssim \log\varepsilon^{-1}\eta_1^{1/p_1 - 1/q_1} \eta_2^{1/p_2 - 1/q_2} \Big(\sum_j 2^{jp_1}|E_1^j|\Big)^{1/q_1} \Big(\sum_k 2^{kp_2}|E_2^k|\Big)^{1/q_2}$$

$$\lesssim \log\varepsilon^{-1}\eta_1^{1/p_1 - 1/q_1} \eta_2^{1/p_2 - 1/q_2}.$$

Combining this estimate with (6.5) gives (6.4), completing the proof of Proposition 6.1, conditional on Lemma 6.2. □

We turn to the proof of Lemma 6.2. We will only prove (6.2), and we will take care that our argument can be adapted to prove (6.3) by interchanging the indices. (The roles of π_1 and π_2 are not *a priori* symmetric, because their roles in defining the weight ρ are not symmetric.) The argument is somewhat long and technical, so we start with a broad overview.

Assume that (6.2) fails. By Proposition 5.1, the $\widetilde{\Omega}^{j,k}$ can be well approximated as the images of ellipsoids (the $Q_{\alpha^{j,k}}^I$) under polynomials of bounded degree (the $\Phi_{x^{j,k}}^I$). The definition of $\mathcal{L}$ ensures that the $\alpha^{j,k}$, and hence the radii of these ellipsoids, live at many different dyadic scales (this is where the minimality condition in Proposition 5.1 will be used). On the other hand, the projections $\pi_1(\widetilde{\Omega}^{j,k})$ must have a large degree of overlap (otherwise, the volume of the union would bound the sum of the volumes). In particular, we can find a large number of $\widetilde{\Omega}^{j,k}$ that all have essentially the same projection. These $\widetilde{\Omega}^{j,k}$ all lie along a single integral curve of X_1. The shapes of the $\widetilde{\Omega}^{j,k}$ are determined by widely disparate parameters, the $\alpha^{j,k}$, and polynomials, the $\Phi_{x^{j,k}}^I$. We can take $x^{j,k} = e^{t^{j,k}X_1}(x_0)$, for a fixed x_0, and we use the condition that the projections are all essentially the same to prove that there exists an associated polynomial $\gamma : \mathbb{R} \to \mathbb{R}^n$ that is transverse to its derivative γ' at more scales than Lemma 11.5 allows.

We begin by making precise the assertion that many $\widetilde{\Omega}^{j,k}$ must have essentially the same projection. The main step is an elementary lemma.

Lemma 6.3. *Let $\{E^k\}$ be a collection of measurable sets, and define $E := \bigcup_k E^k$. Then for each integer $M \geq 1$,*

$$\sum_k |E^k| \lesssim_M |E| + |E|^{\frac{M-1}{M}} \Big(\sum_{k_1 < \cdots < k_M} |E^{k_1} \cap \cdots \cap E^{k_M}| \Big)^{\frac{1}{M}}. \tag{6.6}$$

Proof of Lemma 6.3. We review the argument in the case $M = 2$, which amounts to a rephrasing of an argument from [4]. By Cauchy–Schwarz,

$$\sum_k |E^k| = \int_E \sum_k \chi_{E_k} \le |E|^{\frac{1}{2}} \Big(\int_E |\sum_k \chi_{E_k}|^2 \Big)^{\frac{1}{2}}$$

$$= |E|^{\frac{1}{2}} \Big(\sum_k |E_k| + 2 \sum_{k_1 < k_2} |E_{k_1} \cap E_{k_2}| \Big)^{\frac{1}{2}}$$

$$\le \tfrac{1}{2} \sum_k |E_k| + \tfrac{1}{2}|E| + 2|E|^{\frac{1}{2}} \Big(\sum_{k_1 < k_2} |E^{k_1} \cap E^{k_2}| \Big)^{\frac{1}{2}},$$

and inequality (6.6) follows by subtracting $\frac{1}{2} \sum_k |E^k|$ from both sides.

Now to the case of larger M. Arguing analogously to the $k = 2$ case implies that

$$\sum_k |E^k| \lesssim_M |E|^{\frac{M-1}{M}} \Big(\sum_{i=1}^{M} \sum_{k_1 < ... < k_i} |\bigcap_{l=1}^{i} E^{k_l}| \Big)^{\frac{1}{M}}. \tag{6.7}$$

Suppose that (6.6) is proved for $2, \dots, M-1$. Let $1 < i < M$. For fixed $k_1 < \cdots < k_{i-1}$,

$$\sum_{k_i} |E^{k_1} \cap \cdots \cap E^{k_i}| \lesssim_M |E^{k_1} \cap \cdots \cap E^{k_{i-1}}|$$

$$+ |E^{k_1} \cap \cdots \cap E^{k_{i-1}}|^{\frac{M-i-1}{M-i}} \Big(\sum_{k_i < \cdots < k_M} |E^{k_1} \cap \cdots \cap E^{k_M}| \Big)^{\frac{1}{M-i-1}}$$

$$\lesssim_M |E^{k_1} \cap \cdots \cap E^{k_{i-1}}| + \sum_{k_i < \cdots < k_M} |E^{k_1} \cap \cdots \cap E^{k_M}|.$$

By induction and (6.7),

$$\sum_k |E^k| \lesssim_M |E|^{\frac{M-1}{M}} \Big(\sum_k |E^k| + \sum_{k_1 < \cdots < k_M} |E^{k_1} \cap \cdots \cap E^{k_M}| \Big)^{\frac{1}{M}},$$

which implies (6.6). $\qquad\square$

Our next goal is to reduce the proof of Lemma 6.2, specifically, the proof of (6.2) to the following.

Lemma 6.4. *For $M > M(N)$ sufficiently large and each $A > 0$, there exists $B > 0$ such that for all $0 < \delta \le \varepsilon$, if $j_0 \in \mathbb{Z}$ and $\mathcal{K} \subseteq \mathbb{Z}$ is a $(B \log \delta^{-1})$-separated set with cardinality $\#\mathcal{K} \ge M$ and $\{j_0\} \times \mathcal{K} \subseteq \mathcal{L}$, then*

$$|\bigcap_{k \in \mathcal{K}} \pi_1(\widetilde{\Omega}^{j_0,k})| < A^{-1} \delta^A 2^{-j_0 p_1} \eta_1. \tag{6.8}$$

Proof of Lemma 6.2, conditional on Lemma 6.4. We will only prove inequality (6.2). The obvious analogue of Lemma 6.4, which has the same proof as Lemma 6.4, implies inequality (6.3).

Fix $M = M(N)$ sufficiently large to satisfy the hypotheses of Lemma 6.4 and fix $A > Mp_1$. Now fix $B = B(M, N, A)$ as in the conclusion of Lemma 6.4. Let

$\delta := \min\{\delta_0, \varepsilon\}$, with δ_0 to be determined, and let $\mathcal{K}_0 \subseteq \mathbb{Z}$ be a finite $(B \log \delta^{-1})$-separated set with $\{j_0\} \times \mathcal{K}_0 \subseteq \mathcal{L}$. By Lemma 6.3, Lemma 6.4, then the approximation $\binom{\#\mathcal{K}_0}{M} \sim_M (\#\mathcal{K}_0)^M$ and the definition of $\mathcal{L}$,

$$\sum_{k \in \mathcal{K}_0} |\pi_1(\widetilde{\Omega}^{j_0,k})| \lesssim_M |E_1^{j_0}| + |E_1^{j_0}|^{\frac{M-1}{M}} \big(\sum_{\mathcal{K} \subseteq \mathcal{K}_0; \#\mathcal{K}=M} |\bigcap_{k \in \mathcal{K}} \pi_1(\widetilde{\Omega}^{j_0 k})| \big)^{\frac{1}{M}}$$
$$\lesssim_M |E_1^{j_0}| + |E_1^{j_0}|^{\frac{M-1}{M}} \#\mathcal{K}_0 (A^{-1} \delta^A |E_1^{j_0}|)^{\frac{1}{M}}. \tag{6.9}$$

Quasiextremality and the restricted weak type inequality give

$$\delta |E_1^{j_0}|^{\frac{1}{p_1}} |E_2^k|^{\frac{1}{p_2}} \lesssim |\Omega^{j_0 k}| \sim |\widetilde{\Omega}^{j_0 k}| \lesssim |\pi_1(\widetilde{\Omega}^{j_0 k})|^{\frac{1}{p_1}} |E_2^k|^{\frac{1}{p_2}}, \quad k \in \mathcal{K}_0$$

whence

$$\sum_{k \in \mathcal{K}_0} |\pi_1(\widetilde{\Omega}^{j_0 k})| \gtrsim \#\mathcal{K}_0 \delta^{p_1} |E_1^{j_0}|. \tag{6.10}$$

For $\delta_0 = \delta_0(p_1, A, M)$ sufficiently small, $\delta^{p_1} > C_M (A^{-1} \delta^A)^{\frac{1}{M}}$, with C_M as large as we like, so inserting (6.10) into (6.9) implies

$$\sum_{k \in \mathcal{K}_0} |\pi_1(\widetilde{\Omega}^{j_0,k})| \lesssim_M |E_1^{j_0}|. \tag{6.11}$$

Since $\mathcal{K}_0$ was arbitrary and p_1, M, A, B all ultimately depend only on N alone, (6.11) implies (6.2). $\qquad \square$

It remains to prove Lemma 6.4.

Lemma 6.5. *For $M = M(N)$ sufficiently large and each $A > 0$, there exists $B > 0$ such that the following holds for all $0 < \delta \leq \varepsilon$. Fix $j_0 \in \mathbb{Z}$ and let $\mathcal{K} \subseteq \mathbb{Z}$ be a $(B \log \delta^{-1})$-separated set with cardinality $\#\mathcal{K} = M$ and $\{j_0\} \times \mathcal{K} \subseteq \mathcal{L}$. Let $x^{j_0 k} \in \widetilde{\Omega}^{j_0 k}$, $k \in \mathcal{K}$. Then*

$$|\bigcap_{k \in \mathcal{K}} \pi_1 \big(\bigcap_{\sigma \in S_n} B^{I_\sigma}(x^{j_0,k}, c\delta^C \alpha^{j_0,k}) \big)| < A^{-1} \delta^A 2^{-j_0 p_1} \eta_1. \tag{6.12}$$

We note that once the lemma holds for $M = M(N)$, it immediately holds for all $M > M(N)$ as well.

Proof of Lemma 6.4, conditional on Lemma 6.5. By definition (6.1), each $\widetilde{\Omega}^{j_0 k}$ is covered by $C\varepsilon^{-C}$ balls of the form $\bigcap_{\sigma \in S_n} B^{I_\sigma}(x, c\varepsilon^C \alpha^{j_0,k})$; in fact, by the proof of Proposition 5.1, it is also covered by $C\delta^{-C}$ balls $\bigcap_{\sigma \in S_n} B^{I_\sigma}(x, c\delta^C \alpha^{j_0,k})$, for each $0 < \delta \leq \varepsilon$. Thus $\bigcap_{k \in \mathcal{K}} \pi_1(\widetilde{\Omega}^{j_0,k})$ is covered by $(C\delta^{-C})^M$ M-fold intersections of projections of balls, so (6.12) (with a larger value of A) implies (6.8). $\qquad \square$

The remainder of the section will be devoted to the proof of Lemma 6.5. We will give the proof when $\delta = \varepsilon$; since an ε-quasiextremal $\Omega^{j_0 k}$ is also δ-quasiextremal for every $0 < \delta < \varepsilon$, all of our arguments below apply equally well in the case $\delta < \varepsilon$. (We recall that allowing the more general parameter δ instead of ε gave us slightly more technical flexibility in the proof of Lemma 6.2 from Lemma 6.4.)

The potential failure of π_1 to be a polynomial presents a technical complication. (Coordinate changes are not an option in the non-minimal case.) By reordering the words in $I = (w_1, \ldots, w_n)$, we may assume that $w_n = (1)$. Fix $k_0 \in \mathcal{K}$, and set $x_0 = x^{j_0 k_0}$. We define a "cylinder"

$$\mathcal{C} := \Phi_{x_0}^I(U), \qquad U := \{(t', t_n) : (t', 0) \in Q_{c\varepsilon^C \alpha^{j_0 k_0}}^I\}.$$

Set $U_0 := \{(t', 0) \in U\}$ and define

$$U_+ := \{t \in U : t_n > 0 \text{ and, for all } 0 < s \leq t_n, \ \Phi_{x_0}^I(t', s) \notin \Phi_{x_0}^I(\overline{U}_0)\},$$

$$U_- := \{t \in U : t_n < 0 \text{ and, for all } 0 > s \geq t_n, \ \Phi_{x_0}^I(t', s) \notin \Phi_{x_0}^I(\overline{U}_0)\},$$

and $\mathcal{C}_0 := \Phi_{x_0}^I(\overline{U}_0)$, $\mathcal{C}_\pm := \Phi_{x_0}^I(U_\pm)$.

Lemma 6.6. *The map $\Phi_{x_0}^I$ is nonsingular, with*

$$|\det D\Phi_{x_0}^I| \approx \lambda_I(x_0) \approx (\alpha^{j_0 k_0})^{b - \deg I},$$

on U. The sets $U_\pm$ are open, and $\Phi_{x_0}^I$ is one-to-one on each of them. Finally,

$$\mathcal{C} \subseteq \mathcal{C}_+ \cup \mathcal{C}_- \cup \mathcal{C}_0 \cup \mathcal{C}_\partial,$$

where $\mathcal{C}_\partial := \Phi_{x_0}^I(\partial U)$.

Proof of Lemma 6.6. Since X_1 is divergence-free,

$$\det D\Phi_{x_0}^I(t) = \det D\Phi_{x_0}^I(t', 0),$$

and thus the conclusions about the size of this Jacobian determinant follow from Proposition 5.1.

By conclusion (ii) of Proposition 5.1 and continuity of $\Phi_{x_0}^I$, we see that $U_\pm$ is an open set containing

$$\{(t', t_n) : (t', 0) \in Q_{c\varepsilon^C \alpha^{j_0 k_0}}, 0 < \pm t_n < c\varepsilon^C \alpha_1^{j_0 k_0}\}.$$

Suppose $t, u \in U_+$, $\Phi_{x_0}^I(t) = \Phi_{x_0}^I(u)$, and $u_n \leq t_n$. Then $\Phi_{x_0}^I(t', t_n - u_n) = \Phi_{x_0}^I(u', 0)$. If $t_n = u_n$, then $t = u$, because $\Phi_{x_0}^I$ is one-to-one on $Q_{c\varepsilon^C \alpha^{j_0 k_0}} \ni (u', 0), (t', 0)$. Otherwise, $(t', t_n - u_n) \in U_+$, so $\Phi_{x_0}^I(t', t_n - u_n) = \Phi_{x_0}^I(u', 0)$ is impossible. Thus $\Phi_{x_0}^I$ is indeed one-to-one on $U_\pm$.

Finally, let $t \in U$, with $t_n > 0$ and $t \notin U_+$. We need to show that $\Phi_{x_0}^I(t) \in \mathcal{C}_+ \cup \mathcal{C}_0 \cup \mathcal{C}_\partial$. The curve $\{\Phi_{x_0}^I(t', s) : s \in \mathbb{R}\}$ intersects $\mathcal{C}_0$ a bounded number of times, so, by the definition of U_+, there exists some maximal $0 < s \leq t_n$ such that $\Phi_{x_0}^I(t', s) = \Phi_{x_0}^I(u', 0)$, for some $(u', 0) \in \overline{U}_0$. Thus $\Phi_{x_0}^I(t', t_n) = \Phi_{x_0}^I(u', t_n - s)$. If $s = t_n$, $(u', 0) \in \partial U$, or $(u', t_n - s) \in U_+$, we are done. Otherwise, there exists some $0 < r < t_n - s$ and $(v', 0) \in \overline{U}_0$ such that $\Phi_{x_0}^I(u', r) = \Phi_{x_0}^I(v', 0)$, whence $\Phi_{x_0}(t', s + r) = \Phi_{x_0}^I(v', 0)$, contradicting maximality of s. $\qquad\square$

On $U_\pm$, $\Phi_{x_0}^I$ has a smooth inverse, and we define a map $\widetilde{\pi}_1$ on $\mathcal{C}_\pm$ by

$$\widetilde{\pi}_1 := ((\alpha^{j_0 k_0})^{-\deg w_1}(\Phi_{x_0}^I)_1^{-1}, \ldots, (\alpha^{j_0 k_0})^{-\deg w_{n-1}}(\Phi_{x_0}^I)_{n-1}^{-1}).$$

Lemma 6.7. *Define*

$$F(t') := \pi_1 \circ \Phi_{x_0}^I \big((c\varepsilon^C \alpha^{j_0 k_0})^{\deg w_1} t_1, \cdots, (c\varepsilon^C \alpha^{j_0 k_0})^{\deg w_{n-1}} t_{n-1}, 0\big), \qquad |t'| < 1.$$

Then F is a bounded-to-one local diffeomorphism satisfying $|\det DF| \approx |E_1^{j_0}|$ and $\pi_1|_{\mathcal{C}_\pm} = F \circ \widetilde{\pi}_1|_{\mathcal{C}_\pm}$.

Proof of Lemma 6.7. Let

$$Q_{c\varepsilon^C \alpha^{j_0 k_0}}^\flat := \{t' : (t', 0) \in Q_{c\varepsilon^C \alpha^{j_0 k_0}}\}.$$

We begin by proving that $t' \mapsto \pi_1 \circ \Phi_{x_0}^I(t', 0)$ is bounded-to-one on $Q_{c\varepsilon^C \alpha^{j_0 k_0}}^\flat$. By the implicit function theorem, the definition of X_1, and hypothesis (ii) of Theorem 1.1, for every $y \in \mathbb{R}^{n-1}$, $\pi_1^{-1}(y)$ intersects at most one nonconstant integral curve of X_1. Therefore, since X_1 is nonvanishing on $\mathcal{C}$, if $\pi_1 \circ \Phi_{x_0}^I(u', 0) = \pi_1 \circ \Phi_{x_0}^I(t', 0)$

for some $u' \neq t'$, we may assume that $\Phi^I_{x_0}(t', 0) = \Phi^I_{x_0}(u', u_n)$ for some $u_n > 0$. By Lemma 11.7 and $\det D\Phi^I_{x_0} \neq 0$ on U, given $t' \in Q^\flat_{c\varepsilon^C \alpha^{j_0 k_0}}$, there are only a bounded number of such u'.

