THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Improved Fourier restriction estimates in higher dimensions

OPEN ACCESS

IMPROVED FOURIER RESTRICTION ESTIMATES
IN HIGHER DIMENSIONS

JONATHAN HICKMAN AND KEITH M. ROGERS

ABSTRACT. We consider Guth's approach to the Fourier restriction problem via polynomial partitioning. By writing out his induction argument as a recursive algorithm and introducing new geometric information, known as the polynomial Wolff axioms, we obtain improved bounds for the restriction conjecture, particularly in high dimensions. Consequences for the Kakeya conjecture are also considered.

1. INTRODUCTION

We consider the Fourier transform defined, initially on integrable functions, by

$$\widehat{f}(\xi) := \int_{\mathbb{R}^n} f(x)\, e^{-i\langle x, \xi\rangle}\mathrm{d}x.$$

Letting σ denote the surface measure on a truncated piece of the paraboloid, Stein's restriction conjecture [31] asserts that the *a priori* estimate

$$\|\widehat{f}\,\|_{L^{p'}(d\sigma)} \leqslant C_{n,p}\|f\|_{L^{p'}(\mathbb{R}^n)} \tag{R_p}$$

holds for all $p > \frac{2n}{n-1}$, where $1/p + 1/p' = 1$. This was proved by Fefferman and Stein in two dimensions [10], but remains open in higher dimensions despite extensive study; see, for example, [1, 2, 6, 8, 9, 11, 12, 24, 28, 33, 35, 36, 37, 38, 39, 40, 44] and the references therein.

The strongest partial results are based on the polynomial partitioning method, introduced to the problem by Guth [11, 12]. In this article further progress is obtained by augmenting the method with additional geometric inequalities recently established in work of Katz and the second author [18].

Our results are most easily compared with the previous literature in the high dimensional context. If the restriction conjecture were true, (R_p) would hold for

$$p > 2 + 2n^{-1} + O(n^{-2}),$$

and so we consider $\lambda \geqslant 2$ for which we can confirm that (R_p) holds in the range

$$p > 2 + \lambda n^{-1} + O(n^{-2}). \tag{1}$$

A consequence of the work of Tomas [39] is that λ can be taken to be 4. Although many refinements were made since (including the work of Tao [35] which removed the $O(n^{-2})$-term with $\lambda = 4$), the linear coefficient was not improved for some thirty-five years when Bourgain and Guth [6] showed that it can be lowered to 3. Most recently, Guth [12] proved that λ can be taken to be 8/3. We improve this as follows:

Theorem 1.1. (R_p) *holds in the range* (1) *with* $\lambda = 4/(5 - 2\sqrt{3})$.

We also obtain concrete improvements on the range of exponents for (R_p) in all dimensions $n \geqslant 3$ except $n = 3, 6, 8, 10$ or 12. In these exceptional cases the current best results are due to Wang [40] when $n = 3$ and Guth [12] when $n = 6, 8, 10$ or 12. The current state-of-the-art for the restriction problem in various low dimensions is tabulated below in Figure 1.

The proof of Theorem 1.1 relies on geometric information coming from a recent result in [18]. This geometric information, which we will refer to as the *polynomial Wolff axioms*, bounds the number of direction-separated line segments that can be contained in the neighbourhood of a real algebraic variety.

The present analysis extends that previously performed by Guth [11] in $\mathbb{R}^3$, who proved and applied the polynomial Wolff axiom for a two dimensional variety. Guth's induction argument [11] can be combined with later developments from [12] and thereby directly extended to higher dimensions, using a single application of the $(n-1)$-dimensional polynomial Wolff axioms (see, for example, [9] or [18]), however this yields weaker results than those obtained here. We will take advantage of the polynomial Wolff axioms more often.

By combining the arguments of this article with results from [13], one may also establish a version of Theorem 1.1 for general positively-curved surfaces, including the unit sphere. It is also possible that the methods could be applied to study other oscillatory integral operators, such as those arising in the study of Bochner–Riesz multipliers, but this has not been fully explored. Finally, by a standard argument relating the restriction and Kakeya conjectures, Theorem 1.1 implies estimates for the Kakeya maximal function. This bound is new with $n = 9$, however it does not improve the dimension estimate for Kakeya sets due to Katz–Tao [20]. Perhaps of more interest is the fact that these estimates provide an asymptotic improvement over the classical Wolff bound [42] via a very different approach to that used in [20].

The article is organised as follows:

- A number of reductions are performed in the sequel. Following [6, 11, 12], the problem is reduced to establishing the so-called *k-broad estimates* for the extension operator.
- After setting up some notational conventions in Section 3, a sketch of the proof of the main theorem is provided in Section 4.
- In Sections 5-7, the basics of broad norms, polynomial partitioning and the wave packet decomposition are recalled.
- In Section 8, we show how the polynomial Wolff axiom theorem can be used to improve certain estimates for averaged norms at different scales.
- In Section 9, Guth's polynomial partitioning argument from [11, 12] is reformulated as a recursive algorithm.
- In Section 10, the new estimates are combined with the recursive algorithm to improve the range of estimates for the restriction problem.
- The final section contains a discussion of restriction to other hypersurfaces, some remarks on the numerology, and possible directions in which the argument could be strengthened. Finally, the application to the Kakeya problem is described.

2. Reduction to k-broad estimates

Restriction estimates are typically proven via duality, with the adjoint operator E defined by

$$Eg(x) := \int_{|\xi| \leqslant 1} g(\xi)\, e^{i(x_1\xi_1 + \ldots + x_{n-1}\xi_{n-1} + x_n|\xi|^2)} \mathrm{d}\xi.$$

Noting that now $\xi \in \mathbb{R}^{n-1}$ (and $x \in \mathbb{R}^n$ as before), this is often referred to as the *extension operator*. It follows that the estimate (R_p) for a given value of p is equivalent to the inequality

$$\|Eg\|_{L^p(\mathbb{R}^n)} \leqslant C_{n,p}\|g\|_{L^p(\mathbb{R}^{n-1})}.$$

Moreover, by a now standard *ε-removal argument* (see [34]) and factorisation theory (see [2] or [7, Lemma 1]), this holds for all p in an open range if and only if for all $\varepsilon > 0$ and all $R \gg 1$ the local estimates

$$\|Eg\|_{L^p(B_R)} \leqslant C_{n,p,\varepsilon}R^\varepsilon\|g\|_{L^\infty(\mathbb{R}^{n-1})} \tag{R_p^*}$$

hold in the same range. Here B_R denotes an arbitrary ball of radius R in $\mathbb{R}^n$.

Rather than attempt to prove (R_p^*) directly, it is useful to work with a class of weaker inequalities known as *k-broad estimates*. These inequalities were introduced by Guth [11, 12] and were inspired by the earlier multilinear restriction theory developed in [1]. The k-broad estimates take the form

$$\|Eg\|_{\mathrm{BL}_k^p(B_R)} \leqslant C_{n,p,\varepsilon}R^\varepsilon\|g\|_{L^\infty(\mathbb{R}^{n-1})}, \tag{BL_k^p}$$

where the expression on the left-hand side is known as a *k-broad norm*. The precise definition of the k-broad norm is a little complicated and is deferred until Section 5. We remark, however, that the key advantage of working with $\|Eg\|_{\mathrm{BL}_k^p(B_R)}$ rather than $\|Eg\|_{L^p(B_R)}$ is that the former expression is very small whenever the mass of Eg is concentrated near a $(k-1)$-dimensional set (see Lemma 8.2 below for a precise statement of this property).

The main result of this article is the following theorem.

Theorem 2.1. *Let $2 \leqslant k \leqslant n-1$ and*

$$p \geqslant p_n(k) := 2 + \frac{8(2n-1)}{n(5n+2k-9)+k(k-3)+4}. \tag{2}$$

Then (BL_k^p) *holds for all $\varepsilon > 0$ and $R \gg 1$.*

When $n = 3$ and $k = 2$ this corresponds to the main result from [11] and stronger estimates are now known in this case [40]. In all other dimensions $n \geqslant 4$, Theorem 2.1 offers an improvement over what was previously known. When $n = 4$ and $k = 3$, the range (2) extends that given by [9, Theorem 3.2].[1] When $n \geqslant 5$, Theorem 2.1 strengthens a (corollary of a) theorem of Guth [12] which showed that the inequality (BL_k^p) holds whenever $p \geqslant 2 + \frac{4}{n+k-2}$;[2] observe that the range (2) in Theorem 2.1 is strictly larger than this.

Unfortunately, since the k-broad estimates are weaker than the corresponding linear estimates, it is difficult to pass directly from an inequality of the form (BL_k^p) to one of the form (R_p^*). Nevertheless, a mechanism developed by Bourgain and

---

[1]In [9] it is shown that the $n = 4$ and $k = 3$ case of Theorem 2.1 would follow from a strengthened version of the polynomial Wolff axiom theorem from [18] involving a polynomial dependence on the degree. For the purposes of this article, no such explicit dependence on the degree is required, and therefore the 3-broad inequality in $\mathbb{R}^4$ is established in a larger range than that stated in [9].
[2]In [12], strengthened versions of (BL_k^p) are established with L^2 rather than L^∞ norms appearing on the right-hand side, and so our estimates are stronger in one sense and weaker in another.

$n =$	$p >$		$n =$	$p >$	
2	4	Fefferman–Stein [10]	11	$2 + \frac{14}{55}$	Theorem 2.1
3	$3 + \frac{3}{13}$	Wang [40]	12	$2 + \frac{4}{17}$	Guth [12]
4	$2 + \frac{1407}{1759}$	Theorem 2.1[5]	13	$2 + \frac{100}{471}$	Theorem 2.1
5	$2 + \frac{12}{19}$	Theorem 2.1	14	$2 + \frac{108}{541}$	Theorem 2.1
6	$2 + \frac{1}{2}$	Guth [12]	15	$2 + \frac{116}{637}$	Theorem 2.1
7	$2 + \frac{52}{123}$	Theorem 2.1	16	$2 + \frac{62}{359}$	Theorem 2.1
8	$2 + \frac{4}{11}$	Guth [12]	17	$2 + \frac{4}{25}$	Theorem 2.1
9	$2 + \frac{34}{107}$	Theorem 2.1	18	$2 + \frac{7}{46}$	Theorem 2.1
10	$2 + \frac{2}{7}$	Guth [12]	19	$2 + \frac{1}{7}$	Theorem 2.1

FIGURE 1. The current state-of-the-art for the restriction problem in low dimensions. New results are highlighted and in most cases are deduced by combining Theorem 2.1 with Proposition 2.2.[5]

Guth [6] allows this passage under certain constraints on the exponent p. In particular, the following proposition is a consequence of the method developed in [6], as observed in [12, Proposition 9.1].

Proposition 2.2 (Bourgain–Guth [6], Guth [12]). *Let $n \geqslant 3$ and*

$$2 + \frac{4}{2n - k} \leqslant p \leqslant 2 + \frac{2}{k - 2}.$$

Then (BL_k^p) *implies* (R_p^*).

The original method of Bourgain–Guth [6] was developed to convert certain multilinear inequalities of Bennett–Carbery–Tao [1] into linear estimates. It was later observed by Guth [12] that the method of [6] does not require the full strength of the k-linear theory, but may instead take k-broad estimates as its input (which appear to be somewhat easier to prove[3]).

Theorem 1.1 follows as a direct consequence of Theorem 2.1 and Proposition 2.2. When applying Proposition 2.2, the upper bound on p is unimportant. However the lower bound,

$$p \geqslant 2 + \frac{4}{2n - k}, \tag{3}$$

is a limiting factor in the arguments, along with the condition (2) on the exponents in the k-broad inequality. In order to improve the state-of-the-art for the restriction conjecture one must choose an optimal k so that neither of these two conditions is overly restrictive. For instance, if $n = 5$ and $k = 3$, then

$$2 + \frac{4}{10 - 3} \leqslant p_5(3) = 2 + \frac{12}{19};$$

---

[3]See [13, Section 6.2] for a detailed discussion of the relationship between k-broad and k-linear inequalities.

[4]This is deduced by combining Theorem 2.1 with a more sophisticated version of Proposition 2.2: see Remark 11.1 below.

[5]These computations were carried out using the following Maple [23] code:

```
n := [insert dimension];
p_broad := 2+8*(2n-1)/(n*(5n+2k-9)+k*(k-3) +4):   p_limit :=2+ 4/(2*n-k):
p_seq := [seq(max(eval(p_broad, k = i), eval(p_limit, k = i)), i = 2 .. n)]:
new_exponent := min(p_seq);
```

Theorem 2.1 and Proposition 2.2 therefore imply that the restriction inequality holds for $p > 2 + \frac{12}{19}$ when $n = 5$. Other low dimensional cases can also be analysed directly and some examples can be found in Figure 1 above.

In high dimensions, to derive the λ coefficient featured in Theorem 1.1, we write $k = \nu n + O(1)$ for some $0 < \nu < 1$, so that, asymptotically,

$$p_n(k) = 2 + \frac{16}{5 + 2\nu + \nu^2} n^{-1} + O(n^{-2}). \tag{4}$$

On the other hand, with $k = \nu n + O(1)$, the condition (3) can be rewritten as

$$p \geqslant 2 + \frac{4}{2 - \nu} n^{-1} + O(n^{-2}). \tag{5}$$

The linear coefficients in (4) and (5) are then equal when ν is the positive solution of the quadratic equation

$$x^2 + 6x - 3 = 0.$$

Plugging this solution back into (5) yields (R_p) in the range

$$p > 2 + \lambda n^{-1} + O(n^{-2})$$

with $\lambda = \frac{4}{2-\nu} = \frac{4}{5 - 2\sqrt{3}}$.