Our definition of $\widetilde{\pi}_1$ implies that $\pi_1 = F \circ \widetilde{\pi}_1$ on $\mathcal{C}_0$, and hence on all of $\mathcal{C}$ (since both sides are constant on X_1's integral curves).

Let $A \subseteq \{|t'| < 1\}$. Then

$$B := \widetilde{\pi}_1^{-1}(A) \cap \{\Phi^I_{x_0}(t) : t \in U, \ |t_n| < c\varepsilon^C \alpha^{j_0 k_0}_1\}$$

equals the image

$$\{\Phi^I_{x^0}(t) : ((c\varepsilon^C \alpha^{j_0 k_0})^{-\deg w_1} t_1, \ldots, (c\varepsilon^C \alpha^{j_0 k_0})^{-\deg w_{n-1}} t_n) \in A, \ |t_n| < c\varepsilon^C \alpha^{j_0 k_0}_1\},$$

and hence, by Proposition 5.1, has volume

$$|B| \approx (\alpha^{j_0 k_0})^{\deg I} |\lambda_I(x^0)| |A| \approx |\Omega^{j_0 k_0}| |A|. \tag{6.13}$$

By the definition of B, the coarea formula, (6.13), and the definition of $\alpha^{j_0 k_0}$,

$$|F(A)| = |\pi_1(B)| \approx (\alpha^{j_0 k_0}_1)^{-1} |B| \approx (\alpha^{j_0 k_0}_1)^{-1} |\Omega^{j_0 k_0}| |A| = |E^{j_0}_1| |A|.$$

The estimate on the Jacobian determinant of F follows from the change of variables formula. $\qquad\square$

The next lemma allows us to replace $\mathcal{C}$ with the domain of $\widetilde{\pi}_1$.

Lemma 6.8. *If (6.12) fails for some $M, A, B, \delta = \varepsilon > 0, j_0, \mathcal{K}, \{x^{j_0 k}\}_{k \in \mathcal{K}}$ satisfying the hypothesis of Lemma 6.5, then there exists $\mathcal{K}' \subseteq \mathcal{K}$, of cardinality $\#\mathcal{K}' \sim M$, such that*

$$| \bigcap_{k \in \mathcal{K}'} \widetilde{\pi}_1(\mathcal{C}_+ \cap \bigcap_{\sigma \in S_n} B^{I_\sigma}(x^{j_0, k}, c\varepsilon^C \alpha^{j_0 k}))| \geq A^{-1} \varepsilon^A, \tag{6.14}$$

or such that (6.14) holds with '$-$' in place of '$+$.' Here the quantity A depends on the corresponding quantity in Lemma 6.5 and N.

Proof of Lemma 6.8. For $k \in \mathcal{K}$, set

$$B^k := \bigcap_{\sigma \in S_n} B^{I_\sigma}(x^{j_0 k}; c\varepsilon^C \alpha^{j_0 k}).$$

Since $\pi_1(B^{k_0}) \subseteq \pi_1(B^I(x, c\varepsilon^C \alpha^{j_0 k_0}))$, our hypothesis that π_1 fibers lie on a single integral curve of X_1 implies that $\bigcap_{k \in \mathcal{K}} \pi_1(B^k) = \bigcap_{k \in \mathcal{K}} \pi_1(\mathcal{C} \cap B^k)$. The projection $\pi_1(\mathcal{C}_\partial)$ has measure zero. For a.e. $y \in \bigcap_{k \in \mathcal{K}} \pi_1(B^k)$, $\pi_1^{-1}(y) \cap B^k = \Phi^I_{x^0}(t'_0, J)$ for some set $J \subseteq \mathbb{R}$ having positive measure; thus, $|\bigcap_{k \in \mathcal{K}} \pi_1(B^k)| = |\bigcap_{k \in \mathcal{K}} \pi_1(B^k \backslash \mathcal{C}_0)|$. Putting these two observations together with Lemma 6.6 and using standard set manipulations,

$$| \bigcap_{k \in \mathcal{K}} \pi_1(\mathcal{C} \cap B^k)| = | \bigcup_{\bullet \in \{+,-\}^{\mathcal{K}}} \bigcap_{k \in \mathcal{K}} \pi_1(\mathcal{C}_{\bullet_k} \cap B^k)|.$$

Thus if (6.12) fails, there exists a decomposition $\mathcal{K} = \mathcal{K}_+ \cup \mathcal{K}_-$ such that

$$\min\{| \bigcap_{k \in \mathcal{K}_+} \pi_1(\mathcal{C}_+ \cap B^k)|, | \bigcap_{k \in \mathcal{K}_-} \pi_1(\mathcal{C}_- \cap B^k)|\} > A^{-1} \varepsilon^A 2^{-j_0 p_1} \eta_1.$$

One of $\mathcal{K}_+, \mathcal{K}_-$ must have cardinality $\#\mathcal{K}_\bullet \gtrsim M$; we may assume that the larger is $\mathcal{K}_+ =: \mathcal{K}'$. Inequality (6.14) then follows from Lemma 6.7 and the definition of $\mathcal{L}$. $\qquad\square$

The next lemma verifies that a slightly enlarged version of each B^k has large intersection with $\mathcal{C}_+$.

Lemma 6.9. *Assume that (6.14) holds, and let $k \in \mathcal{K}'$, $y^{j_0 k} \in B^k$, with B^k defined as in the proof of Lemma 6.8. Set*

$$G^k := \mathcal{C}_+ \cap \bigcap_{\sigma \in S_n} B^{I_\sigma}(y^{j_0 k}; c' \varepsilon^{C'} \alpha^{j_0 k}).$$

Then $|G^k| \gtrsim A^{-1} \varepsilon^A |B^k|$ and $B^k \cap \mathcal{C}_+ \subseteq G^k$.

Proof of Lemma 6.9. By conclusion (ii) of Proposition 5.1, $B^k \cap \mathcal{C}_+ \subseteq G^k$. Let $x \in B^k \cap \mathcal{C}_+$. So long as $x \notin \mathcal{C}_\partial$ (which has measure zero), $e^{tX_1}(x) \in \mathcal{C}_+$ for all except finitely many positive values of t (i.e. except for those t for which $e^{tX_1}(x) \in \mathcal{C}_0$). Additionally,

$$e^{tX_1}(x) \in \bigcap_{\sigma \in S_n} B^{I_\sigma}(y^{j_0 k}, c' \varepsilon^{C'} \alpha^{j_0 k}),$$

for all $|t| < c\varepsilon^C \alpha^{j_0 k}$, so $e^{tX_1}(x) \in G^k$ for t in a set of measure $\gtrsim \alpha^{j_0 k}$. By the coarea formula, then Lemma 6.7 and (6.14), the definition of $\alpha^{j_0 k}$, and finally Proposition 5.1,

$$|G^k| \gtrsim \alpha_1^{j_0 k} |\pi_1(B^k \cap \mathcal{C}_+)| \approx A^{-1} \varepsilon^A \alpha_1^{j_0 k} |E_1^{j_0}| \approx A^{-1} \varepsilon^A |\Omega^{j_0 k}| \approx A^{-1} \varepsilon^A |B^k|.$$

$\square$

To motivate the next lemma, we recall that our goal is to show that a certain inequality holds at many points of the form $e^{t^{jk} X_1}(x_0) \in B^k$. This will be possible because the set of $y^{jk} \in B^k$ at which the inequality fails must be very small, and hence have small projection.

Lemma 6.10. *Under the hypotheses and notation of Lemma 6.9, there exists a subset $\widetilde{G}^k \subseteq G^k$ such that $|G^k \setminus \widetilde{G}^k| < D^{-1} \varepsilon^D |G^k|$, with $D = D(N, A)$ sufficiently large for later purposes, such that for all $x \in \widetilde{G}^k$,*

$$|D\widetilde{\pi}_1(x)(\alpha^{j_0 k})^{\deg w_i} X_{w_i}(x)| \lesssim_A 1, \qquad 1 \leq i \leq n - 1. \tag{6.15}$$

Proof of Lemma 6.10. To simplify our notation somewhat, we will say that a subset $\widetilde{G}^k \subseteq G^k$ constitutes the *vast majority* of G^k if $|G^k \setminus \widetilde{G}^k| < D^{-1} \varepsilon^D |G^k|$, with $D = D(N, A)$ as small as we like.

Taking intersections, it suffices to establish the lemma for a single index $1 \leq i \leq n - 1$. We recall that $w_i \neq (1)$. Fix a permutation $\sigma \in S_n$ such that $\sigma(n) = i$. By construction, $G^k \subseteq B^{I_\sigma}(x^{j_0 k}, c' \varepsilon^{C'} \alpha^{j_0 k})$. By Lemma 6.9,

$$|B^{I_\sigma}(x^{j_0 k}, c' \varepsilon^{C'} \alpha^{j_0 k}) \cap \mathcal{C}_+| \gtrsim_A |B^{I_\sigma}(x^{j_0 k}; c' \varepsilon^{C'} \alpha^{j_0 k})|,$$

so our Jacobian bound, $|\det D\Phi_{x^{j_0 k}}^{I_\sigma}| \approx |\lambda_I(x^{j_0 k})|$ on $Q_{c' \varepsilon^{C'} \alpha^{j_0 k}}^{I_\sigma}$, implies that for the vast majority of points $x \in G^k$, $e^{tX_w}(x) \in \mathcal{C}_+$ for all $t \in E_x$, E_x some set of measure $|E_x| \gtrsim_A (\alpha^{j_0 k})^{\deg w_i}$.

By Lemmas 11.8 and 11.9, E_x can be written as a union of a bounded number of intervals on which each component of $\frac{d}{dt} \widetilde{\pi}_1(e^{tX_{w_i}}(x))$ is single signed. Thus, using the semigroup property of exponentiation, we see that for the vast majority of $x \in G^k$, there exists an interval $J_x \ni 0$, of length $|J_x| \gtrsim_A (\alpha^{j_0 k})^{\deg w_i}$, such that $e^{tX_{w_i}}(x) \in \mathcal{C}_+$ and the components of $\frac{d}{dt} \widetilde{\pi}_1(e^{tX_{w_i}}(x))$ do not change sign on J_x.

Let $x \in G^k$ be one of these majority points. By the Fundamental Theorem of Calculus and

$$\widetilde{\pi}_1(e^{tX_{w_i}}(x)) \subseteq \widetilde{\pi}_1(\mathcal{C}_+) \subseteq \{u \in \mathbb{R}^{n-1} : |u| < 1\}, \qquad t \in J_x,$$

combined with the above non-sign-changing condition,

$$\int_{J_x} |\tfrac{d}{dt}\widetilde{\pi}_1(e^{tX_{w_i}}(x))| \, dt \sim |\int_{J_x} \tfrac{d}{dt}\widetilde{\pi}_1(e^{tX_{w_i}}(x)) \, dt| < 1.$$

Thus on the vast majority of J_x,

$$|\tfrac{d}{dt}\widetilde{\pi}_1(e^{tX_{w_i}}(x))| \lesssim_A |J_x|^{-1} \lesssim_A (\alpha^{j_0 k})^{-\deg w_i}.$$

The conclusion of the lemma follows from the Chain Rule and our Jacobian estimate on $\Phi^{I_\sigma}_{x^{j_0 k}}$. $\qquad\square$

Finally, we come to the main step in deriving the promised contradiction.

Lemma 6.11. *Under the hypotheses and notation of Lemma 6.10, there exist a point $y^0 \in B^{k_0}$ and times $t^{j_0 k} \in \mathbb{R}$, $k \in \mathcal{K}'$ such that for any $1 \le j \le n$ and any choice of $1 \le i_1 < \cdots < i_j \le n-1$ and $k \in \mathcal{K}'$,*

$$|\bigwedge_{l=1}^{j} D\widetilde{\pi}_1(e^{t^{j_0 k}X_1}(y^0))(\alpha^{j_0 k})^{\deg w_{i_l}} X_{w_{i_l}}(e^{t^{j_0 k}X_1}(y^0))| \approx_A 1. \qquad (6.16)$$

Proof of Lemma 6.11. Let $\widetilde{G}^k$ be as in Lemma 6.10. For $k \in \mathcal{K}'$, and $D = D(N,A)$ as large as we like,

$$|\widetilde{\pi}_1(G^k \setminus \widetilde{G}^k)| \lesssim (\alpha_1^{j_0 k})^{-1}|G^k \setminus \widetilde{G}^k| < D^{-1}\varepsilon^D(\alpha_1^{j_0 k})^{-1}|G^k| \lesssim D^{-1}\varepsilon^D.$$

Thus if $D = D(N,A)$ is sufficiently large,

$$|\bigcup_{k \in \mathcal{K}'} \widetilde{\pi}_1(G^k \setminus \widetilde{G}^k)| < \tfrac{1}{2}A^{-1}\varepsilon^A,$$

so $\bigcap_{k \in \mathcal{K}'} \widetilde{\pi}_1(\widetilde{G}^k)$ is nonempty. Thus there exists a point $y^0 \in \widetilde{G}^{k_0}$ and times $\{t^{j_0 k}\}$ such that $e^{t^{j_0 k}X_1}(y^0) \in \widetilde{G}^k$, $k \in \mathcal{K}'$. We may assume that $t^{j_0 k_0} = 0$, and we set $y^k := e^{t^{j_0 k}X_1}(y^0)$.

By the Chain Rule and basic linear algebra, and then our Jacobian estimate on $\det D\Phi^I_{x^0}$,

$$|\bigwedge_{j=1}^{n-1} D\widetilde{\pi}_1(y^k)X_{w_j}(y^k)|$$

$$= \alpha_1^{j_0 k_0}(\alpha^{j_0 k_0})^{-\deg I}|\det D(\Phi^I_{x^0})^{-1}(y^k)||\det(X_{w_1}(y^k),\ldots,X_{w_n}(y^k))|$$

$$\sim \alpha_1^{j_0 k_0}(\alpha^{j_0 k_0})^{-\deg I}\frac{|\lambda_I(y^k)|}{|\lambda_I(x^0)|}.$$

By (ii) of Proposition 5.1 and the definition of α^{jk},

$$(\alpha^{j_0 k})^{\deg I}|\lambda_I(y^k)| \approx (\alpha^{j_0 k})^b \approx |\Omega^{j_0 k}| = \alpha_1^{j_0 k}|E_1^{j_0}|,$$

for all k, so

$$\alpha_1^{j_0 k_0}(\alpha^{j_0 k_0})^{-\deg I}\frac{|\lambda_I(y^k)|}{|\lambda_I(x^0)|} \approx \alpha_1^{j_0 k}(\alpha^{j_0 k})^{-\deg I}.$$

Putting these inequalities together

$$| \bigwedge_{i=1}^{n-1} D\widetilde{\pi}_1(y^k)(\alpha^{j_0 k})^{\deg w_i} X_{w_i}(y^k)| \approx 1,$$

and by (6.15), this is possible only if (6.16) holds. $\square$

Finally, we are ready to complete the proof of Lemma 6.5.

Proof of Lemma 6.5. Let

$$\gamma(t) := D\widetilde{\pi}_1(e^{tX_1}(y^0))X_2(e^{tX_1}(y_0)) = D\widetilde{\pi}_1(y^0)De^{-tX_1}(e^{tX_1}(y^0))X_2(e^{tX_1}(y^0)).$$

Then γ is a polynomial, and

$$\gamma'(t) = D\widetilde{\pi}_1(y^0)De^{-tX_1}(e^{tX_1}(y^0))X_{12}(e^{tX_1}(y^0)) = D\widetilde{\pi}_1(e^{tX_1}(y^0))X_{12}(e^{tX_1}(y_0)).$$

Thus by Lemma 6.11,

$$|\gamma(t^k)| \approx_A (\alpha_2^{j_0 k})^{-1}, \qquad |\gamma(t^k) \wedge \gamma'(t^k)| \approx_A |\gamma(t^k)||\gamma'(t^k)|. \qquad (6.17)$$

By the definition of $\mathcal{L}$ and a bit of arithmetic,

$$\alpha_2^{j_0 k} \sim \varepsilon \eta_1^{\frac{1}{p_1}} 2^{-j_0} \eta_2^{-\frac{1}{p_2}} 2^{k\frac{p_2}{p_2}},$$

and thus for B sufficiently large, (6.17) contradicts Lemma 11.5. $\square$

7. ADDING UP THE TORSION SCALES

In this section, we add up the different torsion scales, $\rho \sim 2^{-m}$, thereby completing the proof of Theorem 1.1.

As in the previous section, we consider functions

$$f_i = \sum_k 2^k \chi_{E_i^k}, \qquad \|f_i\|_{p_i} \sim 1, \qquad i = 1, 2,$$

with the E_i^k pairwise disjoint (as k varies) for each i. For $m \in \mathbb{Z}$, we define $U_m := \{\rho \sim 2^{-m}\}$. By rescaling Proposition 6.1, we know that

$$\mathcal{B}_m(f_1, f_2) := \int_{U_m} f_1 \circ \pi_1(x) f_2 \circ \pi_2(x) \rho(x) \, dx \lesssim 1.$$

For $0 < \delta \lesssim 1$, define

$$\mathcal{M}(\delta) := \{m : \mathcal{B}_m(f_1, f_2) \sim \delta\}.$$

Define $\theta := (p_1^{-1} + p_2^{-1})^{-1}$. Then $0 < \theta < 1$. We will prove that for each $0 < \delta \lesssim 1$,

$$\sum_{m \in \mathcal{M}(\delta)} \mathcal{B}_m(f_1, f_2)^\theta \lesssim (\log \delta^{-1})^C. \qquad (7.1)$$

Thus

$$\sum_{m \in \mathcal{M}(\delta)} \mathcal{B}_m(f_1, f_2) \lesssim \delta^{1-\theta}(\log \delta^{-1})^C,$$

which implies Theorem 1.1.