It remains to prove Theorem 2.1, which will be the focus of the remainder of the article.

3. Notational conventions

From now on, we work with smooth, bounded functions f, g or h that map from the unit ball B^{n-1} of $\mathbb{R}^{n-1}$ to the complex numbers, and we sometimes write $x \in \mathbb{R}^n$ as $x = (x', x_n)$ where $x' \in \mathbb{R}^{n-1}$. We call an n-dimensional ball B_r of radius r an r-*ball* and an $(n-1)$-dimensional ball θ of radius $r^{-1/2}$ an $r^{-1/2}$-*cap*. We call a cylinder of length r and radius $r^{1/2}$ an r-*tube*. The δ-neighbourhood of a set E will be denoted by $N_\delta E$.

The arguments will involve the *admissible parameters* n, p and ε and the constants in the estimates will be allowed to depend on these quantities. Given positive numbers $A, B \geqslant 0$ and a list of objects L, the notation $A \lesssim_L B$, $B \gtrsim_L A$ or $A = O_L(B)$ signifies that $A \leqslant C_L B$ where C_L is a constant which depends only on the objects in the list and the admissible parameters. We write $A \sim_L B$ when both $A \lesssim_L B$ and $B \lesssim_L A$. We will also write $A \ll B$ or $B \gg A$ to denote that $A \leqslant C^{-1} B$ for some choice of $C \geqslant 1$ which can be taken to be as large as desired provided it is admissible.

The cardinality of a finite set A is denoted by $\#A$. A set A' is said to be a *refinement* of A if $A' \subseteq A$ and $\#A' \gtrsim \#A$. In many cases it will be convenient to *pass to a refinement* of a set A, by which we mean that the original set A is replaced with some refinement.

4. Overview

4.1. **The polynomial Wolff axioms.** The key new geometric ingredient is the following theorem, which amounts to a confirmation of the Kakeya conjecture in a very specialised 'algebraic' situation. It follows by combining [18, Theorem 1.1] with Wongkew's lemma [45], the latter of which bounds the measure of a neighbourhood of a real algebraic variety over a ball.

Theorem 4.1 (Polynomial Wolff axioms [18]). *Let $\delta > 0$ and $c, r \geqslant 1$. Let $\mathbf{Z} \subseteq \mathbb{R}^n$ denote an m-dimensional algebraic variety and let $\mathbf{T}$ denote a collection of r-tubes*

contained in a ball of radius $2r$. *If the central axes of the tubes point in* $r^{-1/2}$-*separated directions, then*

$$\#\Big\{ T \in \mathbf{T} \,:\, T \subseteq N_{cr^{1/2}}\mathbf{Z} \Big\} \lesssim_{\overline{\deg}\,\mathbf{Z},\delta} c^{n-m} r^{\frac{m-1}{2}+\delta}.$$

This theorem was proven for $n = 3$ by Guth [11] who later conjectured the general statement in [12] (see also [15]). The $n = 4$ case was solved by Zahl before a proof in general dimensions was given in [18]. Theorem 4.1 is referred to as the *polynomial Wolff axioms* since the result can be interpreted as a verification that families of direction-separated tubes satisfy a natural polynomial generalisation of the classical (linear) *Wolff axiom* introduced in [42] (see [15] for further details).

4.2. **A brief description of the proof.** The proof of Theorem 2.1 extends an argument of Guth [11] in $\mathbb{R}^3$, by combining it with the later developments in higher dimensions from [12]. Both the articles [11] and [12] give comprehensive and highly readable introductory overviews of the core arguments; readers unfamiliar with these topics are encouraged to consult these sources for a detailed description of the main ideas. In high dimensions some complications arise which are not present in $\mathbb{R}^3$. For this reason, the proof given in Sections 5 - 10 is structured somewhat differently from the proofs presented in [11, 12]. These differences are highlighted and explained in the following subsection.

The key ingredients of the proof of Theorem 2.1 are as follows:

Wave packet decomposition. The first step of the argument is to employ the standard technique of decomposing the input function f as a sum of localised pieces called *wave packets*. In particular, fixing a large scale $R \gg 1$, one decomposes the domain B^{n-1} as a union of $R^{-1/2}$-balls denoted by θ and referred to as $R^{-1/2}$-*caps*. The function is then written as a sum of pieces $f = \sum_{(\theta,v)} f_{\theta,v}$ where each $f_{\theta,v}$ has support in the cap θ and the inverse Fourier transform of $f_{\theta,v}$ is concentrated in an $R^{1/2}$-ball centred at $v \in \mathbb{R}^{n-1}$. There are two key properties of this decomposition:

- **Orthogonality:** Given any collection of wave packets $\mathbb{W}$ one has

$$\Big\| \sum_{(\theta,v)\in\mathbb{W}} f_{\theta,v} \Big\|_2^2 \sim \sum_{(\theta,v)\in\mathbb{W}} \|f_{\theta,v}\|_2^2. \tag{6}$$

- **Spatial concentration:** On the ball $B(0,R)$, the function $Ef_{\theta,v}$ is essentially supported on an R-tube $T_{\theta,v}$ with direction governed by θ and position governed by v.

More precisely, the direction of $T_{\theta,v}$ is given by the normal direction to the paraboloid at the point $(\xi_\theta, |\xi_\theta|^2)$, where ξ_θ is the centre of θ. Thus, $Ef = \sum_{(\theta,v)} Ef_{\theta,v}$ can be thought of as a sum of oscillating, normalised characteristic functions of tubes, which point in many different directions. Understanding the incidence geometry of these tubes is a key consideration in the restriction problem.

Polynomial partitioning. A useful tool for studying the incidence-geometric problems arising from the wave packet decomposition is the *polynomial partitioning method*. This method was introduced by Guth and Katz [14] in their resolution of the Erdős distance conjecture and was first applied to the restriction problem by Guth in [12] (the latter work also incorporated a refinement to the original partitioning method of [14] due to Solymosi and Tao [30]). The basic idea is a divide-and-conquer-style argument: one begins by finding a polynomial P of low degree which partitions the mass of $\|Ef\|_{\mathrm{BL}_k^p(B_R)}$ into equal size pieces. More precisely, let $Z(P) := \{z \in \mathbb{R}^n : P(z) = 0\}$ denote the zero set of P and cell(P) the set of connected components of $\mathbb{R}^n \setminus Z(P)$. These connected components are referred to

as *cells*. The polynomial P can then be chosen so that the $\|Ef\|_{\mathrm{BL}_k^p(O')}$ are (essentially) equal as O' varies over cell(P). Due to geometric (and underlying uncertainty principle) considerations, one actually works with a 'blurred out' version of the variety $Z(P)$ given by the $R^{1/2}$-neighbourhood $W := N_{R^{1/2}} Z(P)$ and referred to as the *wall*. Defining the collection of slightly shrunken cells $\mathcal{O} := \{O' \setminus W : O' \in \mathrm{cell}(P)\}$, the following simple, yet vital, geometric property holds:

Whenever $T_{\theta,v}$ enters a shrunken cell $O = O' \setminus W$, the core line necessarily enters the original cell O'. $\qquad$ (7)

Unlike the original cells, collectively the $O \in \mathcal{O}$ may only account for a small proportion of the mass of $\|Ef\|_{\mathrm{BL}_k^p(B_R)}$. There are two cases to consider:

Cellular case: The mass of $\|Ef\|_{\mathrm{BL}_k^p(B_R)}$ concentrates on the $O \in \mathcal{O}$ in the sense that

$$\|Ef\|_{\mathrm{BL}_k^p(B_R)}^p \lesssim \sum_{O \in \mathcal{O}} \|Ef\|_{\mathrm{BL}_k^p(O)}^p.$$

In this situation, one defines $f_O := \sum_{(\theta,v) \in \mathbb{T}_O} f_{\theta,v}$ where $\mathbb{T}_O$ denotes the collection of wave packets (θ, v) for which $T_{\theta,v} \cap O \neq \emptyset$. By the spatial concentration property of the wave packets

$$\|Ef\|_{\mathrm{BL}_k^p(B_R)}^p \lesssim \sum_{O \in \mathcal{O}} \|Ef_O\|_{\mathrm{BL}_k^p(O)}^p.$$

The key observation here is that each (θ, v) can only belong to a small number (in particular, $\deg P + 1$) of the sets $\mathbb{T}_O$. This is due to (7) and the fact that, by the fundamental theorem of algebra (or Bézout's theorem), the core line of a tube $T_{\theta,v}$ can only enter $\deg P + 1$ cells from cell(P). This observation can be interpreted as saying the sets $\mathbb{T}_O$ are 'almost disjoint' which implies, via (6), that the f_O are 'almost orthogonal'. Consequently, one can pass to the cells and analyse them individually. This forms the basis of a recursive procedure.[6]

Algebraic case: The mass of $\|Ef\|_{\mathrm{BL}_k^p(B_R)}$ concentrates on the wall in the sense that

$$\|Ef\|_{\mathrm{BL}_k^p(B_R)}^p \lesssim \|Ef\|_{\mathrm{BL}_k^p(W)}^p.$$

Here it suffices to consider only those wave packets (θ, v) for which $T_{\theta,v} \cap W \neq \emptyset$. A tube $T_{\theta,v}$ can intersect W in one of two ways: either *tangentially* or *transversally*. The analysis is further divided into two subcases depending on whether the main contribution to $\|Ef\|_{\mathrm{BL}_k^p(B_R)}^p$ arises from tangential or transverse wave packets.

In the transversal subcase the tubes can be thought of as passing directly through the wall. This situation can be treated in a manner similar to the cellular case, this time using a continuum version of Bézout's theorem to show that any given tube can intersect W transversally in relatively few places.

It remains to study the tangential subcase. Here the $T_{\theta,v}$ can be thought of as being contained in W and making a small angle with tangent spaces at nearby points of the variety $Z(P)$.

Dimensional reduction. The polynomial partitioning argument sketched above can be interpreted as a dimensional reduction. If either the cellular or the transverse algebraic case holds, then one can obtain acceptable estimates for $\|Ef\|_{\mathrm{BL}_k^p(B_R)}$. Thus, it suffices to consider the situation where the wave packets of f are all tangent

---

[6]This part of the argument is fairly delicate and the almost orthogonal property needs to be precisely quantified in terms of $\deg P$. For the purposes of this sketch, the full details are omitted.

to some variety of dimension $n - 1$. By iterating this dimensional reduction procedure,[7] it becomes important to understand what can be said when the wave packets of f are all tangent to some variety of dimension m for any value $0 \leqslant m \leqslant n - 1$.

Key estimates in the tangential case. The reduction to tangential situations, as outlined above, can be exploited in a number of ways:

- **Vanishing property of the k-broad norms:** The definition of the k-broad norms implies that if the wave packets of f are all tangential to a variety of dimension $m < k$, then $\|Ef\|_{\mathrm{BL}_k^p(B_R)}$ essentially vanishes. Thus, one need only consider tangency properties with respect to varieties of dimension at least k. Using this fact alone, one may prove k-broad estimates in the range $p > \frac{2k}{k-1}$ corresponding to the Bennett–Carbery–Tao multilinear restriction theorem [1].[8]
- **Transverse equidistribution estimates:** These inequalities were introduced by Guth [12] and heavily exploit the curvature properties of the paraboloid, allowing for k-broad estimates beyond the $p > \frac{2k}{k-1}$ range. The basic idea behind the transverse equidistribution estimates is recalled below in Section 8.3.
- **The polynomial Wolff axioms:** Given a family $\mathbf{T}$ of R-tubes lying in the $R^{1/2}$-neighbourhood of a variety, the polynomial Wolff axioms limit the number of different directions in which the $T \in \mathbf{T}$ can lie. Thus, if the wave packets of f are all tangent to some low dimensional variety, then f must be supported on very few caps θ (since the caps θ correspond to the directions of the tubes $T_{\theta,v}$). The small support of f can be exploited via Hölder's inequality to obtain favourable k-broad estimates.

4.3. **Induction versus recursion.** When applying the polynomial Wolff axioms to the restriction problem in high dimensions, a number of complications arise which are not present in the $\mathbb{R}^3$ case treated in [11]. The root of these complications lies in the fact that, in contrast with $\mathbb{R}^3$ where one only need consider tangency conditions with respect to 2-surfaces, in higher dimensions one must consider tangency conditions with respect to surfaces of many different dimensions.

The core argument sketched in the previous subsection can be implemented as either an induction or a recursion argument. The original articles [11] and [12] make heavy use of mathematical induction (inducting on a number of quantities including the choice of scale R); this has the advantage of yielding a clean and concise argument, but unfortunately useful structural properties are potentially hidden. From the perspective of a recursive algorithm one may gain a more detailed understanding of the argument at each stage of the iterative process; this is the approach taken in the present article.