The remainder of this section will be devoted to the proof of (7.1) for some fixed $\delta > 0$. We will use the notation $A \lesssim B$ to mean that $A \leq C\delta^{-C}B$ for some C depending on N.

For $m \in \mathcal{M}(\delta)$ and $\varepsilon, \eta_1, \eta_2 \lesssim 1$, define

$$\mathcal{L}_m(\varepsilon, \eta_1, \eta_2) := \{(j, k) : \mathcal{B}_m(\chi_{E_1^j}, \chi_{E_2^k}) \sim \varepsilon |E_1^j|^{\frac{1}{p_1}} |E_2^k|^{\frac{1}{p_2}},$$

$$2^{jp_1}|E_1^j| \sim \eta_1,\ 2^{kp_2}|E_2^k| \sim \eta_2\}.$$

By (6.4), the sum over all (j,k) lying in any $\mathcal{L}_m(\varepsilon, \eta_1, \eta_2)$ with ε, η_1, or η_2 much smaller than δ^C contributes a negligible amount to $\mathcal{B}_m(f_1, f_2)$:

$$\sum_{\min\{\varepsilon, \eta_1, \eta_2\} < c\delta^C} \sum_{(j,k)\in\mathcal{L}_m(\varepsilon, \eta_1, \eta_2)} 2^{j+k}\mathcal{B}_m(\chi_{E_1^j}, \chi_{E_2^k}) < c\delta < \tfrac{1}{2}\mathcal{B}_m(f_1, f_2).$$

Thus the majority of each $\mathcal{B}_m(f_1, f_2)$ is contributed by the $C(\log \delta^{-1})^3$ parameters $\varepsilon, \eta_1, \eta_2 \approx 1$. By the triangle inequality and pigeonholing, there exists (at least) one such triple for which

$$\sum_{m\in\mathcal{M}(\delta)} \mathcal{B}_m(f_1, f_2)^\theta \lesssim (\log \delta^{-1})^3 \sum_{m\in\mathcal{M}(\delta)} \sum_{(j,k)\in\mathcal{L}_m(\varepsilon, \eta_1, \eta_2)} \mathcal{B}_m(2^j\chi_{E_1^j}, 2^k\chi_{E_2^k})^\theta.$$

Henceforth, we will abbreviate $\mathcal{L}_m := \mathcal{L}_m(\varepsilon, \eta_1, \eta_2)$, for this choice of $\varepsilon, \eta_1, \eta_2$.

For $m \in \mathcal{M}(\delta)$ and $(j,k) \in \mathcal{L}_m$, we set

$$\Omega^{jkm} := U_m \cap \pi_1^{-1}(E_1^j) \cap \pi_2^{-1}(E_2^k), \qquad \alpha^{jkm} := \Big(\tfrac{|\Omega_m^{jk}|}{|E_1^j|}, \tfrac{|\Omega_m^{jk}|}{|E_2^k|}\Big).$$

There exist finite sets $\mathcal{A}^{jkm} \subseteq \Omega^{jkm}$, satisfying the conclusions of Proposition 5.1, appropriately rescaled. We proceed under the assumption that the minimal n-tuple I for all of these sets are the same; the general case follows by taking a sum over all possible minimal n-tuples. We set

$$\widetilde{\Omega}^{jkm} := \bigcup_{x\in\mathcal{A}^{jkm}} \bigcap_{\sigma\in S_n} B^{I_\sigma}(x; c\delta^C\alpha^{jkm}).$$

We recall that on these balls,

$$(\alpha^{jkm})^{\deg I}|\lambda_I| \approx (\alpha^{jkm})^b 2^{-m(|b|-1)} \approx |\Omega^{jkm}| \sim |\widetilde{\Omega}^{jkm}|.$$

(The factor $|b| - 1$ in the exponent is due to the form of the weight ρ.)

As in the preceding section, we let $q_i := \theta^{-1}p_i$. By the definition of Ω^{jkm}, $|\widetilde{\Omega}^{jkm}| \gtrsim |\Omega^{jkm}|$, the restricted weak type inequality (2.1), and Hölder's inequality,

$$\sum_{m\in\mathcal{M}(\delta)} \sum_{(j,k)\in\mathcal{L}_m} \mathcal{B}_m(2^j\chi_{E_1^j}, 2^k\chi_{E_2^k})^\theta$$

$$\lesssim \sum_{m\in\mathcal{M}(\delta)} \sum_{(j,k)\in\mathcal{L}_m} (2^{j+k}|\pi_1(\widetilde{\Omega}^{jkm})|^{\frac{1}{p_1}}|\pi_2(\widetilde{\Omega}^{jkm})|^{\frac{1}{p_2}})^\theta$$

$$\lesssim \Big(\sum_{m\in\mathcal{M}(\delta)} \sum_{(j,k)\in\mathcal{L}_m} 2^{jp_1}|\pi_1(\widetilde{\Omega}^{jkm})|\Big)^{\frac{1}{q_1}} \Big(\sum_{m\in\mathcal{M}(\delta)} \sum_{(j,k)\in\mathcal{L}_m} 2^{kp_2}|\pi_2(\widetilde{\Omega}^{jkm})|\Big)^{\frac{1}{q_2}}.$$

Thus the inequalities

$$\sum_{m\in\mathcal{M}(\delta)} \sum_{k:(j,k)\in\mathcal{L}_m} |\pi_1(\widetilde{\Omega}^{j,k,m})| \lesssim (\log \delta^{-1})^C |E_1^j|, \qquad j\in\mathbb{Z} \qquad (7.2)$$

$$\sum_{m\in\mathcal{M}(\delta)} \sum_{j:(j,k)\in\mathcal{L}_m} |\pi_2(\widetilde{\Omega}^{j,k,m})| \lesssim (\log \delta^{-1})^C |E_2^k|, \qquad k\in\mathbb{Z}, \qquad (7.3)$$

together would imply (7.1). The rest of the section will be devoted to the proof of (7.2), the proof of (7.3) being similar.

The proof is similar to the proof of Lemma 6.2; so we will just review that argument, giving the necessary changes. Let $\mathcal{K} \subseteq \mathbb{Z}^2$ be a finite set such that

$(j, k) \in \mathcal{L}_m$ for all $(k, m) \in \mathcal{K}$ and such that the following sets are all $(B \log \delta^{-1})$-separated for some $B = B(N)$ sufficiently large for later purposes:

$$\{k : (k, m) \in \mathcal{K}, \text{ for some m}\}, \qquad \{m : (k, m) \in \mathcal{K}, \text{ for some k}\},$$
$$\{m + \tfrac{p_2}{p_2'} k : (k, m) \in \mathcal{K}\}.$$

(In the case of the last set, we recall that $\frac{p_2}{p_2'}$ is rational.) It suffices to prove that

$$\sum_{(k,m) \in \mathcal{K}} |\pi_1(\widetilde{\Omega}^{jkm})| \lesssim |E_1^j|. \tag{7.4}$$

By the proof of Lemma 6.2, failure of (7.4) implies that there exists a subset $\mathcal{K}' \subseteq \mathcal{K}$ of cardinality $\#\mathcal{K}' \geq M$, with $M = M(N)$ sufficiently large for later purposes, and points $x^{jkm} \in \mathcal{A}^{jkm}$ such that

$$| \bigcap_{(k,m) \in \mathcal{K}'} \pi_1(\bigcap_{\sigma \in S_n} B^{I_\sigma}(x^{jkm}, c\delta^C \alpha^{jkm}))| \gtrsim |E_1^j|, \tag{7.5}$$

with $I = (w_1, \ldots, w_n)$ minimal and $w_n = (1)$. By rescaling Lemma 6.5 to torsion scale $\rho \sim 2^{-m}$, for each m, $\#(\mathbb{Z} \times \{m\}) \cap \mathcal{K} \lesssim 1$. Thus we may assume that

$$\mathcal{K}' = \{(k_1, m_1), \ldots, (k_M, m_M)\},$$

with the m_i all distinct. Set $\alpha^i := \alpha^{jk_i m_i}$.

As in the proof of Lemma 6.5, we can construct a submersion $\widetilde{\pi}_1$ and find points $y^i = e^{t^i X_1}(y)$ such that

$$2^{-m_i(|b|-1)} \sim \rho(y^i)^{|b|-1} \approx (\alpha^i)^{-b} \max_{I'}(\alpha^i)^{\deg I'} |\lambda_{I'}(y^i)|, \tag{7.6}$$

$$| \bigwedge_{s=1}^{L} D\widetilde{\pi}_1(y) De^{t^i X_1}(y^i)(\alpha^i)^{\deg w_{l_s}} X_{w_{i_l}}(y^i)| \approx 1, \tag{7.7}$$

for all $1 \leq i \leq M$ and $1 \leq l_1 < \cdots < l_L \leq n - 1$.

By construction, the m_i are all $(B \log \delta^{-1})$-separated. Thus by Lemma 3.2 and (7.6), for B sufficiently large,

$$|t^i - t^{i'}| \gtrsim \alpha_1^i + \alpha_1^{i'}, \qquad \text{for each } i \neq i'; \tag{7.8}$$

otherwise, two distinct balls would share a point in common, whence $2^{m_i} \approx 2^{m_{i'}}$, a contradiction. With $\gamma(t) := D\widetilde{\pi}_1(y) De^{tX_1}(e^{tX_1}(y)) X_2(e^{tX_1}(y))$, (7.7) gives

$$|\gamma(t^i)| \approx (\alpha_2^i)^{-1}, \quad |\gamma'(t^i)| \approx (\alpha_1^i \alpha_2^i)^{-1}, \quad |\gamma'(t^i) \wedge \gamma'(t^i)| \approx |\gamma(t^i)||\gamma'(t^i)|. \tag{7.9}$$

Since

$$\alpha_2^i \sim \varepsilon \eta_1^{\frac{1}{p_1}} \eta_2^{-\frac{1}{p_2'}} 2^{-j} 2^{m_i + k_i \frac{p_2}{p_2'}},$$

and the set of values $m_i + k_i \frac{p_2}{p_2'}$ takes on is $(B \log \delta^{-1})$-separated, by Lemma 11.5, we may assume that $m_i + k_i \frac{p_2}{p_2'}$ is constant as i varies. Thus we may fix α_2 so that $\alpha_2^i \sim \alpha_2$ for all i.

We note that

$$\alpha_1^i \sim \varepsilon \eta_1^{-\frac{1}{p_1'}} \eta_2^{\frac{1}{p_2}} 2^{-j \frac{p_1}{p_1'}} 2^{m_i - k_i}.$$

Since

$$m_i - k_i = -\tfrac{p_2'}{p_2}(m_i + k_i \tfrac{p_2}{p_2'}) + p_2' m_i = -\tfrac{p_2'}{p_2}(m_1 + k_1 \tfrac{p_2}{p_2'}) + p_2' m_i,$$

our prior deductions imply that the $m_i - k_i$ are all distinct, $(B \log \delta^{-1})$-separated. Reindexing, we may assume that $m_1 - k_1 < \cdots < m_M - k_M$. Thus $\alpha_1^1 < \cdots < \alpha_1^M$.

By Lemma 11.4 (after a harmless time translation), we may assume that all of the t^i lie within a single interval $I \subseteq (0, \infty)$ on which

$$|\tfrac{1}{k!}\gamma^{(k)}(0)t^k| < c_N |\tfrac{1}{k_0!}\gamma^{(k_0)}(0)t^{k_0}|, \qquad k \neq k_0, \tag{7.10}$$

with c_N sufficiently small. As we have seen, $|\gamma(t^i)| \approx \alpha_2^{-1}$, for all i. On the other hand, for c_N sufficiently small, and any subinterval $I' \subseteq I$,

$$|\int_{I'} \gamma'(t)\,dt| \sim |I| \max_{t \in I} |\gamma'(t)|.$$

(We can put the norm outside of the integral by (7.10).) Specializing to the case when I' has endpoints t_1, t_2, and using (7.8),

$$\alpha_1^2 (\alpha_1^1 \alpha_2)^{-1} \lesssim |t_1 - t_2||\gamma'(t_2)| \lesssim |\gamma(t_2) - \gamma(t_1)| \approx (\alpha_2)^{-1},$$

i.e. $\alpha_1^2 \lesssim \alpha_1^1$, which is impossible for B sufficiently large. Thus we have a contradiction, and tracing back, (7.2) must hold. This completes the proof of Theorem 1.1.

8. Nilpotent Lie algebras and polynomial flows

In the next section, we will generalize Theorem 1.1 by relaxing the hypothesis that the flows of the vector fields X_j must be polynomial. In this section, we lay the groundwork for that generalization by reviewing some results from Lie group theory. In short, we will see that if M is a smooth manifold and $\mathfrak{g}_M \subseteq \mathcal{X}(M)$ is a nilpotent Lie algebra, then there exist local coordinates for M in which the flows of the elements of $\mathfrak{g}$ are polynomial. These results have the advantage over the analogous results in [13] that the lifting of the vector fields is by a local diffeomorphism, rather than a submersion; this will facilitate the global results in the next section.

Throughout this section, M will denote a connected n-dimensional manifold, and $\mathfrak{g}_M \subseteq \mathcal{X}(M)$ will denote a Lie subalgebra of the space $\mathcal{X}(M)$ of smooth vector fields on M. We assume throughout that $\mathfrak{g}_M$ is nilpotent, and we let $N := \dim \mathfrak{g}_M$. We further assume that the elements of $\mathfrak{g}_M$ span the tangent space to M at every point. We will say that a quantity is bounded if it is bounded by a finite, nonzero constant depending only on N, and our implicit constants will continue to depend only on N.

For the moment, we will largely forget about the manifold M.

Let G denote the unique connected, simply connected Lie group with Lie algebra $\mathfrak{g}_M$. For clarity, we denote the Lie algebra of right invariant vector fields on G by $\mathfrak{g}$, and we fix an isomorphism $X \mapsto \widehat{X}$ of $\mathfrak{g}_M$ onto $\mathfrak{g}$. Under the natural identification of G as a subgroup of $\mathrm{Aut}(G)$, $G = \exp(\mathfrak{g})$, and the group law is given by $e^{\widehat{X}} \cdot e^{\widehat{Y}} = e^{\widehat{X}} \circ e^{\widehat{Y}} = e^{\widehat{X} * \widehat{Y}}$, where $X * Y$ a Lie polynomial in X and Y, which is given explicitly by the Baker–Campbell–Hausdorff formula.

Let S be a Lie subgroup of G. The Lie algebra $\mathfrak{z}$ of S is a Lie subalgebra of $\mathfrak{g}$, and $Z := \exp(\mathfrak{z})$ is the connected component of S containing the identity. In addition, Z is a normal subgroup of S. Let $n := N - \dim \mathfrak{z}$. (Later on, we will set $\mathfrak{z} = \mathfrak{z}_{x_0} := \{\widehat{X} \in \mathfrak{g} : X(x_0) = 0\}$ and $S = S_{x_0} := \{e^{\widehat{X}} : e^X(x_0) = x_0\}$.)

Let $\Pi : G \to G/Z$ denote the quotient map. For $g \in G$ and $s \in S$, left multiplication by g and right multiplication by s have well-defined pushforwards; in other words, there exist automorphisms $\Pi_* l_g$, $\Pi_* r_s$ on G/Z such that

$$(\Pi_* l_g)(hZ) = (gh)Z, \qquad (\Pi_* r_s)(hZ) = (hs)Z,$$

for every $h \in G$.

Our next task is to find good coordinates on G.

Lemma 8.1 ([7, Theorem 1.1.13]). *There exists an ordered basis $\{\widehat{X}_1, \ldots, \widehat{X}_N\}$ of $\mathfrak{g}$, such that for each k, the linear span $\mathfrak{g}_k$ of $\{\widehat{X}_{k+1}, \ldots, \widehat{X}_N\}$ is a Lie subalgebra of $\mathfrak{g}$ and such that $\mathfrak{g}_n = \mathfrak{z}$.*

We will not replicate the proof.

Such a basis is called a weak Malcev basis of $\mathfrak{g}$ through $\mathfrak{z}$. As we will see, the utility of weak Malcev bases is that they give coordinates for G and G/Z in which the flows of our vector fields are polynomial. We will say that a function q is a polynomial diffeomorphism on $\mathbb{R}^N$ if $q : \mathbb{R}^N \to \mathbb{R}^N$ is a polynomial having a well defined inverse $q^{-1} : \mathbb{R}^N \to \mathbb{R}^N$ that is also a polynomial. Polynomial diffeomorphisms must have constant Jacobian determinant; we will say that they are volume-preserving if this constant equals 1.

Fix a weak Malcev basis $\{\widehat{X}_1, \ldots, \widehat{X}_N\}$ for $\mathfrak{g}$ through $\mathfrak{z}$. For convenience, we will use the notation $x \cdot \vec{X} := \sum_{j=1}^N x_j \widehat{X}_j$, for $x \in \mathbb{R}^N$. Define

$$\psi(x) := e^{x_1 \widehat{X}_1} \cdots e^{x_N \widehat{X}_N}.$$

Lemma 8.2. *There exists a polynomial diffeomorphism p on $\mathbb{R}^N$ such that $\psi(x) = \exp(p(x) \cdot \vec{X})$. In particular, ψ is a diffeomorphism of $\mathbb{R}^N$ onto G. In these coordinates, the right and left exponential maps are polynomial. More precisely, for $x^1, x^2 \in \mathbb{R}^N$,*

$$e^{x^2 \cdot \vec{X}} \psi(x^1) = \psi(q(x^1, x^2)), \qquad \psi(x^1) e^{x^2 \cdot \vec{X}} = \psi(r(x^1, x^2)),$$

where $q, r : \mathbb{R}^{2N} \to \mathbb{R}^N$ are polynomials, $q(\cdot, x^2)$ and $r(\cdot, x^2)$ are volume-preserving polynomial diffeomorphisms for each x^2, and for each $1 \le i \le N$, $q_i(x^1, x^2)$ only depends on $x_1^1, \ldots, x_i^1$, and x^2.