There is certainly a precedent for the recursive approach: for instance, in the fourth section of [6], Bourgain and Guth reformulate their key induction-on-scale argument as a recursive procedure to allow for the use of additional information coming from X-ray transform estimates (see also [22, 38] for an elaboration of this argument). Similarly, in a recent article of Wang [40], the induction-on-scale procedure of [11] was rewritten as a recursion; this permitted a more detailed

---

[7]A number of serious complications arise in implementing this iteration scheme and, in particular, in dealing with the transverse algebraic case. This part of the argument requires what are known as *transverse equidistribution estimates*: these inequalities are briefly mentioned below, see the introductory discussion in [12] for further details.

[8]Indeed, this follows by applying the argument of [12] but ignoring gains coming from *transverse equidistribution* (see the following bulletpoint).

analysis of the underlying geometry of the extension operator and led to the current best known bounds for the restriction conjecture in $\mathbb{R}^3$.

When written in the form of a recursive algorithm, the polynomial partitioning argument of [12] can be interpreted as a structural statement. Following the discussion in the previous subsection, one may think of the input function f as being broken into many different pieces where, roughly, each piece is made up of wave packets tangential to a low dimensional variety at some scale (there may be other pieces which do not have this property, but they arise from the cellular or transverse algebraic cases and satisfy favourable estimates). Thus, the structural statement allows one to focus on estimating the 'tangential' pieces $\{f_{\text{tang}}\}$ of the function. This 'tangential reduction' is then exploited via the key estimates described in the previous subsection.

In high dimensional cases, however, the f_{tang} tend to enjoy further structural properties which one could potentially utilise in order to improve the range of estimates guaranteed by Theorem 2.1. Indeed, typically a given f_{tang} is not only tangent to a single variety $\mathbf{Z}$ at a single scale r, but it satisfies certain tangency conditions with respect to a whole sequence of scales $r_m < \cdots < r_n$ and a corresponding sequence of varieties $\mathbf{Z}_m, \ldots, \mathbf{Z}_n$ with $\dim \mathbf{Z}_i = i$ for $m \leqslant i \leqslant n$. These 'nested' conditions could potentially lead to further gains for the restriction exponent. To carry out such a programme, however, one would have to effectively analyse properties of the f_{tang} across many distinct scales $r_m < \cdots < r_n$; this situation lends itself more naturally to a recursive algorithm, rather than an inductive argument.

5. Broad norms

Here we recall the definition and basic properties of the k-broad norms from [11] and [12]. For a detailed motivation of this definition and its relation to the multilinear restriction theory of [1] the reader is referred to [12] and [13, Section 6.2].

Fix some large $R \gg 1$ and a ball $B_R \subset \mathbb{R}^n$. Decompose the unit ball B^{n-1} into finitely-overlapping balls τ of radius K^{-1}, where K is a large constant satisfying $1 \ll K \ll R$. These $(n-1)$-dimensional balls are referred to as K^{-1}-caps. Given a function f, supported on B^{n-1}, we write $f = \sum_\tau f_\tau$ where $f_\tau := f\psi_\tau$ for $(\psi_\tau)_\tau$ a partition of unity subordinate to the caps τ. Let $G \colon B^{n-1} \to S^{n-1}$ denote the Gauss map associated to the paraboloid, given explicitly by

$$G(\xi) := \frac{1}{(1 + 4|\xi|^2)^{1/2}} \left(-2\xi, 1 \right). \tag{8}$$

Given a pair of non-zero vectors $v, v' \in \mathbb{R}^n$, let $\angle(v, v')$ denote the (unsigned) angle between them. If $V \subseteq \mathbb{R}^n$ is a linear subspace, then let $\angle(G(\tau), V)$ denote the minimum of $\angle(v, v')$ over all pairs of non-zero vectors $v \in V$ and $v' \in G(\tau)$.

The spatial ball B_R is also decomposed into relatively small balls B_{K^2} of radius K^2. In particular, fix $\mathcal{B}_{K^2}$ a collection of finitely-overlapping K^2-balls which are centred in and cover B_R. Then, for $B_{K^2} \in \mathcal{B}_{K^2}$, define

$$\mu_{Ef}(B_{K^2}) := \min_{V_1, \ldots, V_A \in \mathrm{Gr}(k-1, n)} \left(\max_{\tau : \angle(G(\tau), V_a) > K^{-1} \text{ for } 1 \leqslant a \leqslant A} \|Ef_\tau\|^p_{L^p(B_{K^2})} \right); \tag{9}$$

here $\mathrm{Gr}(k-1, n)$ is the Grassmannian manifold of all $(k-1)$-dimensional subspaces in $\mathbb{R}^n$. For $U \subseteq \mathbb{R}^n$ the k-broad norm over U can then be defined as

$$\|Ef\|_{\mathrm{BL}^p_{k,A}(U)} := \left(\sum_{B_{K^2} \in \mathcal{B}_{K^2}} \frac{|B_{K^2} \cap U|}{|B_{K^2}|} \mu_{Ef}(B_{K^2}) \right)^{1/p}. \tag{10}$$

With this definition, the inequality (BL_k^p) from Section 2 is understood to hold for $\|Ef\|_{\mathrm{BL}_k^p(U)} := \|Ef\|_{\mathrm{BL}_{k,A}^p(U)}$ for some choice of $A \sim 1$.

Before continuing it is perhaps useful to clarify the relative sizes of the parameters. Given any p and ε, when proving a broad norm estimate (BL_k^p) it is always assumed that K and A are large but admissible (that is, they depend only on n, p and ε). The parameter K must be chosen large in order for Proposition 2.2 to hold (see [12] and [6]) whilst the parameter A must be chosen large in order to facilitate multiple applications of Lemma 5.2 and Lemma 5.3, as described below. Nevertheless, it is always possible to make admissible choices of K and A. The parameter R, on the other hand, is an arbitrarily large number which will be, in general, non-admissible.

As mentioned in Section 2, the key advantage of working with k-broad norms rather than the classical L^p-norms is that, roughly, they vanish whenever the mass of Ef is concentrated around a set of dimension less than k. This property is fundamental to the proof of Theorem 2.1, but to make it precise requires a number of preliminary definitions and therefore the details are postponed until Lemma 8.2 below.

5.1. **Basic properties.** It is easy to see that $\|Ef\|_{\mathrm{BL}_{k,A}^p(U)}$ is not a norm in any traditional sense. Nevertheless, as noted in [12], it does satisfy weak variants of certain key properties of L^p-norms.

Lemma 5.1 (Finite subadditivity). *Let $U_1, U_2 \subseteq \mathbb{R}^n$, $1 \leqslant p < \infty$, and $A \in \mathbb{N}$. Then*

$$\|Ef\|_{\mathrm{BL}_{k,A}^p(U_1 \cup U_2)}^p \leqslant \|Ef\|_{\mathrm{BL}_{k,A}^p(U_1)}^p + \|Ef\|_{\mathrm{BL}_{k,A}^p(U_2)}^p$$

holds for all integrable $f \colon B^{n-1} \to \mathbb{C}$.