Proof. The assertion on p is just Proposition 1.2.8 of [7]. That q and r are polynomial just follows by taking compositions:

$$\exp(q(x^1, x^2)) = \exp(x^2 \cdot \vec{X}) \psi(x^1) = \exp((x^2 \cdot \vec{X}) * p(x^1)) = \psi(p^{-1}((x^2 \cdot \vec{X}) * p(x^1)));$$

similarly for r.

The inverse of $r(\cdot, x^2)$ is $r(\cdot, -x^2)$, also a polynomial. Since $r(r(x^1, x^2), -x^2) \equiv x^1$, $\det(D_{x^1} r)(r(x^1, x^2), -x^2) \det D_{x^1} r(x^1, x^2) \equiv 1$, and since both determinants are polynomial in x^1 and x^2, both must be constant. Finally, since $r(x^1, 0)$ is the identity, this constant must be 1.

We turn to the dependence of q_i on x^2 and the first i entries of x^1. Set $G_k := \exp(\mathfrak{g}_k)$ (in the notation of Lemma 8.1). Our coordinates ψ on G give rise to diffeomorphisms

$$\phi_k : \mathbb{R}^k \to G/G_k, \qquad \phi_k(y) := \psi(y, 0) G_k.$$

In these coordinates, the projections $\Pi_k : G \to G/G_k$ may be expressed as coordinate projections: $\phi_k^{-1} \circ \Pi_k \circ \psi(y, z) = y$. Since left multiplication pushes forward via Π_k,

$$(q_1, \ldots, q_i)(y, z, x^2) = \phi_k^{-1} \circ \Pi_k(l_{e^{x^2 \cdot \vec{X}}} \psi(y, z)) = \phi_k^{-1}((\Pi_k)_* l_{e^{x^2 \cdot \vec{X}}} \Pi_k \psi(y, z))$$
$$= \phi_k^{-1}((\Pi_k)_* l_{e^{x^2 \cdot \vec{X}}} \phi_k(y)),$$

which is independent of z. $\qquad\square$

Recalling that $Z = G_n$, we set $\phi := \phi_n$. The pushforwards $\Pi_* \widehat{X}$, $\widehat{X} \in \mathfrak{g}$, are well-defined and have polynomial flows; indeed,

$$\exp(\Pi_*(x \cdot \vec{X}))(\phi(y)) = \phi(q_1((y, 0), x), \ldots, q_n((y, 0), x)).$$

Furthermore, Π_* is a Lie group homomorphism of $\mathfrak{g}$ onto a Lie subgroup of $\mathcal{X}(G/Z)$, and, since Π_* is a submersion and $\mathfrak{g}$ spans the tangent space to $\mathbb{R}^N$ at every point, $\Pi_* \mathfrak{g}$ spans the tangent space to $\mathbb{R}^n$ at every point.

Next we examine the pushforwards $\Pi_* r_s$ of right multiplication by $s \in S$. First, a preliminary remark. Since Z is a normal subgroup of S, S acts on Z by conjugation. Replacing G with Z, Lemma 8.2 implies that the pushforward $\psi_* dz$ of $(N - n)$-dimensional Hausdorff measure on Z is a bi-invariant Haar measure on Z. We may uniquely extend this to a bi-invariant Haar measure on S. Both Z and this Haar measure on S are invariant under the conjugation action, so $\psi_* dz$ is invariant under the conjugation action of S.

Lemma 8.3. *In the coordinates given by ϕ, the pushfoward $\Pi_* r_s$ is a volume-preserving polynomial diffeomorphism.*

Proof. By Lemma 8.2, for each $s \in S$, there exists a polynomial $r^s : \mathbb{R}^N \to \mathbb{R}^N$ such that $r_s(\psi(x)) = \psi(r^s(x))$. From the definition of the pushforward,

$$\Pi_* r_s(\phi(y)) = \Pi(r_s(\psi(y, 0))) = \Pi(\psi(r^s(y, 0))) = \phi((r_1^s, \ldots, r_n^s)(y, 0)),$$

and taking the composition with ϕ^{-1} yields a polynomial. Since $(r^s)^{-1} = r^{-s}$, this is also a polynomial diffeomorphism. It remains to verify that this diffeomorphism is volume-preserving.

For simplicity, we will use vertical bars to denote the pushforward by ϕ of Lebesgue measure on $\mathbb{R}^n$ to G/Z and also the pushforwards by ψ of Lebesgue measure on $\mathbb{R}^N$ to G and Hausdorff measure on $\mathbb{R}^{N-n} \times \{0\}$ to Z. Fix an open, unit volume set $B \subseteq Z$. By the remarks preceding the statement of Lemma 8.3, $|s^{-1} Bs| = |B| = 1$. Let $U \subseteq G/Z$ be measurable, and let $\sigma : G/Z \to G$ denote the section $\sigma(u) := \psi(\phi^{-1}(u), 0)$. By the coarea formula,

$$|\Pi_* r_s U| = |\sigma(\Pi_* r_s U)(s^{-1} Bs)|.$$

Of course, $\sigma(\Pi_* r_s U)(s^{-1} Bs) = (\sigma(U)B)s$, so using the fact that right multiplication by s is volume-preserving, and using the coarea formula a second time,

$$|\Pi_* r_s U| = |\sigma(U)B| = |U|.$$

$\square$

Now we are ready to return to our n-dimensional manifold M from the opening of this section. Fix $x_0 \in M$, and set $\mathfrak{z} = \mathfrak{z}_{x_0} := \{\widehat{X} \in \mathfrak{g} : X(x_0) = 0\}$ and $Z = Z_{x_0} := \exp(\mathfrak{z})$.

We consider the smooth manifold $H = H_{x_0} := \mathbb{R}^n \times M$, and view $\mathfrak{g} \simeq \mathfrak{g}_H$ as a tangent distribution on H, with elements $(\phi^* \Pi_* \widehat{X}) \oplus X \in \mathfrak{g}_H$. By the Frobenius theorem, there exists a smooth submanifold $(0, x_0) \in L = L_{x_0} \subseteq H$ whose tangent space equals the span of the elements of $\mathfrak{g}_H$ at each point. The dimension of this leaf equals n; indeed, the map $\phi^* \Pi_* \widehat{X}(0) \mapsto X(x_0)$ is an isomorphism, so its graph, $T_{(0,x_0)} H$, has dimension n.

We let $p_1 : L \to \mathbb{R}^n$ and $p_2 : L \to M$ denote the restrictions to L of the coordinate projections of H onto $\mathbb{R}^n$ and M, respectively. These restrictions are smooth, because $\phi^* \Pi_* \mathfrak{g}$ and $\mathfrak{g}_M$ span the tangent spaces to $\mathbb{R}^n$ and M, respectively,

at every point. For this same reason, they are in fact submersions, and hence local diffeomorphisms. Composition of p_2 with a local inverse for p_1 immediately yields the following.

Lemma 8.4. *Let $x_0 \in M$ and fix a weak Malcev basis $\{\widehat{X}_1, \ldots, \widehat{X}_N\}$ of $\mathfrak{g}$ through $\mathfrak{z}_{x_0}$. Then there exist neighborhoods V_{x_0} of 0 in $\mathbb{R}^n$ and U_{x_0} of x_0 such that the map*

$$\Phi_{x_0}(y) := e^{y_1 X_1} \cdots e^{y_n X_n}(x_0)$$

is a diffeomorphism of V_{x_0} onto U_{x_0}, and, moreover, the pullbacks $\widetilde{X} := (\Phi_{x_0})^* X$, $X \in \mathfrak{g}_M$ may be extended to globally defined vector fields on $\mathbb{R}^n$ for which each exponentiation $(t, x) \mapsto e^{t\widehat{X}}(x_0)$ is a polynomial of bounded degree.

We would like to remove the restriction to small neighborhoods of points in M from the preceding.

Lemma 8.5. *The projection $p_2 : L_{x_0} \to M$ is a covering map.*

Proof of Lemma 8.5. That p_2 is surjective follows from Hörmander's condition and connectedness of M. Indeed, any point of the form $e^{X_1} \cdots e^{X_K}(x_0)$ (here we assume that each of the exponentials is defined) is in the range of p_2, and the set of such points is both open and closed in M. (This is Chow's theorem.)

Let $x \in M$. Fix a weak Malcev basis $\{\widehat{W}_1, \ldots, \widehat{W}_N\}$ of $\mathfrak{g}$ through $\mathfrak{z}_x$. Then there exist neighborhoods $0 \in V_x \subseteq \mathbb{R}^n$ and $x \in U_x \subseteq M$ such that

$$\Phi_x(w) := e^{w_1 W_1} \cdots e^{w_n W_n}(x)$$

is a diffeomorphism of V_x onto U_x, so

$$p_2^{-1}(U_x) = \bigcup_{y:(y,x) \in L_{x_0}} \{(e^{w_1 \widetilde{W}_1} \cdots e^{w_n \widetilde{W}_n}(y), \Phi_x(w)) : w \in V\},$$

where $\widetilde{W}_n := \phi^* \Pi_* \widehat{W}_n$, and the restriction of p_2 to each set in this union is a diffeomorphism. $\square$

Lemma 8.6. *Assume that the exponential $e^X(x_0)$ is defined for every $X \in \mathfrak{g}_M$. Then the projection $p_1 : L_{x_0} \to G/Z_{x_0}$ is one-to-one.*

Proof. The projection p_1 fails to be one-to-one if and only if there exist $\widehat{X}_1, \ldots, \widehat{X}_K \in \mathfrak{g}$ such

$$e^{X_1} \cdots e^{X_K}(x_0)$$

is defined and not equal to x_0, but $\widehat{X}_1 * \cdots * \widehat{X}_K = 0$. Thus it suffices to show that if $e^{X_1} \cdots e^{X_K}(x_0)$ is defined, it equals $e^{X_1 * \cdots * X_K}(x_0)$. By induction, it suffices to prove this when $K = 2$.

Assume that $e^X e^Y(x_0)$ is defined, and let

$$E := \{t \in [0, 1] : e^{sX} e^Y(x_0) = e^{(sX)*Y}(x_0), \ s \in [0, t]\}.$$

Let $Y_t := (tX) * Y$, $t \in [0, 1]$. It suffices to prove that there exists $\delta > 0$ such that for each $t \in [0, 1]$ and $0 < s < \delta$, $e^{sX} e^{Y_t}(x_0) = e^{(sX)*Y_t}(x_0)$. From our initial remark, p_1 is one-to-one on each of the sets

$$\Gamma_t := \{e^{s\widehat{Y}_t}(0, x_0) : s \in [0, 1]\}, \qquad \widehat{Y}_t := \widetilde{Y}_t \oplus Y_t \in \mathfrak{g}_H.$$

Since p_1 is a local diffeomorphism, the Γ_t are compact, and t varies in a compact interval, there exists $\delta > 0$ such that p_1 is a diffeomorphism on the neighborhoods

$$N_\delta(\Gamma_t) := \{e^{\widehat{Z}}(z) : \widehat{Z} \in \mathfrak{g}_H, |\widehat{Z}| < \delta, z \in \Gamma_t\}.$$

Since $p_2 \circ p_1^{-1}|_{N_\delta(\Gamma_t)}$ is a diffeomorphism, for s sufficiently small (independent of t),

$$e^{sX}e^{Y_t}(x_0) = p_2 \circ p_1^{-1}(e^{s\widetilde{X}}e^{\widetilde{Y}_t}(0)) = p_2 \circ p_1^{-1}(e^{(s\widetilde{X})*\widetilde{Y}_t}(0)) = e^{(sX)*Y_t}(x_0).$$

$\square$

Taking the composition $p_2 \circ p_1^{-1}$, we obtain the following.

Proposition 8.7. *Let $x_0 \in M$, and assume that $e^X(x_0)$ is defined for each $X \in \mathfrak{g}_M$. Fix a weak Malcev basis $\{\widehat{X}_1, \ldots, \widehat{X}_N\}$ of $\mathfrak{g}$ through $\mathfrak{z}_{x_0}$. Then the map*

$$\Phi_{x_0}(y) := e^{y_1 X_1} \cdots e^{y_n X_n}(x_0)$$

is a local diffeomorphism of $\mathbb{R}^n$ onto M, which is also a covering map. For each $X \in \mathfrak{g}_M$, the flow $(t, x) \mapsto e^{t\widetilde{X}}(x)$ of the pullback $\widetilde{X} := \Phi_{x_0}^* X$ is polynomial. Finally, the covering is regular, and elements of the deck transformation group are volume-preserving.

Much of the proposition has already been proved; our main task is the following.

Lemma 8.8. *Let $S := \{e^{\widehat{X}} \in G : e^X(x_0) = x_0\}$. Then the deck transformation group $\mathrm{Aut}(\Phi_{x_0})$ of Φ_{x_0} coincides with the group $\mathcal{S} \subseteq \mathrm{Diff}(\mathbb{R}^n)$ whose elements are the pushforwards $\widehat{r}_s := \phi^* \Pi_* r_s$ of right multiplication by elements of S.*

Proof of Lemma 8.8. Let $s = e^{\widehat{X}} \in S$. Then

$$\Phi_{x_0} \circ \phi^{-1} \circ (\Pi_* r_s) \circ \phi(y) = e^{y_1 Y_1} \cdots e^{y_n Y_n} e^X(x_0) = \Phi_{x_0}(y),$$

so $\mathcal{S} \subseteq \mathrm{Aut}(\Phi_{x_0})$. If $y_0 \in \Phi_{x_0}^{-1}(x_0)$, then we may write $y_0 = e^{\widetilde{X}}(0)$, with $e^{\widehat{X}} \in S$, so $\mathcal{S}$ acts transitively on the fiber $\Phi_{x_0}^{-1}(x_0)$.

Let $f \in \mathrm{Aut}(\Phi_{x_0})$, and set $y_0 := f(0)$. By the preceding, there exists an element $r \in \mathcal{S}$ such that $r(0) = y_0$. We claim that $f = r$. The set of points where the maps coincide is closed by continuity. If $f(y) = r(y)$, then the maps must coincide on a neighborhood of y, because Φ_{x_0} is a covering map. Thus the set of points where the maps coincide is also open. Since $f(0) = r(0)$, $f \equiv r$. $\square$

Proof of Proposition 8.7. It remains to prove that the covering Φ_{x_0} is regular, and that the elements of its deck transformation group are volume-preserving. By Lemma 8.3, the deck transformations are all volume-preserving, and as seen in the proof of Lemma 8.8, $\mathrm{Aut}(\Phi_{x_0})$ acts transitively on $\Phi_{x_0}^{-1}(x_0)$, which is to say that Φ_{x_0} is regular. $\square$

9. Generalizations of Theorem 1.1

In [13], which sparked our interest in this problem, Gressman established unweighted, local, endpoint restricted weak type inequalities, subject to the hypotheses that the $\pi_j : \mathbb{R}^n \supseteq U \to \mathbb{R}^{n-1}$ are smooth submersions and that there exist smooth, nonvanishing vector fields Y_1, Y_2 on U that are tangent to the fibers of the π_j and generate a nilpotent Lie algebra. Thus the results of [13] are more general than Theorem 1.1 in two respects: The hypotheses are made on vector fields parallel to the fibers, and these vector fields are only assumed to generate a nilpotent

Lie algebra, not to have polynomial flows. In this section, we address both of these generalizations.

Changes of variables, changes of measure, and the affine arclengths. The above mentioned generalizations will be achieved by using the results of the previous section, so we begin by observing how the weights ρ_β transform under compositions of the π_j with diffeomorphisms. We note that the same computations also show how the ρ_β transform under smooth changes of the measures on M and the N_j. (Changes of measure change the vector fields associated to the maps π_1, π_2 by the coarea formula.)

Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a diffeomorphism, and let $G_j : \mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$ be a smooth map, $j = 1, 2$. Define $\widehat{\pi}_j := G_j \circ \pi_j \circ F$. These maps give rise to associated vector fields $\widehat{X}_j$, and a simple computation shows that

$$\widehat{X}_j = [(\det DG_j) \circ \pi_j \circ F](\det DF)F^* X_j, \tag{9.1}$$

where F^* denotes the pullback $F^* X_j := (DF)^{-1} X_j \circ F$. We continue to let $\Psi_{F(x_0)}(t)$ denote the map obtained by iteratively flowing along the X_i and let $\widehat{\Psi}_{x_0}(t)$ denote the map obtained by iteratively flowing along the $\widehat{X}_i$.

By naturality of the Lie bracket and the Chain Rule, we thus have for any multiindex β that

$$\partial^\beta \det D\widehat{\Psi}_{x_0}(0) = \sum_{\beta' \preceq \beta} G^\beta_{\beta'}(F(x_0))\partial^{\beta'} \det D\Psi_{F(x_0)}(0). \tag{9.2}$$

Here '$\preceq$' denotes the coordinate-wise partial order on multiindices,

$$G^\beta_\beta(F(x_0)) := (\det DF(x_0))^{b_1+b_2-1}(\det DG_1 \circ \pi_1 \circ F(x_0))^{b_1}(\det DG_2 \circ \pi_2 \circ F(x_0))^{b_2},$$

and for $\beta' \prec \beta$, $G^\beta_{\beta'}$ is a smooth function involving derivatives of the Jacobian determinants $\det DF$, $\det DG_i$.

This allows us to bound the weight associated to the maps $\widehat{\pi}_1, \widehat{\pi}_2$ and multiindex β:

$$|\widehat{\rho}_\beta| \leq |\det DF||(\det DG_1) \circ \pi_1 \circ F|^{\frac{1}{p_1}}|(\det DG_2) \circ \pi_2 \circ F|^{\frac{1}{p_2}}\rho_\beta \circ F$$

$$+ \sum_{\beta' \prec \beta} g^{\beta'}_\beta \rho_{\beta'}^{\frac{|b'|-1}{|b|-1}} \circ F, \tag{9.3}$$

where the $g^{\beta'}_\beta$ are continuous and equal zero if $\det DF$, $\det DG_1$, and $\det DG_2$ are constant, and $b' = b(\beta')$ and $\rho_{\beta'}$ are as in (1.6), (1.7), respectively, p is as in (1.8), and vertical bars around b's denote the ℓ^1 norm.

We turn to an estimate for

$$\int \prod_{j=1}^2 |f_j \circ \widehat{\pi}_j|\widehat{\rho}_\beta \, a(x) \, dx,$$

with $|a| \leq 1$ a cutoff function (possibly identically 1). We begin with the contribution from the main term of (9.3). Assuming (1.9), the change of variables formula gives

$$\int \Big(\prod_{j=1}^2 |f_j \circ \widehat{\pi}_j| \, |\det DG_j \circ \pi_j \circ F|^{\frac{1}{p_j}}\Big)|\det DF|\,\rho_\beta \circ F \, a \, dx \lesssim \prod_{j=1}^2 \|f_j\|_{p_j}. \tag{9.4}$$

Now we turn to the error terms. Fix $\beta' \prec \beta$ and assume that a has compact support. The analogue of (1.9), with β' in place of β, together with the change of variables formula, yields

$$\left|\int (\prod_{j=1}^{2} f_j \circ \widehat{\pi}_j)(g_\beta^{\beta'})^{1/\theta} \rho_{\beta'} \circ F\, a\, dx\right| \lesssim_{F,G_1,G_2} \prod_{j=1}^{2} \|f_j\|_{q_j}, \qquad (9.5)$$

where $q = p(b') = (\frac{|b'|-1}{b_1}, \frac{|b'|-1}{b_2})$ and $\theta = \frac{|b'|-1}{|b|-1}$. Provided that the π_j are submersions on the support of a, Hölder's inequality gives

$$\left|\int (\prod_{j=1}^{2} f_j \circ \widehat{\pi}_j)\, a\, dx\right| \lesssim_{F,G_1,G_2,\pi_1,\pi_2} \operatorname{diam}(\operatorname{supp} a) \prod_{j=1}^{2} \|f_j\|_{r_j}, \qquad (9.6)$$

where $(r_1, r_2) = (\frac{|b|-|b'|}{b_1-b_1'}, \frac{|b|-|b'|}{b_2-b_2'})$. Since $(p_1^{-1}, p_2^{-1}) = \theta(q_1^{-1}, q_2^{-1}) + (1-\theta)(r_1^{-1}, r_2^{-1})$, complex interpolation gives

$$\left|\int (\prod_{j=1}^{2} f_j \circ \widehat{\pi}_j) g_\beta^{\beta'} (\rho_{\beta'} \circ F)^\theta\, a\, dx\right| \lesssim_{F,G_1,G_2,\pi_1,\pi_2} \operatorname{diam}(a)^{1-\theta} \prod_{j=1}^{2} \|f_j\|_{p_j}, \qquad (9.7)$$

so the error terms are harmless for sufficiently local estimates **in the special case** that the π_j are submersions on the support of a.

Uniform local estimates. For simplicity, we will give our local estimates in coordinates. Let $U \subseteq \mathbb{R}^n$ be an open set, let $\pi_1, \pi_2 : U \to \mathbb{R}^n$ be smooth maps, and let X_1, X_2 denote the vector fields associated to the π_j by (1.3). Assume that:
(i) For $j = 1, 2$ and a.e. $y \in \pi_j(U)$, $\pi_j^{-1}(y)$ is contained in a single integral curve of X_j;
(ii) The Lie algebra generated by X_1, X_2 spans the tangent space to $\mathbb{R}^n$ at every point of U;
(iii) There exist smooth, nonvanishing functions h_1, h_2 such that the vector fields $Y_j := h_j X_j$, $j = 1, 2$, generate a nilpotent Lie algebra of step at most N.

We note that even if one knows that (i-iii) hold, it may be very difficult to find h_1, h_2. Our next proposition allows one to use the "wrong" vector fields (the X_j), at least locally, and for certain β.

Proposition 9.1. *Fix $x_0 \in U$. If β is minimal in the sense that $\beta' \prec \beta$ implies $\rho_{\beta'} \equiv 0$, or if $d\pi_1(x_0)$ and $d\pi_2(x_0)$ both have full rank, then there exists a neighborhood U_{x_0} of x_0, depending on x_0 and the π_j, such that for all $f_1, f_2 \in C^0(U)$,*

$$\left|\int_{U_{x_0}} \prod_{j=1}^{2} f_j \circ \pi_j(x)\, \rho_\beta(x)\, dx\right| \leq C_N \prod_{j=1}^{2} \|f_j\|_{p_j}; \qquad (9.8)$$

here ρ_β is the weight (1.7), defined using the X_j, not the Y_j.

Proposition 9.1 implies a uniform, strong type endpoint version of the restricted weak type result in [13]. We remark that uniform bounds are impossible if we define the weight ρ_β using the Y_i. This can be seen by replacing Y_i with λY_i and sending $\lambda \to \infty$. In Section 10, we will give a counter-example showing the impossibility of global bounds under these hypotheses in the case that β is non-minimal.

Proof of Proposition 9.1. By Lemma 8.4, we may find neighborhoods U_{x_0} of x_0 and V_{x_0} of 0, and a diffeomorphism $\Phi_{x_0} : V_{x_0} \to U_{x_0}$ such that the pullbacks $\widehat{Y}_j$ of the

Y_j with respect to Φ_{x_0} extend to global vector fields with polynomial flows. Let $\widehat{Z}_j$ denote the vector field associated to $\widehat{\pi}_j := \pi_j \circ \Phi_{x_0}$, via the natural analogue of (1.3). Then

$$\widehat{Z}_j = (\det D\Phi_{x_0})\Phi_{x_0}^* X_j = \tfrac{1}{\widehat{h}_j}\widehat{Y}_j, \qquad \widehat{h}_j = h_j \circ \Phi_{x_0}.$$

Lemma 9.2. *There exist functions g_j on $\pi_j(U_{x_0})$ such that $\widehat{h}_j = g_j \circ \widehat{\pi}_j$, a.e. on U_{x_0}.*

Proof of Lemma 9.2. Since $\widehat{Y}_j$ is polynomial, it is divergence free, and since $\widehat{Z}_j$ is defined by (1.3), it is also divergence free. Since

$$0 = \operatorname{div}\widehat{Z}_j = \tfrac{1}{\widehat{h}_j}\operatorname{div}\widehat{Y}_j + \widehat{Y}_j(\tfrac{1}{\widehat{h}_j}) = \widehat{Y}_j(\tfrac{1}{\widehat{h}_j}),$$

$\widehat{h}_j$ is constant on the integral curves of $\widehat{Y}_j$. By our hypothesis on the fibers of the π_j, the lemma follows. $\qquad\square$

If $\Omega \subseteq V_{x_0}$,

$$|\Omega| = \int_{\widehat{\pi}_j(\Omega)} \int_{\pi_j^{-1}(y)} \chi_\Omega(t)\,|\widehat{Z}_j(t)|^{-1} d\mathcal{H}^1(t)\,dy$$

$$= \int_{\widehat{\pi}_j(\Omega)} \int_{\pi_j^{-1}(y)} \chi_\Omega(t)\,|\widehat{Y}_j(t)|^{-1} d\mathcal{H}^1(t)\,g_j(y)\,dy.$$

Thus the change of variables formula and the proof of Theorem 1.1 (c.f. the argument leading to (9.11)) imply that

$$|\int_{U_{x_0}} \prod_{j=1}^2 f_j \circ \pi_j\,\widehat{\rho}_\beta \circ \Phi_{x_0}^{-1}\,|\det D\Phi_{x_0}|^{-1}\,dx| = |\int_{V_{x_0}} \prod_{j=1}^2 f_j \circ \widehat{\pi}_j\,\widehat{\rho}_\beta\,dx| \tag{9.9}$$

$$\lesssim \prod_{j=1}^2 \|f_j\|_{L^{p_j}(g_j\,dy)},$$

where $\widehat{\rho}_\beta$ is defined using $\widehat{Y}_1$ and $\widehat{Y}_2$. We have seen that $\frac{1}{h_j} = g_j \circ \pi_j$, so computations similar to those leading up to (9.3) give

$$|\rho_\beta| \leq |\det D\Phi_{x_0}|^{-1}|g_1 \circ \pi_1|^{\frac{1}{p_1}}|g_2 \circ \pi_2|^{\frac{1}{p_2}}\widehat{\rho}_\beta \circ \Phi_{x_0}^{-1} + \sum_{\beta \prec \beta'} g_\beta^{\beta'}\,\rho_{\beta'}^{\frac{|b'|-1}{|b|-1}} \circ \Phi_{x_0}^{-1},$$

where the $g_\beta^{\beta'}$ are continuous and involve derivatives of $\det D\Phi_{x_0}$, g_1, and g_2.

Finally, (9.8) follows from (9.4) and (9.7) in the case that β is minimal or $d\pi_1(x_0)$ and $d\pi_2(x_0)$ both have full rank. $\qquad\square$

A "global" version on manifolds. Let M be a smooth n-dimensional manifold, let P_1, P_2 be smooth $(n-1)$-dimensional manifolds, and assume that $\pi_j : M \to P_j$ are smooth maps with a.e. surjective differentials. Assume that we are given measures μ, ν_1, ν_2 on M, P_1, P_2 that have smooth, nonvanishing densities in local coordinates.

For instance, in the setting of (1.1) and (1.2), we are given Riemannian manifolds $(P_1, h_1), (P_2, h_2)$ and a map

$$P_2 \times \mathbb{R} \ni (x,t) \mapsto \gamma_x(t) \in P_1;$$

here the measures ν_1, ν_2 are the Riemannian volume elements, the manifold M is simply $M = P_2 \times \mathbb{R}$, and $d\mu = |\gamma'_{x_2}(t)|_{h_1} \, d\nu_2 \, dt$.

By (9.1), we may define (up to a sign) vector fields $X_1, X_2 \in \mathcal{X}(M)$ such that in any choice of local coordinates,

$$X_j = d\pi_j^1 \wedge \cdots \wedge d\pi_j^{n-1} \left(\tfrac{d\nu_j}{dy} \circ \pi_j\right)\left(\tfrac{dx}{d\mu}\right).$$

We observe that μ is invariant under the flow of the X_j, and hence is also invariant under the flow of elements of the Lie algebra $\mathfrak{g}_M$ generated by X_1 and X_2. We assume that:

(i) The Lie algebra $\mathfrak{g}_M$ generated by X_1, X_2 is nilpotent of step N, and the flows of its elements are complete;

(ii) For a.e. $y \in P_j$, $\pi_j^{-1}(y)$ is contained in a single integral curve of X_j.

Let M_0 denote the (open) submanifold of M on which $\mathfrak{g}_M$ spans the tangent space to M, and decompose M_0 into its connected components, $M_0 = \bigcup_k M_{0,k}$. By the Frobenius Theorem, $\mathfrak{g}_M \subseteq \mathcal{X}(M_{0,k})$ for each k. We now put local coordinates on M_0 by fixing points $x_k \in M_{0,k}$ and letting $\Phi_k := \Phi_{x_k} : \mathbb{R}^n \to M_{0,k}$ be the covering map guaranteed by Proposition 8.7.

Fix k. Then Φ_k is a local diffeomorphism, and the pullbacks of vector fields in $\mathfrak{g}_M$ by Φ_k have polynomial flows. By composing Φ_k with an isotropic dilation, we may assume that for $U_k \subseteq \mathbb{R}^n$ open with $\Phi_k|_{U_k}$ one-to-one, $(\Phi_k|_{U_k})_*(dx) = d\mu$ on $\Phi_k(U_k)$. (Such a dilation exists because $d\mu$ and the pushforward of dx are both invariant under the flows of the X_j and hence differ from one another by a constant by Chow's theorem.) The vector fields $\widehat{X}_j := \Phi_k^* X_j$ are divergence-free and tangent to the fibers of $\widehat{\pi}_1 := \pi_1 \circ \Phi_k$ and $\widehat{\pi}_2 := \pi_2 \circ \Phi_k$, respectively. For β a multiindex, the $\widehat{X}_j$ give rise to a measure $\widehat{\rho}_\beta \, dx$ on $\mathbb{R}^n$. If $r \in \mathrm{Aut}(\Phi)$ is an element of the deck transformation group, then $\widehat{\rho}_\beta \circ r = \widehat{\rho}_\beta$, and thus we can define a measure μ_β on $M_{0,k}$ by setting $\mu_\beta|_{\Phi_k(U_k)} := (\Phi_k)_*(\widehat{\rho}_\beta \, dx|_{U_k})$ whenever $\Phi_k|_{U_k}$ is a diffeomorphism. We extend this to a measure on M by setting $\mu_\beta = 0$ on $M \setminus M_0$.

The measure μ plays a slightly lesser role than the ν_j in the construction of the μ_β. The measure μ affects the definition of the vector fields X_j, and hence the nilpotency hypothesis, but in the minimal case (that $\rho_{\beta'} \equiv 0$ for all $\beta' \prec \beta$), all choices of μ lead to the same definition of μ_β by (9.2). Moreover, in the case that β is minimal, by (9.2), the analogous construction carried out with respect to any choice of local coordinates on M would give rise to the same measure μ_β. When β is non minimal, the measure depends on the choice of coordinates, but in any coordinates, the analogue of μ_β would vanish on $M \setminus M_0$.

Theorem 9.3. *Under the notation and hypotheses above, let $V \subseteq M$ be an open set. For each k, let $V_k := V \cap M_{0,k}$, let $U_k \subseteq \mathbb{R}^n$ be an open set, and assume that for a.e. $x \in V_k$, $\#(\Phi_k^{-1}(x) \cap U_k) \geq A_k$, and for a.e. $y \in \pi_j(V_k)$, $U_k \cap \Phi_k^{-1}(\pi_j^{-1}\{y\})$ is contained in the union of at most $B_{j,k}$ integral curves of $\widehat{X}_j$, with $0 < A_k, B_{1,k}, B_{2,k} < \infty$. Then*

$$\left| \int_V f_1 \circ \pi_1(x) f_2 \circ \pi_2(x) \, d\mu_\beta(x) \right| \lesssim \left(\sup_k \frac{B_{1,k}^{1/p_1} B_{2,k}^{1/p_2}}{A_k}\right) \|f_1\|_{L^{p_1}(P_1;\nu_1)} \|f_2\|_{L^{p_2}(P_2;\nu_2)}.$$

(9.10)

Here the exponent pair $p = (p_1, p_2)$ is defined as in (1.8).

The quantities $A_k, B_{1,k}, B_{2,k}$ count (in the absence of the polynomial hypothesis) oscillations naturally associated to the π_j. The pre-image $\Phi_k^{-1}(x)$ is in one-to-one

correspondence with $\mathrm{Aut}(\Phi_k)$ and may be viewed as the set of distinct paths of the form $t \mapsto e^{tX}(x)$, $0 \le t \le 1$, $X \in \mathfrak{g}_M$, that start and end at x. Assume that $\pi_j(x) = y$. The set of $\widehat{X}_j$ integral curves containing $\Phi_k^{-1}(\pi_j^{-1}(y))$ equals the set of $\widehat{X}_j$ integral curves containing $\Phi_k^{-1}(x)$, and thus is in one-to-one correspondence with the set of distinct paths of the form $t \mapsto \pi_j(e^{tX}(x))$, $0 \le t \le 1$, $X \in \mathfrak{g}_M$, that start and end at y. Both of these sets are either singletons (the trivial loop) or are countable; intersecting with the set U as in the hypothesis of the theorem makes other finite cardinalities possible. As seen in the next section, the analogue of Theorem 9.3 without some accounting for oscillations is false.

Proof of Theorem 9.3. Let $P_{k,j} := \pi_j(M_{0,k})$. By hypothesis (ii), for each $j = 1, 2$, the $P_{k,j}$ have measure zero (pairwise) intersection. Therefore by Hölder's inequality and $p_1^{-1} + p_2^{-1} > 1$,

$$\sum_k \|f_1\|_{L^{p_1}(P_{k,1};\nu_1)} \|f_2\|_{L^{p_2}(P_{k,2};\nu_2)} \le \Big(\sum_k \|f_1\|_{L^{p_1}(P_{k,1};\nu_1)}^{p_1}\Big)^{\frac{1}{p_1}} \Big(\sum_k \|f_2\|_{L^{p_2}(P_{k,2};\nu_2)}^{p_2}\Big)^{\frac{1}{p_2}}$$
$$= \|f_1\|_{L^{p_1}(P_1;\nu_1)} \|f_2\|_{L^{p_2}(P_2;\nu_2)},$$

so it suffices to prove (9.10) when $V = V_k$ for some k. This reduces matters to consideration of the special case when M is connected and the elements of $\mathfrak{g}_M$ span the tangent space to M at every point, and we may henceforth omit the subscript k from the various objects in the setup of the theorem.

Define $\widetilde{P}_j := \mathbb{R}^n/[x \sim e^{t\widehat{X}_j}(x)]$, i.e. the set of all $\widehat{X}_j$ integral curves in $\mathbb{R}^n$, let $\widetilde{\pi}_j : \mathbb{R}^n \to \widetilde{P}_j$ denote the quotient map, and endow $\widetilde{P}_j$ with the quotient topology. The image $\widetilde{\pi}_j(\mathbb{R}^n \setminus \{\widehat{X}_j = 0\})$ is then a smooth $(n-1)$-dimensional manifold. Since $\pi_j \circ \Phi$ is constant along integral curves of $\widehat{X}_j$ (and hence on the level sets of $\widetilde{\pi}_j$), we may define a map $\widetilde{\Phi}_j : \widetilde{P}_j \to P_j$ by $\widetilde{\Phi}_j(\widetilde{\pi}_j(x)) := \pi_j(\Phi(x))$. We observe that $\widetilde{\Phi}_j$ is a local diffeomorphism from $\widetilde{\pi}_j(\mathbb{R}^n \setminus \{\widehat{X}_j = 0\})$ onto $\pi_j(M \setminus \{X_j = 0\})$. Our hypothesis on the B_j is precisely the statement that for a.e. $y \in \pi_j(V)$, $\#(\widetilde{\Phi}_j^{-1}(y) \cap \widetilde{\pi}_j(U)) \le B_j$.