This is an immediate consequence of the definition of the k-broad norms. A slightly less trivial observation is that $\|Ef\|_{\mathrm{BL}_{k,A}^p(U)}$ also satisfies weak versions of the triangle and logarithmic convexity inequalities.

Lemma 5.2 (Triangle inequality). *Let $U \subseteq \mathbb{R}^n$, $1 \leqslant p < \infty$ and $A \in \mathbb{N}$. Then*

$$\|E(f_1 + f_2)\|_{\mathrm{BL}_{k,2A}^p(U)} \lesssim \|Ef_1\|_{\mathrm{BL}_{k,A}^p(U)} + \|Ef_2\|_{\mathrm{BL}_{k,A}^p(U)}$$

holds for all integrable $f_1, f_2 \colon B^{n-1} \to \mathbb{C}$.

Lemma 5.3 (Logarithmic convexity). *Let $U \subseteq \mathbb{R}^n$, $1 \leqslant p, p_0, p_1 < \infty$ and $A \in \mathbb{N}$. Suppose that $0 \leqslant \alpha \leqslant 1$ satisfies*

$$\frac{1}{p} = \frac{1 - \alpha}{p_0} + \frac{\alpha}{p_1}.$$

Then

$$\|Ef\|_{\mathrm{BL}_{k,2A}^p(U)} \lesssim \|Ef\|_{\mathrm{BL}_{k,A}^{p_0}(U)}^{1-\alpha} \|Ef\|_{\mathrm{BL}_{k,A}^{p_1}(U)}^{\alpha}$$

holds for all integrable $f \colon B^{n-1} \to \mathbb{C}$.

The proofs of these estimates are entirely elementary and can be found in [12]. The parameter A appears in the definition of the k-broad norm to allow for these weak triangle and logarithmic convexity inequalities.

5.2. **Linear versus k-broad estimates.** Any k-broad estimate is weaker than the corresponding linear estimate. For instance, assuming that the local extension estimate (R_p^*) holds, given $\varepsilon > 0$ and $1 \leqslant r \leqslant R$, it follows that

$$\|Ef\|_{\mathrm{BL}_{k,A}^p(B_r)} \leqslant \left(\sum_{\tau: K^{-1}-\mathrm{cap}} \sum_{\substack{B_{K^2} \in \mathcal{B}_{K^2} \\ B_{K^2} \cap B_r \neq \emptyset}} \|Ef_\tau\|_{L^p(B_{K^2})}^p \right)^{1/p} \lesssim K^{O(1)} r^\varepsilon \|f\|_\infty;$$

since K is just a constant (in particular, it is chosen independently of $R \gg 1$), this implies (BL_k^p).

From the preceding observation, L^p estimates for the extension operator translate into k-broad inequalities. In view of this, it is useful to briefly recall some standard L^2 estimates for the extension operator. Plancherel's theorem implies the familiar conservation of energy identity

$$\int_{\mathbb{R}^{n-1}} |Ef(x', x_n)|^2 \, \mathrm{d}x' = (2\pi)^{n-1} \|f\|_2^2 \tag{11}$$

and one may integrate in the x_n variable and take square roots to conclude that

$$\|Ef\|_{L^2(B_r)} \lesssim r^{1/2} \|f\|_2$$

for any r-ball B_r. Arguing as above, one immediately arrives at the k-broad variant

$$\|Ef\|_{\mathrm{BL}_{k,A}^2(B_r)} \lesssim r^{1/2} \|f\|_2, \tag{12}$$

valid for all $r \geqslant 1$.

6. POLYNOMIAL PARTITIONING

6.1. **Basic partitioning.** In this section the relevant algebraic and topological ingredients for the proof of Theorem 2.1 are reviewed. In particular, the key polynomial partitioning theorem is stated, which is adapted from previous works of Guth [11, 12] on the restriction conjecture.

Definition 6.1. Given any collection of polynomials $P_1, \ldots, P_{n-m} : \mathbb{R}^n \to \mathbb{R}$ the common zero set

$$Z(P_1, \ldots, P_{n-m}) := \left\{ x \in \mathbb{R}^n \, : \, P_1(x) = \cdots = P_{n-m}(x) = 0 \right\}$$

will be referred to as a *variety*.[9] Given a variety $\mathbf{Z} = Z(P_1, \ldots, P_{n-m})$, define its *(maximum) degree* to be the number

$$\overline{\deg} \, \mathbf{Z} := \max\{\deg P_1, \ldots, \deg P_{n-m}\}.$$

It will often be convenient to work with varieties which satisfy the additional property that

$$\bigwedge_{j=1}^{n-m} \nabla P_j(z) \neq 0 \qquad \text{for all } z \in \mathbf{Z} = Z(P_1, \ldots, P_{n-m}). \tag{13}$$

In this case the zero set forms a smooth m-dimensional submanifold of $\mathbb{R}^n$ with a (classical) tangent space $T_z \mathbf{Z}$ at every point $z \in \mathbf{Z}$. A variety $\mathbf{Z}$ which satisfies (13) is said to be an *m-dimensional transverse complete intersection*.

Of particular interest is the case of hypersurfaces, where $m = n - 1$. Given a polynomial $P : \mathbb{R}^n \to \mathbb{R}$ consider the collection $\mathrm{cell}(P)$ of connected components of $\mathbb{R}^n \setminus Z(P)$. As in Section 4, each $O \in \mathrm{cell}(P)$ is referred to as a *cell* cut out by the variety $Z(P)$ and the cells are thought of as partitioning the ambient euclidean space into a finite collection of disjoint regions.

---

[9]The ideal generated by the P_j is *not* required to be irreducible.

Theorem 6.2 (Guth [11]). *Fix $d \in \mathbb{N}$ and suppose $F \in L^1(\mathbb{R}^n)$ is non-negative. Then there exists a polynomial $P \colon \mathbb{R}^n \to \mathbb{R}$ of degree at most d such that:*

  i) $\#\mathrm{cell}(P) \sim d^n$;
  ii) The integrals $\int_O F$ for $O \in \mathrm{cell}(P)$ are all equal.

This theorem is based on an earlier discrete partitioning result which played a central role in the resolution of the Erdős distance conjecture [14]. The proof is essentially topological, involving the polynomial ham sandwich theorem of Stone–Tukey [32], which is itself a consequence of the Borsuk–Ulam theorem.