Because the flows of the $\widehat{X}_j$ preserve Lebesgue measure, we may define Borel measures $\widetilde{\nu}_j$ on the $\widetilde{P}_j$ by setting $\widetilde{\nu}_j(\widetilde{\pi}_j(\{\widehat{X}_j = 0\})) = 0$ and, for every finite measure $\Omega \subseteq \{\widehat{X}_j \ne 0\}$,

$$|\Omega| = \int_{\widetilde{\pi}_j(\Omega)} \int_{\widetilde{\pi}_j^{-1}(y)} \chi_\Omega(t)\, |\widehat{X}_j(t)|^{-1} d\mathcal{H}^1(t)\, d\widetilde{\nu}_j(y).$$

Equivalently, if V is open and $\widetilde{\Phi}_j|_V$ is a diffeomorphism, $\widetilde{\nu}_j = (\widetilde{\Phi}_j|_V)_*^{-1}\nu_j$.

By the proof of Theorem 1.1, which did not use the global Euclidean structure of $\mathbb{R}^{n-1}$, nor its algebraic properties, nor the specific measure dx, but only the local Euclidean structure of $\pi_j(\{X_j \ne 0\})$,

$$\int_{\mathbb{R}^n} \prod_{j=1}^2 |\widetilde{f}_j \circ \widetilde{\pi}_j|\, \widehat{\rho}_\beta(x)\, dx \lesssim \prod_{j=1}^2 \|\widetilde{f}_j\|_{L^{p_j}(\widetilde{P}_j;\widetilde{\nu}_j)}, \tag{9.11}$$

for all pairs of continuous, compactly supported functions $\widetilde{f}_j$ on $\widetilde{P}_j$, $j = 1, 2$.

Taking V, U, A, B_1, B_2, Φ as in the hypothesis of the theorem and f_j a continuous function with compact support on P_j, $j = 1, 2$, we set $\widetilde{f}_j := (f_j \circ \widetilde{\Phi}_j)\chi_{\widetilde{\pi}_j(U)}$. By

construction,

$$\Big| \int_V \prod_{j=1}^2 f_j \circ \pi_j \, d\mu_\beta \Big| \leq \frac{1}{A} \int_U \prod_{j=1}^2 |\widetilde{f}_j \circ \pi_j \circ \Phi| \, \widehat{\rho}_\beta(x) \, dx = \frac{1}{A} \int_U \prod_{j=1}^2 |\widetilde{f}_j \circ \widetilde{\pi}_j| \, \widehat{\rho}_\beta(x) \, dx,$$

$$\text{and} \quad \int_{\widetilde{P}_j} |\widetilde{f}_j|^{p_j} \, d\widetilde{\nu}_j \leq B_j \int_{P_j} |f_j|^{p_j} \, d\nu_j.$$

Together with (9.11), the preceding two inequalities imply (9.10). $\qquad \square$

10. Examples, counter-examples, and open questions

The translation invariant case. We begin with a concrete example. The weights ρ_β were originally conceived in [25] as a generalization of the affine arclength measure associated to curves, and the results of this article include, as a special case, results on translation invariant averages on curves with affine arclength measure. Let $\gamma : \mathbb{R} \to \mathbb{R}^d$ be a polynomial of degree at most N. Consider the maps $\pi_j : \mathbb{R}^{d+1} \to \mathbb{R}^d$ given by

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

The vector fields associated to these maps are

$$X_1 = \tfrac{\partial}{\partial t}, \quad X_2 = \tfrac{\partial}{\partial t} + \gamma'(t) \cdot \nabla_x,$$

and X_1, X_2 generate a nilpotent Lie algebra on $\mathbb{R}^{d+1}$ whose elements have polynomial flows. As discussed in Section 4, it is slightly easier to compute determinants of vector fields arising as iterated Lie brackets of the X_i, rather than derivatives of Jacobian determinants, so we look to Theorem 4.2. Provided that the polytope $\mathcal{P}$ associated to X_1, X_2 via (4.1) is nonempty,

$$\mathcal{P} = \text{ch}\big\{ \big((\text{d}, 1 + \tfrac{\text{d}(\text{d}-1)}{2}) + [0, \infty)^2\big) \cup \big((1 + \tfrac{\text{d}(\text{d}-1)}{2}, \text{d}) + [0, \infty)^2\big) \big\}.$$

Thus minimal elements b of $\mathcal{P}$ lie on the line segment joining $(d, 1 + \tfrac{d(d-1)}{2})$ and $(1 + \tfrac{d(d-1)}{2}, d)$. The corresponding Lebesgue exponents are those (p_1, p_2) with (p_1^{-1}, p_2^{-1}) lying on the line segment joining

$$\Big(\tfrac{2d}{d(d+1)}, \tfrac{2 + d(d-1)}{d(d+1)} \Big), \qquad \Big(\tfrac{2 + d(d-1)}{d(d+1)}, \tfrac{2d}{d(d+1)} \Big),$$

and the corresponding weights are all equal:

$$|\lambda_I|^{\frac{1}{b_1 + b_2 - 1}} = |\det(\gamma'(t), \dots, \gamma^{(d)}(t))|^{\frac{2}{d(d+1)}}.$$

Theorem 4.2 thus states that

$$\Big| \iint g(x) f(x - \gamma(t)) \, |\det(\gamma'(t), \dots, \gamma^{(d)}(t))|^{\frac{2}{d(d+1)}} \, dt \, dx \Big| \lesssim_N \|g\|_{p_1} \|f\|_{p_2},$$

for all p_1, p_2 as above, which is precisely the main theorem of [23]. One may analogously obtain the main result of [10], which considered the X-ray transform restricted to polynomial curves, as a special case of Theorem 4.2.

Independence and necessity of Hypotheses (i) and (ii). Hypothesis (ii) of Theorem 1.1 certainly does not imply (i); nor does (i) imply (ii), as can be seen by considering, on the domain $U := (1, \infty) \times \mathbb{R} \times \mathbb{R}$, the maps

$$\pi_1(x, y, z) := (y, z), \qquad \pi_2(x, y, z) := (x \cos(y + \tfrac{z}{x}), x \sin(y + \tfrac{z}{x})), \qquad (10.1)$$

for which $X_1 = \partial_x$ and $X_2 = \partial_y - x\partial_z$.

Hypothesis (ii) can be weakened to the assumption that a bounded number of integral curves constitute each fiber; this can be carried out by factoring the π_j through the quotients of $\mathbb{R}^n$ by the integral curves of the X_j, as in the proof of Theorem 9.3. The necessity of some hypothesis in this direction follows from the example (10.1) above. Indeed, with this choice of π_1, π_2, (1.9) would suggest

$$|\int_U f_1 \circ \pi_1(x) \, f_2 \circ \pi_2(x) \, dx| \lesssim \|f_1\|_{3/2} \|f_2\|_{3/2},$$

which can be seen to fail for $f_1 := \chi_{\{|y| < R\}}$, $f_2 := \chi_{\{1 < |y| < 2\}}$, as $R \to \infty$.

We expect that hypothesis (i) can weakened substantially, though at a cost of losing some uniformity (as will be seen momentarily). Indeed, in the translation invariant case, this has been done [18, 11, 14]. That being said, the conclusions of the theorem are false if we completely omit this hypothesis. To see this, we consider first Sjölin's [19] counter-example

$$\pi_1(x) := (x_1, x_2) \qquad \pi_2(x) := (x_1, x_2) - (x_3, \phi(x_3)),$$

$$\phi(x_3) := \sin(x_3^{-k}) e^{-1/x_3}, \quad x \in \mathbb{R}^2 \times (0, \infty),$$

for k sufficiently large. Inequality (1.9) would suggest

$$|\int_{\{0 < x_3 < 1\}} f_1 \circ \pi_1(x) \, f_2 \circ \pi_2(x) \, |\phi''(x_3)|^{1/3} \, dx| \lesssim \|f_1\|_{3/2} \|f_2\|_{3/2},$$

but this can be seen to fail for the characteristic functions $f_j = \chi_{E_j}$,

$$E_1 := \{y \in \mathbb{R}^2 : \delta < y_1 < \delta + \delta^2, \quad |y_2| \le e^{-1/\delta}\}$$

$$E_2 := \{y \in \mathbb{R}^2 : |y_1| \lesssim \delta^2, \quad |y_2| \lesssim e^{-1/\delta}\},$$

as $\delta \searrow 0$.

Malcev coordinates and the linear operator. For simplicity, we consider the Euclidean case. We recall that we were initially interested in bilinear forms arising in the study of averages on curves, $\mathcal{B}(f_1, f_2) = \langle f_1, Tf_2 \rangle$, where

$$Tf_2(x) := \int f_2(\gamma_x(t)) \, d\mu_{\gamma_x}(t).$$

Thus we are particularly interested in the case when π_1 is a coordinate projection, and dualizing the linear operator corresponds to changing variables so that π_2 is a coordinate projection. As we will see, weak Malcev coordinates are sometimes useful in carrying this out.

Fix a nilpotent Lie algebra $\mathfrak{g}$ generated by vector fields $X_1, X_2 \in \mathfrak{g}$. Let N denote the dimension of $\mathfrak{g}$, and let $\mathfrak{z}$ denote an $(N-n)$-dimensional Lie subalgebra of $\mathfrak{g}$. As we have seen, there exists a weak Malcev basis $\{W_1, \ldots, W_N\}$ for $\mathfrak{g}$, with $\{W_{n+1}, \ldots, W_N\}$ a basis for $\mathfrak{z}$, and, in the coordinates

$$(x_1, \ldots, x_n, z_{n+1}, \ldots, z_N) \mapsto e^{x_1 W_1} \cdots e^{x_n W_n} e^{z_{n+1} W_{n+1}} \cdots e^{z_N W_N}$$

for the associated Lie group $G := \exp(\mathfrak{g})$, the flows of the elements of $\mathfrak{g}$ are polynomial, and, moreover, the projection map $(x, z) \mapsto x$ defines a Lie group isomorphism of $\mathfrak{g}$ onto a Lie subgroup of $\mathbb{R}^n$, in which $\mathfrak{z}$ pushes forward to $\mathfrak{z}_0$, the algebra consisting of vector fields in (the pushforward of) $\mathfrak{g}$ that vanish at 0. Thus we may identify $\mathbb{R}^n$ with G/Z, where $Z := \exp(\mathfrak{z})$.

If $W_1 = X_1$, then we define $\pi_1(x) := (x_2, \ldots, x_n)$. (Alternately, there are local coordinates in which π_1 may be written in this form.) If there exists another weak Malcev basis $\{\widetilde{W}_1, \ldots, \widetilde{W}_N\}$ for $\mathfrak{g}$ through $\mathfrak{z}$ with $\widetilde{W}_1 = X_2$, then the map $F : x \mapsto \widetilde{x}$ is a polynomial diffeomorphism, and so $\pi_2(x) := (\widetilde{x}_2, \ldots, \widetilde{x}_n)$ is also a polynomial. The map F, being a polynomial diffeomorphism, has constant Jacobian determinant. By scaling the $\widetilde{W}_j$, we may assume that this constant is 1. Our bilinear form is

$$B(f_1, f_2) = \int f_1 \circ \pi_1(x) \, f_2 \circ \pi_2(x) \, \rho_\beta(x) \, dx$$

$$= \int f_1 \circ \pi_1 \circ F^{-1}(x) \, f_2 \circ \pi_2 \circ F^{-1}(x) \, \rho_\beta \circ F^{-1}(x) \, dx.$$

Thus the associated linear and adjoint operators are

$$Tf(y) = \int f(\pi_2(t, y)) \, \rho_\beta(t, y) \, dt, \quad T^*g(y) = \int g(\pi_1 \circ F^{-1}(t, y)) \, \rho_\beta \circ F^{-1}(t, y) \, dt,$$

averages along curves parametrized by polynomials.

It is therefore natural to ask when it is possible to find a weak Malcev basis of $\mathfrak{g}$ through $\mathfrak{z}$ whose first element is X_1.

Initially fix any weak Malcev basis $\{W_1, \ldots, W_N\}$. Let $\mathfrak{g}^{(2)} := [\mathfrak{g}, \mathfrak{g}]$, and let $\mathfrak{h} := \mathfrak{g}^{(2)} + \mathfrak{z}$. Then $\mathfrak{h}$ is an ideal in $\mathfrak{g}$. In fact, it is a proper ideal, because the linear span $\mathbb{R}W_2 + \cdots + \mathbb{R}W_N$ is an ideal (being a codimension 1 subalgebra) of $\mathfrak{g}$ that contains both $\mathfrak{g}^{(2)}$ and $\mathfrak{z}$. Since $\mathfrak{h} \subseteq \mathbb{R}W_2 + \cdots + \mathbb{R}W_N$, if $X_1 \in \mathfrak{h}$, we cannot take $W_1 = X_1$. If $X_1 \notin \mathfrak{h}$, then there exists a weak Malcev basis $\{W_k, \ldots, W_N\}$ of $\mathfrak{h}$ through $\mathfrak{z}$, which we may complete to a basis $\mathcal{B} := \{X_1, W_2, \ldots, W_N\}$ of $\mathfrak{g}$. For each $2 \leq j \leq k$, the linear span $\mathbb{R}W_j + \cdots + \mathbb{R}W_N$ is an ideal in $\mathfrak{g}$, so $\mathcal{B}$ is a weak Malcev basis of $\mathfrak{g}$ through $\mathfrak{z}$.

Since X_1, X_2 generate $\mathfrak{g}$, both cannot lie in the proper subideal $\mathfrak{h}$, and so there does exist a weak Malcev basis with either X_1 or X_2 as the first element.

Malcev coordinates aside, we can ask when it is possible to express π_1 as a coordinate projection and π_2 as a polynomial, without changing the Lie algebra. The authors have not strenuously endeavored to determine necessary and sufficient conditions, but it is clear that it is not possible in general, even locally around points where both maps are submersions. Indeed, local polynomial maps extend to global ones generating the same Lie algebra, and a necessary condition for π_1 to be a coordinate projection is that

$$X_1 \notin \{X * Z * (-X) : Z \in \mathfrak{z}_0, \, X \in \mathfrak{g}\},$$

since $\mathfrak{z}_{e^X(0)} = \{X * Z * (-X) : Z \in \mathfrak{z}_0\}$.

Optimality of the weight. It is proved in [25] that if b is an extreme point of the Newton polytope $\mathcal{P}$ defined in (4.1), then the corresponding weight ρ_β is (up to summing weights corresponding to the same degree) the largest possible weight for which (1.9) can hold. It is also shown that if b is not on the boundary of $\mathcal{P}$,

then it is not possible to establish a pointwise bound on weights ρ for which (1.9) might hold.

Changes of speed and failure of global bounds. The analogue of Proposition 9.1 with U_{x_0} replaced by the full region U can fail if β is not minimal and the Hodge-star vector fields are not themselves nilpotent, even when the Hodge-star vector fields are real analytic and have flows satisfying natural convexity hypotheses. We see this by considering the example $U := \{x \in \mathbb{R}^3 : x_3 > 0\}$ and

$$\pi_1(x) := (x_1, x_2), \qquad \pi_2(x) := (x_1, x_2) - (\log x_3, (\log x_3)^2).$$

The Hodge-star vector fields are

$$X_1 = \partial_3, \qquad X_2 = \frac{1}{x_3}\partial_1 + \frac{2}{x_3}\log x_3 \partial_2 + \partial_3.$$

Taking $Y_1 := x_3 X_1$ and $Y_2 := x_3 X_2$, we have $Y_{12} = 2\partial_2$, and all higher order commutators are zero. Taking $\beta = (0, 2, 0)$,

$$\partial_t^\beta|_{t=0} \det D_t\, e^{t_3 X_1} \circ e^{t_2 X_2} \circ e^{t_1 X_1}(x) = -\frac{3}{x_3^4}.$$

Thus (9.8) would suggest the bound

$$|\int_U f_1 \circ \pi_1(x)\, f_2 \circ \pi_2(x)\, x_3^{-1}\, dx| \lesssim \|f_1\|_2\|f_2\|_{4/3}. \tag{10.2}$$

Changing variables, (10.2) becomes

$$|\int_{\mathbb{R}^3} f_1(x_1, x_2)\, f_2(x_1 - t, x_2 - t^2)\, dt\, dx| \lesssim \|f_1\|_2\|f_2\|_{4/3},$$

which is easily seen to be false by scaling.

It is still conceivable that global bounds are possible in the real analytic case when β is minimal and some convexity/non-oscillation assumption is made.

Failure of strong type bounds in dimension 2. The hypothesis $n \geq 3$ in Theorem 1.1 cannot be omitted. Indeed, consider $\pi_1(x_1, x_2) := x_1$, $\pi_2(x_1, x_2) := x_2^k$. Then $X_1 = \frac{\partial}{\partial x_2}$, $X_2 = kx_2^{k-1}\frac{\partial}{\partial x_1}$, which together generate a nilpotent Lie algebra with polynomial flows. Moreover, if we take $\beta = (k - 1, 0)$, then the corresponding weight is $\rho_\beta \sim 1$, so (1.9) would suggest

$$|\int f_1(x_1)f_2(x_2^k)\, dx_1\, dx_2| \lesssim \|f_1\|_1\|f_2\|_k,$$

which is false in general (take e.g. $f_2(y) = (y^{\frac{1}{k}}\log y)^{-1}\chi_{(0,1]}$).

We recall, however, that the argument in [13] (and also the proof of Proposition 2.1) did not require the hypothesis $n \geq 3$ to obtain the restricted weak type inequality on the single scale $\rho_\beta \sim 1$.

Failure of global estimates for an analogue on manifolds. An interesting question that we do not investigate is whether there are natural, simple hypotheses leading to global estimates in Theorem 9.3. Without further hypotheses, such a result is false, as can be seen in the following counterexample.

Let $M = \mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z} \times \mathbb{R}$ and $P_1 = P_2 = \mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z}$, all equipped with Lebesgue measure. Define projections

$$\pi_1(\theta_1, \theta_2, t) = (\theta_1, \theta_2), \qquad \pi_2(\theta_1, \theta_2, t) = (\theta_1 + t, \theta_2 + t^2).$$

Then the π_j naturally give rise to the vector fields

$$X_1 = \tfrac{\partial}{\partial t}, \qquad X_2 = \tfrac{\partial}{\partial t} - \tfrac{\partial}{\partial \theta_1} - 2t\tfrac{\partial}{\partial \theta_2}.$$

These generate a nilpotent Lie algebra obeying the Hörmander condition, and, moreover, each point of P_j has its π_j-preimage contained in a unique integral curve of X_j. A naive analogue of Theorem 1.1 might suggest

$$\int_M |f_1 \circ \pi_1 \, f_2 \circ \pi_2| \lesssim \|f_1\|_{\frac{3}{2}} \|f_2\|_{\frac{3}{2}},$$

but this is obviously false, as can be seen by taking $f_1 \equiv f_2 \equiv 1$.

Multilinear averages on curves. In the multilinear case considered in [24, 25], the natural generalization of the map Ψ_x used to define ρ involves iteratively exponentiating the vector fields in some specified order, and the single-scale restricted weak type inequality is known to hold under the natural analogue of the hypotheses of Theorem 1.1. Indeed, the proof in Section 2 readily generalizes. Unfortunately, the analogy breaks down in Section 5, where we need to use the gain coming from nonzero entries of the multiindex β. To rule out such examples in the multilinear case would require rather more complicated hypotheses, particularly if we want a theory that includes examples such as the perturbed Loomis–Whitney inequality, where the endpoint bounds are known to hold [1].

We record here two multilinear examples that may be of interest in future explorations of this topic.

The first is a Loomis–Whitney inspired variant on the above two-dimensional example. Define

$$\pi_i(x) := (x_1, \ldots, \widehat{x_i}, \ldots, x_n), \quad 1 \le i \le n-1, \qquad \pi_n(x) := (x_1^k, x_2, \ldots, x_{n-1}).$$

Our vector fields are $X_i = \tfrac{\partial}{\partial x_i}$, $1 \le i \le n-1$, and $X_n = kx_1^{k-1}\tfrac{\partial}{\partial x_n}$, and the endpoint inequality

$$\left| \int \prod_i f_i \circ \pi_i \, dx \right| \lesssim \prod_{i=1}^n \|f_i\|_{p_i}, \qquad p_1 = \tfrac{n+k-2}{k}, \; p_i = n+k-2, \; i = 2, \ldots, n$$

is false for $k > 1$, as can be seen by considering $f_1 := \chi_{B_1}$ and $f_i(x_1, x') := |x_1|^{-\frac{1}{n+k-2}} |\log|x_1||^{-\frac{1}{n-1}} \chi_{B_1}(x)$, where B_1 denotes the unit ball.

The second is a hybrid of a well-studied convolution operator with this example. For $(x, t, s) \in \mathbb{R}^{n+1+1}$, let $\pi_1(x, t, s) := (x, s)$, $\pi_2(x, t, s) := (x - \gamma(t), s)$, $\pi_3(x, t, s) := (x, t^k)$, where $\gamma(t) := (t, t^2, \ldots, t^n)$. Our vector fields are $X_1 = \tfrac{\partial}{\partial t}$, $X_2 = \tfrac{\partial}{\partial t} - \gamma'(t) \cdot \nabla_x$, $X_3 = kt^{k-1}\tfrac{\partial}{\partial s}$. From the preceding examples, we might guess that the endpoint inequality

$$\left| \int \prod_{i=1}^3 f_i \circ \pi_i \, dx \right| \lesssim \prod_{i=1}^3 \|f_i\|_{p_i},$$

$$p_1 := \frac{2k + n(n+1)}{2k + n(n-1)}, \qquad p_2 := \frac{2k + n(n+1)}{2n}, \qquad p_3 := \frac{2k + n(n+1)}{2}$$

fails. In fact, this inequality is true, as can be seen from Hölder's inequality and Theorem 2.3 of [11].

11. Appendix: Polynomial lemmas

In this section, we collect together a number of lemmas on the size and injectivity of polynomials.

The next lemma shows that if a polynomial bounds a monomial, then the monomial must in fact be bounded by two terms of the polynomial; this facilitates a complex interpolation argument used in the deduction of Theorem 4.2 from Theorem 1.1.

Lemma 11.1. *Let $p(t) = \sum_{n=0}^{N} a_n t^n$ be a polynomial with nonnegative coefficients, and let $k \in \mathbb{Z}_{\geq 0}$. If $t^k \leq p(t)$ for all $t > 0$, then $a_k \gtrsim 1$ or there exist $n_1 < k < n_2$ such that $(a_{n_1})^{\frac{n_2-k}{n_2-n_1}}(a_{n_2})^{\frac{k-n_1}{n_2-n_1}} \gtrsim 1$. Conversely, if $a_k \geq 1$ or $(a_{n_1})^{\frac{n_2-k}{n_2-n_1}}(a_{n_2})^{\frac{k-n_1}{n_2-n_1}} \geq 1$ for some $n_1 < k < n_2$, then $t^k \leq p(t)$ for all $t > 0$.*

Proof. If $t^k \leq p(t)$ for all $t \geq 0$, but $a_k \leq \frac{1}{2}$, then $t^k \leq 2(p(t) - a_k t^k)$, so we may as well assume that $a_k = 0$.

Let $p_{lo}(t) := \sum_{n<k} a_n t^n$ and $p_{hi}(t) := \sum_{n>k} a_n t^n$. By considering small t, we see that $p_{lo} \not\equiv 0$, and by considering large t, we see that $p_{hi} \not\equiv 0$. By a routine application of the Intermediate Value Theorem, there exists a unique $t_0 > 0$ such that $p_{lo}(t_0) = p_{hi}(t_0)$. We may choose $n_1 < k < n_2$ such that $p_{lo}(t_0) \sim a_{n_1} t_0^{n_1}$ and $p_{hi}(t_0) \sim a_{n_2} t_0^{n_2}$. Thus $t_0^k \lesssim a_{n_1} t_0^{n_1} \sim a_{n_2} t_0^{n_2}$, from which we learn that $t_0 \sim (\frac{a_{n_1}}{a_{n_2}})^{\frac{1}{n_2-n_1}}$ and, consequently, $1 \lesssim (a_{n_1})^{\frac{n_2-k}{n_2-n_1}}(a_{n_2})^{\frac{k-n_1}{n_2-n_1}}$. In the converse direction, if $(a_{n_1})^{\frac{n_2-k}{n_2-n_1}}(a_{n_2})^{\frac{k-n_1}{n_2-n_1}} \geq 1$, then at $t_0 := (\frac{a_{n_1}}{a_{n_2}})^{\frac{1}{n_2-n_1}}$, $t_0^k \leq a_{n_1} t_0^{n_1} = a_{n_2} t_0^{n_2}$, so $t^k \leq a_{n_1} t^{n_1}$ for all $t \leq t_0$ and $t^k \leq a_{n_2} t^{n_2}$ for all $t \geq t_0$. $\qquad\square$

A lemma in [9] states that if $\mathcal{P}$ is a finite collection of polynomials on $\mathbb{R}$, each of degree at most N, then there exists a decomposition $\mathbb{R} = \bigcup_{j=1}^{C(\#\mathcal{P},N)} I_j$, with each I_j an interval, such that on I_j, each p has roughly the same size as some fixed monomial, centered at a point that depends only on j, not p:

$$|p(t)| \sim a_{p,j}(t - b_j)^{k_{p,j}}, \qquad a_{p,j} \in [0,\infty), \quad b_j \notin I_j, \quad k_{p,j} \geq 0.$$

Our next lemma strengthens this to show that we may take each monomial to be an entry of the Taylor polynomial centered at b_j of the polynomial p and ensures that the other entries of that Taylor polynomial are as small as we like.

Lemma 11.2. *Let $\mathcal{P}$ denote a finite collection of polynomials on $\mathbb{R}$, each having degree at most N, and let $\varepsilon > 0$. There exist a collection of nonoverlapping open intervals $I_1, \ldots, I_{N'}$, with $N' = N'(N, \#\mathcal{P}, \varepsilon)$ and $\mathbb{R} = \bigcup_j \overline{I_j}$, and centers $b_1, \ldots, b_{N'}$, with $b_j \notin I_j$, such that for each j and $p \in \mathcal{P}$, there exists an integer $k_{j,p}$ such that*

$$|\tfrac{1}{k!}p^{(k)}(b_j)(t - b_j)^k| \leq \varepsilon \tfrac{1}{k_{j,p}!}|p^{(k_{j,p})}(b_j)(t - b_j)^{k_{j,p}}|, \qquad k \neq k_{j,p}, \; t \in I_j. \quad (11.1)$$

In particular, provided we take $\varepsilon < \frac{1}{2N}$,

$$|p(t)| \sim A_{j,p}|t - b_j|^{k_{j,p}}, \text{ for } t \in I_j, \text{ where } A_{j,p} := |\tfrac{1}{k_{j,p}!}p^{(k_{j,p})}(b_j)|. \quad (11.2)$$

Proof of Lemma 11.2. We modify the approach from [9]. We will allow the integer N' to change from line to line, subject to the constraint $N' = N'(N, \#\mathcal{P}, \varepsilon)$.

Without loss of generality, all elements of $\mathcal{P}$ are nonconstant, and $\mathcal{P}$ contains all nonconstant derivatives of its elements. Let $\{z_1, \ldots, z_{N'}\}$ denote the union of the

(complex) zero sets of the elements in $\mathcal{P}$. Set

$$S_i := \{t \in \mathbb{R} : |t - z_i| \le |t - z_j|, \ j \ne i\}.$$

Then $\mathbb{R} = \bigcup_i S_i$. We will further decompose each S_i, and by reindexing, it suffices to further decompose S_1. Reindexing, we may assume that $|z_1 - z_2| \le \cdots \le |z_1 - z_N|$. Define

$$T_j := \{t \in S_1 : \tfrac{1}{2}|z_1 - z_j| \le |t - z_1| < \tfrac{1}{2}|z_1 - z_{j+1}|\}, \qquad j = 1, \ldots, N' - 1,$$
$$T_{N'} := S_1 \setminus T_{N'-1}.$$

If $t \in T_j$, then by the triangle inequality,

$$|t - z_1| \le |t - z_{j'}| \le 3|t - z_1|, \quad j' \le j, \qquad \tfrac{1}{2}|z_1 - z_{j'}| \le |t - z_{j'}| < \tfrac{3}{2}|z_1 - z_{j'}|, \quad j' > j.$$

Writing $p(t) = A_p \prod_{j' \in \mathcal{J}_p} (t - z_j)^{n_{j',p}}$, where $\mathcal{J}_p$ denotes the set of indices corresponding to zeros of p,

$$|p(t)| \sim |A_p \prod_{j < j' \in \mathcal{J}_p} (z_1 - z_{j'})^{n_{j',p}} \prod_{j > j' \in \mathcal{J}_p} (t - z_1)^{n_{j',p}}|.$$

Thus p is comparable to a complex monomial. Let $b_1 := \operatorname{Re} z_1$, $c_1 := |\operatorname{Im} z_1|$. On $\{|t - b_1| \le c_1\}$, $|t - z_1| \sim c_1$, and on $\{|t - b_1| \ge c_1\}$, $|t - z_1| \sim |t - b_1|$, so subdividing one more time, we obtain intervals on which each polynomial is comparable to a real monomial.

More precisely, at this point, we have simply reproved the lemma from [9]: There exists a decomposition $\mathbb{R} = \bigcup_{j=1}^{N'} I_j$ such that $|p(t)| \sim a_{p,j}|t - b_j|^{n_{p,j}}$, $p \in \mathcal{P}$ and $t \in I_j$. We want a bit more, which requires us to subdivide further. Reindexing, it suffices to subdivide I_1. Translating, we may assume that $b_1 = 0$, and by symmetry, we may assume that $I_1 = (\ell, r) \subseteq (0, \infty)$. To fix our notation,

$$|p(t)| \sim a_p|t|^{n_p}, \qquad t \in I := I_1. \tag{11.3}$$

If $I = (0, \infty)$, then each p must in fact be a monomial, and we are done. Otherwise, by rescaling, we may assume that either $r = 1$ or that $\ell = 1, r = \infty$.

Case I: $I = (\ell, 1)$. By construction, z_1, which is purely imaginary, is no further from 1 than any zero of any nonzero derivative of any element of $\mathcal{P}$. Thus no element of $\mathcal{P}$ (nor any nonzero derivative of any element) has a zero inside the disk $\{|z - 1| < 1\}$. Therefore for each $p \in \mathcal{P}$, $|p|$ is monotone on $(0, 2)$. If $|p|$ is decreasing, by equivalence of norms,

$$|p(0)| = \|p\|_{C^0(0,2)} \sim \|p\|_{C^0(1,2)} = |p(1)|.$$

Thus either $|p|$ is increasing or $|p| \sim c_p$ on all of $(0, 1)$. In either case, for $t \in (0, 1)$,

$$|p(t)| \sim \|p\|_{L^\infty((0,t))} \sim \sum_{j=0}^N \tfrac{1}{j!}|p^{(j)}(0)|t^j \sim \sum_{j=0}^N \tfrac{1}{j!}\|p^{(j)}\|_{L^\infty((0,t))}t^j. \tag{11.4}$$

Let $\varepsilon^{-1} \le j < \varepsilon^{-2}$, and let $n \ge 1$. By (11.4), $|p^{(n)}(j\varepsilon^2)|(j\varepsilon^2)^n \lesssim |p(j\varepsilon^2)|$, so $|p^{(n)}(j\varepsilon^2)|(\varepsilon^2)^n \lesssim \varepsilon^n|p(j\varepsilon^2)|$. Therefore (11.1) holds for $t \in [j\varepsilon^2, (j+1)\varepsilon^2]$ with $b_j = j\varepsilon^2$ and $k_{j,p} = 0$.

It remains to decompose (ℓ, ε), supposing this interval is nonempty. Evaluating (11.4) at $t = 1$, and recalling (11.3),

$$\sum_{n > n_p} \tfrac{1}{n!}|p^{(n)}(0)| \lesssim a_p.$$

Thus for $0 < t < \varepsilon$ and $n > n_p$, $|p^{(n)}(0)|t^n < \varepsilon a_p t^{n_p}$. Evaluating (11.4) at $t = \ell$,

$$\sum_{n < n_p} |\tfrac{1}{n!} p^{(n)}(0)| \ell^n \lesssim a_p \ell^{n_p},$$

so for $t > \varepsilon^{-1}\ell$ and $n < n_p$, $|p^{(n)}(0)|t^n < \varepsilon a_p t^{n_p}$. Therefore (11.1) holds on $(\varepsilon^{-1}\ell, \varepsilon)$ with $b_j = 0$ and $k_{j,p} = n_p$. This leaves us to decompose $(\ell, \varepsilon^{-1}\ell)$. By scaling, this is equivalent to decomposing $(\varepsilon, 1)$, which we have already shown how to do.

Case II: $I = (1, \infty)$. By construction, z_1 is nearer to each $t > 1$ than any zero of any derivative of any element of $\mathcal{P}$. Thus no element of $\mathcal{P}$ has a zero with positive real part, so each $|p(t)|$ is nonvanishing with nonvanishing derivative on $(0, \infty)$, and thus must be increasing on $(0, \infty)$. Therefore (11.4) holds for each $t \in (0, \infty)$. Taking limits, we see that for $0 \neq p \in \mathcal{P}$, $n_p = \deg p$ and $a_p = \tfrac{1}{n_p!} p^{(n_p)}(0)$. Evaluating at 1, $\sum_n \tfrac{1}{n!}|p^{(n)}(0)| \lesssim \tfrac{1}{n_p!}|p^{(n_p)}(0)|$, so for $t > \varepsilon^{-1}$ and $n < n_p$, $|p^{(n)}(0)|t^n < \varepsilon |p^{(n_p)}(0)|t^{n_p}$. This leaves us to decompose $(1, \varepsilon^{-1})$, which rescales to $(\varepsilon, 1)$, so the proof is complete. $\qquad\square$

The next lemma applies Lemma 11.2 to make precise the heuristic that products of polynomials must vary at least as much as the original polynomials.

Lemma 11.3. *Let p_1 and p_2 be polynomials on $\mathbb{R}$ of degree at most N, and let a_1, a_2 be positive integers. The number of integers k for which there exists $t_k \in \mathbb{R}$ such that*

$$|p_1(t_k)| \sim 2^{a_1 k}, \qquad |p_2(t_k)| \sim 2^{-a_2 k} \tag{11.5}$$

is bounded by a constant depending only on N.

Proof of Lemma 11.3. The conclusion is trivial for monomials, so by Lemma 11.3, it follows for arbitrary polynomials. $\qquad\square$

The next lemma extends Lemma 11.2 to polynomial curves $\gamma : \mathbb{R} \to \mathbb{R}^n$, allowing us to closely approximate a given polynomial curve by a constant vector multiple of a monomial.