Under the hypotheses of Theorem 6.2, it trivially follows that

$$\int_{\mathbb{R}^n} F = \sum_{O \in \mathrm{cell}(P)} \int_O F = \#\mathrm{cell}(P) \int_{O_*} F \qquad \text{for any } O_* \in \mathrm{cell}(P). \qquad (14)$$

In view of the forthcoming applications of the polynomial partitioning theorem, precise equality is not required in (14), but merely comparability. By relaxing the inequality, one may, for instance, ensure that $Z(P)$ is given by a finite union of transverse complete intersections: see Theorem 5.5 of [12]. Furthermore, often one may freely pass to some refinement of the collection of cells which satisfy additional properties. This observation naturally lends itself to pigeonholing arguments, and two examples along these lines are discussed presently.

Passing to shrunken cells. It will be necessary to work with a 'blurred out' version of the variety $Z(P)$ given by the $r^{1/2+\delta_\circ}$-neighbourhood $N_{r^{1/2+\delta_\circ}} Z(P)$ for different choices of $r > 0$ and small parameter $\delta_\circ > 0$. The set $N_{r^{1/2+\delta_\circ}} Z(P)$ is referred to as the *wall*. A simple pigeonholing argument shows that at least one of two cases hold:

<u>Cellular case</u>. One may pass to a refinement of $\mathrm{cell}(P)$ such that if $\mathcal{O}$ denotes the collection of $r^{1/2+\delta_\circ}$-*shrunken cells*

$$\mathcal{O} := \left\{ O' \setminus N_{r^{1/2+\delta_\circ}} Z(P) \, : \, O' \in \mathrm{cell}(P) \right\}, \qquad (15)$$

then the mass of F is essentially evenly distributed across these shrunken cells:

$$\int_O F \sim d^{-n} \int_{\mathbb{R}^n} F \qquad \text{for all } O \in \mathcal{O}.$$

<u>Algebraic case</u>. The contribution to the integral from the wall dominates:

$$\int_{\mathbb{R}^n} F \lesssim \int_{N_{r^{1/2+\delta_\circ}} Z(P)} F.$$

Controlling the size of the cells. A simple but useful observation, appearing in [40], is that one may also apply a pigeonholing argument to yield some natural control on the size of the cells. Here the analysis is localised to a fixed r-ball B_r and, in particular, it is assumed that $\mathrm{supp}\, F \subset B_r$. In this situation one may, after passing to various refinements and relaxing the equalities in (14), assume that each $O \in \mathrm{cell}(P)$ has diameter at most r/d. In the present article, this reduction is made more for convenience rather than out of necessity and only a bound of $r/2$ is needed on the diameter of the cells; the precise details of this argument are therefore omitted (see [40] for further information).

6.2. **Partitioning over lower dimensional sets.** Theorem 6.2 alone is insufficient for the purposes of this article and a more involved partitioning result, which is implicit in [12], will be used.

Theorem 6.3 (Guth [12]). *Fix $r \gg 1$, $d \in \mathbb{N}$ and suppose $F \in L^1(\mathbb{R}^n)$ is non-negative and supported on $B_r \cap N_{r^{1/2+\delta_\circ}} \mathbf{Z}$ for some $0 < \delta_\circ \ll 1$, where $\mathbf{Z}$ is an m-dimensional transverse complete intersection of degree at most d. At least one of the following cases holds:*

<u>Cellular case</u>. *There exists a polynomial $P \colon \mathbb{R}^n \to \mathbb{R}$ of degree $O(d)$ with the following properties:*

i) *$\#\mathrm{cell}(P) \sim d^m$ and each $O \in \mathrm{cell}(P)$ has diameter at most $r/2$.*
ii) *One may pass to a refinement of $\mathrm{cell}(P)$ such that if $\mathcal{O}$ is defined as in (15), then*

$$\int_O F \sim d^{-m} \int_{\mathbb{R}^n} F \qquad \text{for all } O \in \mathcal{O}.$$

<u>Algebraic case</u>. *There exists an $(m-1)$-dimensional transverse complete intersection $\mathbf{Y}$ of degree at most $O(d)$ such that*

$$\int_{B_r \cap N_{r^{1/2+\delta_\circ}} \mathbf{Z}} F \lesssim \int_{B_r \cap N_{r^{1/2+\delta_\circ}} \mathbf{Y}} F.$$

The choice of scales r and $r^{1/2+\delta_\circ}$ is not particularly special in the sense that the theorem holds true in greater generality: the result is presented in this specific case only in anticipation of later applications.

The statement of this theorem does not explicitly appear in [12], but it can be easily deduced from the argument described in Section 8.1 of that article together with the simple pigeonholing arguments discussed earlier in this subsection. The key difference between Theorem 6.3 and Theorem 6.2 is that in the latter one has the additional hypothesis that F is supported in a $r^{1/2+\delta_\circ}$-neighbourhood of the m-dimensional variety $\mathbf{Z}$. This allows one to construct a partitioning polynomial which cuts out only $O(d^m)$ cells rather than the $O(d^n)$ guaranteed by Theorem 6.2.

Theorem 6.3 is then applied to the relevant broad norm by taking

$$F = \sum_{B_{K^2} \in \mathcal{B}_{K^2}} \mu_{Ef}(B_{K^2}) \frac{1}{|B_{K^2}|} \mathbf{1}_{B_{K^2} \cap B_r \cap N_{r^{1/2+\delta_\circ}} \mathbf{Z}}$$

for some $0 < \delta_\circ \ll 1$.

- If the cellular case holds, then it follows that

$$\|Ef\|^p_{\mathrm{BL}^p_{k,A}(B_r \cap N_{r^{1/2+\delta_\circ}} \mathbf{Z})} \lesssim d^m \|Ef\|^p_{\mathrm{BL}^p_{k,A}(O)} \qquad \text{for all } O \in \mathcal{O}$$

  where $\mathcal{O}$ is the collection of cells produced by the theorem.
- If the algebraic case holds, then it follows that

$$\|Ef\|^p_{\mathrm{BL}^p_{k,A}(B_r \cap N_{r^{1/2+\delta_\circ}} \mathbf{Z})} \lesssim \|Ef\|^p_{\mathrm{BL}^p_{k,A}(B_r \cap N_{r^{1/2+\delta_\circ}} \mathbf{Y})}$$

  where $\mathbf{Y}$ is the variety produced by the theorem.