Lemma 11.4. *Let N be an integer and let $\varepsilon > 0$. Let $\gamma : \mathbb{R} \to \mathbb{R}^n$ be a polynomial of degree at most N. There exist nonoverlapping open intervals $I_1, \ldots, I_{N'}$, with $N' = N'(N, n, \varepsilon)$ and $\mathbb{R} = \bigcup_j \overline{I_j}$, and centers $b_1, \ldots, b_{N'}$, with $b_j \notin I_j$, such that for each j, there exists an integer k_j such that*

$$|\tfrac{1}{k!}\gamma^{(k)}(b_j)(t - b_j)^k| \leq \varepsilon |\tfrac{1}{k_j!}\gamma^{(k_j)}(b_j)(t - b_j)^{k_j}|, \qquad k \neq k_j, t \in I_j. \tag{11.6}$$

Proof of Lemma 11.4. By Lemma 11.2, it suffices to decompose an interval $I \subseteq \mathbb{R}$ for which there exists a point $b \notin I$ and integers $0 \leq k_1, \ldots, k_n \leq N$ such that the coordinates of γ satisfy

$$|\tfrac{1}{k!}\gamma_i^{(k)}(b)(t - b)^k| < \varepsilon |\tfrac{1}{k_i!}\gamma_i^{(k_i)}(b)(t - b)^{k_i}|, \qquad t \in I, k \neq k_i. \tag{11.7}$$

Making a finite decomposition of I and reindexing our coordinates if needed, we may assume that

$$|\tfrac{1}{k_1!}\gamma_1^{(k_1)}(b)(t - b)^{k_1}| \geq |\tfrac{1}{k_i!}\gamma_i^{(k_i)}(b)(t - b)^{k_i}|, \qquad i = 2, \ldots, n, t \in I.$$

Thus for $t \in I$,

$$|\tfrac{1}{k!}\gamma^{(k)}(b)(t - b)^k| \lesssim |\tfrac{1}{k_1!}\gamma^{(k_1)}(b)(t - b)^{k_1}|, \qquad |\gamma(t)| \sim |\tfrac{1}{k_1!}\gamma^{(k_1)}(b)(t - b)^{k_1}|. \tag{11.8}$$

Translating, reflecting, and rescaling, we may assume that $b = 0$ and that $I = (\ell, r)$.

Define curves

$$\gamma_{lo}(t) := \sum_{k<k_1} \frac{1}{k!}\gamma^{(k)}(0)t^k, \qquad \gamma_{hi}(t) := \sum_{k>k_1} \frac{1}{k!}\gamma^{(k)}(0)t^k, \qquad \gamma_{k_1}(t) := \frac{1}{k_1!}\gamma^{(k_1)}(0)t^{k_1}$$

and intervals

$$I_{lo}^0 := (\ell\varepsilon^{-1}, r), \qquad I_{lo}^j := (\ell\varepsilon^2 j, \ell\varepsilon^2(j+1)), \quad \varepsilon^{-2} \le j < \varepsilon^{-3}$$
$$I_{hi}^0 := (\ell, r\varepsilon), \qquad I_{hi}^j := (r\varepsilon^2 j, r\varepsilon^2(j+1)), \quad \varepsilon^{-1} \le j < \varepsilon^{-2}.$$

Since I is a bounded union of 'lo' intervals and also a bounded union of 'hi' intervals, I may be written as a bounded union of intersections of one 'lo' interval with one 'hi' interval. We will show that such intersections have the properties claimed in the lemma.

Evaluating (11.8) at ℓ, $\frac{1}{k!}|\gamma^{(k)}(0)|\ell^k \lesssim \frac{1}{k_1!}|\gamma^{(k_1)}(0)|\ell^{k_1}$. Therefore

$$|\gamma_{lo}(t)| \lesssim \varepsilon|\gamma_{k_1}(t)|, \qquad t \in I_{lo}^0.$$

By (11.8) and a Taylor expansion of $\gamma^{(m)}$ about 0,

$$|\gamma^{(m)}(t)|t^m \lesssim |\gamma^{(k_1)}(0)|t^{k_1} \lesssim |\gamma(t)|, \qquad t \in I.$$

Thus for $m \ge 1$, $\varepsilon^{-2} \le j \le \varepsilon^{-3}$ and $t \in I_{lo}^j$,

$$|\gamma^{(m)}(\ell\varepsilon^2 j)|(t - \ell\varepsilon^2 j)^m \le j^{-m}|\gamma^{(m)}(\ell\varepsilon^2 j)|(\ell\varepsilon^2 j)^m \lesssim \varepsilon|\gamma(\ell\varepsilon^2 j)|.$$

Arguing analogously,

$$|\gamma_{hi}(t)| \lesssim \varepsilon|\gamma_{k_1}(t)|, \qquad t \in I_{hi}^0,$$
$$|\gamma^{(m)}(r\varepsilon^2 j)|(t - r\varepsilon^2 j)^m \lesssim \varepsilon|\gamma(r\varepsilon^2 j)|, \qquad m \ge 1, \, t \in I_{hi}^j, \, \varepsilon^{-1} \le j < \varepsilon^{-2}.$$

Putting these inequalities together, (11.6) holds:
-On $I_{lo}^0 \cap I_{hi}^0$ with center $b_0 = 0$ and $k_0 = k_1$
-On $I_{lo}^{j_1} \cap I_{hi}^{j_2}$, for $(j_1, j_2) \ne (0,0)$, with center $b_j = \ell\varepsilon^2 j$ and $k_0 = 0$.
$\square$

The next lemma applies Lemma 11.4 to make precise the heuristic that, for $\gamma : \mathbb{R} \to \mathbb{R}^n$, since the derivative γ' drives the curve forward, γ and γ' are typically almost parallel. This result is crucial to proving Proposition 6.1.

Lemma 11.5. *There exists $M = M(N)$ sufficiently large that for all $\varepsilon > 0$, there exists $\delta > 0$ such that if*

$$|\gamma(t_i)| < \delta|\gamma(t_{i+1})|, \qquad i = 1, \ldots, M-1, \tag{11.9}$$

then

$$|\gamma(t_i) \wedge \gamma'(t_i)| < \varepsilon|\gamma(t_i)||\gamma'(t_i)|, \tag{11.10}$$

for some $1 \le i \le M$.

Proof of Lemma 11.5. Performing a harmless translation and applying Lemma 11.4, it suffices to prove that there exists M such that (11.10) holds whenever (11.9) holds with all t_i lying in some interval I on which

$$\left|\frac{1}{k!}\gamma^{(k)}(0)t^k\right| \le \left|\frac{1}{k_0!}\gamma^{(k_0)}(0)t^{k_0}\right| \sim |\gamma(t)|, \tag{11.11}$$

for all $0 \le k \le N$ and some $0 \le k_0 \le N$. Moreover, by (11.9), we may assume that $k_0 \ne 0$.

For $\delta > 0$ sufficiently small, and each $k_0 \neq k$, by (11.9) and (11.11) the inequality

$$\varepsilon |\tfrac{1}{k_0!} \gamma^{(k_0)}(0) t_i^{k_0}| < |\tfrac{1}{k!} \gamma^{(k)}(0) t_i^k|$$

can only hold for a bounded number of t_i, so we may assume further that

$$|\tfrac{1}{k!} \gamma^{(k)}(0) t_i^k| < \varepsilon |\tfrac{1}{k_0!} \gamma^{(k_0)}(0) t_i^{k_0}|, \qquad k \neq k_0.$$

Therefore

$$|\gamma(t_i) \wedge \gamma'(t_i)| \leq |(\sum_{k \neq k_0} \tfrac{1}{k!} \gamma^{(k)}(0) t_i^k \wedge \tfrac{k_0}{k_0!} \gamma^{(k_0)}(0) t_i^{k_0 - 1}|$$

$$+ |\tfrac{1}{k_0!} \gamma^{(k_0)}(0) t_i^{k_0} \wedge (\sum_{k \neq k_0} \tfrac{k}{k!} \gamma^{(k)}(0) t_i^k)|$$

$$+ |(\sum_{k \neq k_0} \tfrac{1}{k!} \gamma^{(k)}(0) t_i^k) \wedge (\sum_{k \neq k_0} \tfrac{k}{k!} \gamma^{(k)}(0) t_i^k)|$$

$$\lesssim \varepsilon |\gamma(t_i)| |\gamma'(t_i)|.$$

$\square$

Next, we use basic facts from algebraic geometry to prove several lemmas about polynomials of n variables of degree at most N. We say that a quantity is bounded if it is bounded from above by a constant depending only on the dimension n and the degree N, not on the particular polynomials in question.

Our main tool for lemmas below is the following theorem from algebraic geometry.

Theorem 11.6 ([12]). *Let $f_1, \ldots, f_k : \mathbb{C}^n \to \mathbb{C}$ be polynomials of degree at most N and let $Z \subseteq \mathbb{C}^n$ be the associated variety, i.e.*

$$Z := \{z \in \mathbb{C}^n : f_1(z) = \cdots = f_k(z) = 0\}.$$

Then we may decompose

$$Z = \bigcup_{i=1}^{C(k,n,N)} Z_i, \tag{11.12}$$

where each Z_i is an irreducible variety.

In particular, the decomposition in (11.12) involves a bounded number of dimension zero irreducible subvarieties. We recall, and will repeatedly use the fact that the irreducible subvariety containing an isolated point of Z must be a singleton.

Theorem 11.6 follows from the refined version of Bezout's Theorem, Example 12.3.1 of [12], which implies that $\sum_{i=1}^s \deg(Z_i) \leq \prod_{i=1}^k \deg(f_i)$. Since $\deg Z_i \geq 1$ for each i, this suffices.

Lemma 11.7. *Let $P : \mathbb{R}^n \to \mathbb{R}^n$ be a polynomial. Then, with respect to Lebesgue measure on $\mathbb{R}^n$, almost every point in $P(\mathbb{R}^n)$ has a bounded number of preimages.*

Proof. It suffices to show that if $y \notin P(\{\det DP \neq 0\})$, then y has a bounded number of preimages. For such a point y, define

$$Z_y := \{z \in \mathbb{C}^n : P(z) - y = 0\}.$$

By the Inverse Function Theorem and our hypothesis on y, real points $x \in Z_y \cap \mathbb{R}^n$ are isolated (complex) points of Z_y. By Theorem 11.6 and the fact that dimension zero irreducible varieties are singletons, Z_y contains a bounded number of isolated points. $\square$

Lemma 11.8. *Let T denote the tube*

$$T := \{x = (x', x_n) \in \mathbb{R}^n : |x'| < 1\}.$$

Let $P : \mathbb{R}^n \to \mathbb{R}^n$ be a polynomial of degree at most N and assume that $\det DP$ is nonvanishing on $\overline{T}$. If $\gamma : \mathbb{R} \to \mathbb{R}^n$ is a polynomial of degree at most N, then $\gamma^{-1}[P(T)]$ is a union of a bounded number of intervals.

Proof. Consider the complex varieties

$$C := \{v \in \mathbb{C}^n : \sum_{i=1}^{n-1} v_i^2 = 1\}$$

$$Z := \{(u, v) \in \mathbb{C}^{1+n} : \gamma(u) = P(v), \, v \in C\}.$$

Suppose that $(t, x) \in Z \cap \mathbb{R}^n$ is a regular point of some subvariety $Z' \subseteq Z$, with $\dim Z' > 0$. If $\det DP(x) \neq 0$, then by the implicit function theorem, Z' can have complex dimension at most one, and, moreover, if the dimension of Z' is one, then there exists a complex neighborhood U of t such that $\gamma(U) \subseteq P(C)$. Shrinking U if necessary, and again using $\det DP(x) \neq 0$, $\gamma(U \cap \mathbb{R}) \subseteq P(C \cap \mathbb{R}^n) = P(\partial T)$. Thus a boundary point of $\gamma^{-1}[P(T)]$ must be a regular point of a dimension zero subvariety $Z' \subseteq Z$, so by Theorem 11.6, the number of boundary points is bounded. $\square$

Lemma 11.9. *Let $P : \mathbb{R}^n \to \mathbb{R}^n$, $\gamma : \mathbb{R} \to \mathbb{R}^n$, and $Q : \mathbb{R}^n \to \mathbb{R}^n$ be polynomials of degree at most N, and assume that P^{-1} is defined and differentiable on a neighborhood of the image $\gamma(I)$, for some open interval I. Then no coordinate of the vector $[(DP^{-1}) \circ \gamma](Q \circ \gamma)$ can change sign more than a bounded number of times on I.*

Proof. Multiplying the vector $[(DP^{-1}) \circ \gamma](Q \circ \gamma)$ by $(\det DP)^2$ and using Cramer's rule, it is enough to prove that if $R : \mathbb{R}^n \to \mathbb{R}^n$ is a polynomial of bounded degree, then

$$(R \circ P^{-1} \circ \gamma) \cdot (Q \circ \gamma) \tag{11.13}$$

changes sign a bounded number of times on I.

We consider the complex variety

$$Z := \{(u, v, w) \in \mathbb{C} \times \mathbb{C}^n \times \mathbb{C}^n : \gamma(u) = P(v) = w, \, R(v) \cdot Q(w) = 0\}.$$

If (11.13) vanishes at $t \in I$, then $(t, P^{-1}(\gamma(t)), \gamma(t)) =: (t, x, y) \in Z$ and $\det DP(x) \neq 0$.

Let $Z' \subseteq Z$ denote an irreducible subvariety from the decomposition (11.12) for which (t, x, y) is a regular point. By the Implicit Function Theorem and $\det DP(x) \neq 0$, either Z' has dimension zero, or Z' has (complex) dimension one and (11.13) vanishes on a (complex) neighborhood of t. Only a bounded number of points can lie on dimension zero subvarieties, and if (11.13) vanishes on a neighborhood of t, then it vanishes on all of I by analyticity. Either way, the number of sign changes is bounded. $\square$

References

[1] J. Bennett, A. Carbery, J. Wright, *A non-linear generalisation of the Loomis–Whitney inequality and applications.* Math. Res. Lett. 12 (2005), no. 4, 443-457.

[2] C. Biswas, *Extremals for a certain convolution operator.* Preprint.

[3] M. Christ, *Convolution, curvature, and combinatorics: a case study.* Internat. Math. Res. Notices 1998, no. 19, 1033-1048.

[4] M. Christ, *Quasiextremals for a Radon-like transform.* arXiv:1106.0722.

[5] M. Christ, *Lebesgue space bounds for one-dimensional generalized Radon transforms.* Unpublished manuscript.

[6] M. Christ, A. Nagel, E. M. Stein, S. Wainger, *Singular and maximal Radon transforms: analysis and geometry.* Ann. of Math. (2) 150 (1999), no. 2, 489–577.

[7] L. J. Corwin, F. P. Greenleaf, *Representations of nilpotent Lie groups and their applications. Part I. Basic theory and examples.* Cambridge Studies in Advanced Mathematics, 18. Cambridge University Press, Cambridge, 1990.

[8] S. Dendrinos, N. Laghi, J. Wright, *Universal L^p improving for averages along polynomial curves in low dimensions.* J. Funct. Anal. 257 (2009), no. 5, 1355-1378.

[9] S. Dendrinos, J. Wright, *Fourier restriction to polynomial curves I: a geometric inequality.* Amer. J. Math. 132 (2010), no. 4, 1031-1076.

[10] S. Dendrinos, B. Stovall, *Uniform estimates for the X-ray transform restricted to polynomial curves.* J. Funct. Anal. 262 (2012), no. 12, 4986-5020.

[11] S. Dendrinos, B. Stovall, *Uniform bounds for convolution and restricted X-ray transforms along degenerate curves.* J. Funct. Anal. 268 (2015), no. 3, 585-633.

[12] W. Fulton, *Intersection theory.* Second ed. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics], 2. Springer-Verlag, Berlin, 1998.

[13] P. T. Gressman, *L^p-improving properties of averages on polynomial curves and related integral estimates.* Math. Res. Lett. 16 (2009), no. 6, 971-989.

[14] P. T. Gressman, *Uniform sublevel Radon-like inequalities.* J. Geom. Anal. 23 (2013), no. 2, 611-652.

[15] N. Laghi, *A note on restricted X-ray transforms.* Math. Proc. Cambridge Philos. Soc. 146 (2009), no. 3, 719-729.

[16] W. Littman, *$L^p - L^q$-estimates for singular integral operators arising from hyperbolic equations. Partial differential equations*, (Proc. Sympos. Pure Math., Vol. XXIII, Univ. California, Berkeley, Calif., 1971), Amer. Math. Soc., Providence, R.I., (1973) 479–481.

[17] A. Nagel, E. M. Stein, S. Wainger, *Balls and metrics defined by vector fields. I. Basic properties.* Acta Math. 155 (1985), no. 1-2, 103–147.

[18] D. M. Oberlin, *Convolution with measures on flat curves in low dimensions*, J. Funct. Anal. **259** (2010), no. 7, 1799–1815.

[19] P. Sjölin, *Fourier multipliers and estimates of the Fourier transform of measures carried by smooth curves in $\mathbb{R}^2$.* Studia Math. **51** (1974), 169–182.

[20] E. M. Stein, *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals.* With the assistance of Timothy S. Murphy. Princeton Mathematical Series, 43. Monographs in Harmonic Analysis, III. Princeton University Press, Princeton, NJ, 1993.

[21] B. Stovall, *Endpoint bounds for a generalized Radon transform.* J. Lond. Math. Soc. (2) 80 (2009), no. 2, 357-374.

[22] B. Stovall, *Quasi-extremals for convolution with surface measure on the sphere*, Illinois J. Math. **53** (2009), no. 2, 391–412.

[23] B. Stovall, *Endpoint $L^p \rightarrow L^q$ bounds for integration along certain polynomial curves.* J. Funct. Anal. 259 (2010), no. 12, 3205-3229.

[24] B. Stovall, *L^p improving multilinear Radon-like transforms.* Rev. Mat. Iberoam. 27 (2011), no. 3, 1059-1085.

[25] B. Stovall, *Uniform L^p-improving for weighted averages on curves.* Anal. PDE 7 (2014), no. 5, 1109-1136.

[26] B. Street, *Multi-parameter Carnot-Carathéodory balls and the theorem of Frobenius.* Rev. Mat. Iberoam. 27 (2011), no. 2, 645–732.

[27] T. Tao, J. Wright, *L^p improving bounds for averages along curves.* J. Amer. Math. Soc. 16 (2003), no. 3, 605-638.

Department of Mathematics, University of California, Berkeley
Email address: mchrist@berkeley.edu

School of Mathematical Sciences, University College Cork, Cork, Ireland
Email address: sd@ucc.ie

Department of Mathematics, University of Wisconsin–Madison
Email address: stovall@math.wisc.edu

Department of Mathematics, University of Wisconsin–Madison
Email address: street@math.wisc.edu