7. Wave packet decompositions

7.1. **Definition and basic properties.** Let $r \gg 1$ and cover the domain B^{n-1} by a family Θ_r of finitely-overlapping balls of radius $r^{-1/2}$. As noted in Section 4, these $(n-1)$-dimensional balls are referred to as $r^{-1/2}$-*caps* and ξ_θ is used to denote the centre of θ. Fix $(\psi_\theta)_{\theta \in \Theta_r}$ a smooth partition of unity for B^{n-1}, subordinate to the cover Θ_r, such that each function $\xi \mapsto \psi_\theta(\xi_\theta + r^{-1/2}\xi)$ is supported in $[-\pi, \pi]^{n-1}$ and

$$\|\partial_x^\alpha \psi_\theta\|_{L^\infty(\mathbb{R}^{n-1})} \lesssim_\alpha r^{|\alpha|/2} \qquad \text{for all } \alpha \in \mathbb{N}_0^{n-1}.$$

Given our smooth, bounded input function $f \colon B^{n-1} \to \mathbb{C}$, by performing a Fourier series decomposition, we have

$$f \cdot \psi_\theta(\xi) = \Big(\frac{r^{1/2}}{2\pi}\Big)^{n-1} \sum_{v \in r^{1/2}\mathbb{Z}^{n-1}} e^{i\langle v, \xi\rangle} (f \cdot \psi_\theta)^\wedge(v).$$

Writing $\mathbb{T}[r] := \Theta_r \times r^{1/2}\mathbb{Z}^{n-1}$, this yields

$$f = \sum_{(\theta,v)\in\mathbb{T}[r]} f_{\theta,v} \tag{16}$$

where

$$f_{\theta,v}(\xi) := \Big(\frac{r^{1/2}}{2\pi}\Big)^{n-1} e^{i\langle v, \xi\rangle} (f \cdot \psi_\theta)^\wedge(v) \tilde\psi_\theta(\xi)$$

for $\tilde\psi_\theta$ a bump function which is also adapted to $\theta \in \Theta_r$ but which is equal to 1 on the support of ψ_θ. The sum (16) is referred to as the *wave packet decomposition of f at scale r*. The functions $f_{\theta,v}$ and the pairs $(\theta, v) \in \mathbb{T}[r]$ will both be referred to as *(scale r) wave packets*.

The key properties of this decomposition are as follows:

Orthogonality between the wave packets. Recall that the ψ_θ have almost disjoint supports. Combining this observation with the Plancherel identity for Fourier series, one concludes that

$$\Big\| \sum_{(\theta,v)\in\mathbb{W}} f_{\theta,v} \Big\|_2^2 \sim \sum_{(\theta,v)\in\mathbb{W}} \|f_{\theta,v}\|_2^2$$

for any collection of wave packets $\mathbb{W} \subseteq \mathbb{T}[r]$.

It is worth noting that there is a local version of this orthogonality relation. In particular, for $1 \leqslant \rho \leqslant r$ and a $\rho^{-1/2}$-cap θ_*, one may readily verify that

$$\Big\| \sum_{(\theta,v)\in\mathbb{W}} f_{\theta,v} \Big\|_{L^2(\theta_*)}^2 \lesssim \sum_{(\theta,v)\in\mathbb{W}} \|f_{\theta,v}\|_{L^2(3\theta_*)}^2$$

where the right-hand norm is over the cap $3\theta_*$ concentric to θ_* but with thrice the radius. A reverse form of this inequality also holds (with θ_* on the left and $3\theta_*$ on the right-hand side), and together they imply the more symmetric estimate

$$\max_{\theta_*:\rho^{-1/2}-\mathrm{cap}} \Big\| \sum_{(\theta,v)\in\mathbb{W}} f_{\theta,v} \Big\|_{L^2(\theta_*)}^2 \sim \max_{\theta_*:\rho^{-1/2}-\mathrm{cap}} \sum_{(\theta,v)\in\mathbb{W}} \|f_{\theta,v}\|_{L^2(\theta_*)}^2,$$

where the maximum is over all $\rho^{-1/2}$-caps.

Spatial concentration. Given any wave packet $(\theta, v) \in \mathbb{T}[r]$, on the ball $B(0, r)$ the function $Ef_{\theta,v}$ is essentially supported on the tube

$$\Big\{ x \in B(0, r) \,:\, |x' + 2x_n\xi_\theta + v| \leqslant r^{1/2} \Big\}$$

in the sense that $|Ef_{\theta,v}(x)|$ decays rapidly as $x \in B(0, r)$ moves away from this set. More precisely, a simple stationary phase analysis shows that

$$|Ef_{\theta,v}(x)| \lesssim_N r^{-\frac{n-1}{4}} (1 + r^{-1/2}|x' + 2x_n\xi_\theta + v|)^{-N} \|f_{\theta,v}\|_2$$

for all $N \in \mathbb{N}$ and $x \in \mathbb{R}^n$ with $|x_n| < r$; see, for example, [35, Lemma 4.1]. In particular, given $0 < \delta \ll 1$, the function $|Ef_{\theta,v}|$ is very small away from the slightly fattened tube

$$T_{\theta,v} := \Big\{ x \in B(0, r) \,:\, |x' + 2x_n\xi_\theta + v| \leqslant r^{1/2+\delta} \Big\},$$

satisfying

$$|Ef_{\theta,v}(x)\mathbf{1}_{B(0,r)\setminus T_{\theta,v}}(x)| \lesssim_{\delta,N} r^{-N}\|f_{\theta,v}\|_2 \tag{17}$$

for all $N \in \mathbb{N}$ and $x \in \mathbb{R}^n$ with $|x_n| < r$. Note that $T_{\theta,v}$ as defined above is a tube with direction $G(\xi_\theta)$ (where G is the Gauss map as defined in (8)) which passes through the point $(-v, 0) \in \mathbb{R}^n$.

Rapidly decaying terms of the kind seen in (17) are a regular feature of the forthcoming analysis and it is convenient to introduce the notation $\mathrm{RapDec}(r)$ to denote a non-negative term which is rapidly decreasing in r: that is,

$$\mathrm{RapDec}(r) \lesssim_{\delta,N} r^{-N} \qquad \text{for all } N \in \mathbb{N}.$$

Thus, with this definition, the estimate in (17) can be succinctly written as

$$|Ef_{\theta,v}(x)\mathbf{1}_{B(0,r)\setminus T_{\theta,v}}(x)| = \mathrm{RapDec}(r)\|f_{\theta,v}\|_2$$

for all $x \in \mathbb{R}^n$ with $|x_n| < r$.

7.2. **Comparing wave packet decompositions at different scales.** For r as above, consider a smaller scale ρ satisfying $r^{1/2} \leqslant \rho \leqslant r$ and a ball $B(y, \rho)$ with centre $y \in B(0, r)$. We decompose f into wave packets over the ball $B(y, \rho)$ at this smaller spatial scale. The first step is to apply a transformation to recentre $B(y, \rho)$ at the origin. In particular, write $Ef(x) = E\tilde{f}(\tilde{x})$ where $x = y + \tilde{x}$ for some $\tilde{x} \in B(0, \rho)$ and

$$\tilde{f}(\xi) := e^{i(\langle y', \xi\rangle + y_n|\xi|^2)} f(\xi).$$

The function $\tilde{f}$ is now decomposed into scale ρ wave packets;

$$\tilde{f} = \sum_{(\tilde{\theta},\tilde{v}) \in \mathbb{T}[\rho]} \tilde{f}_{\tilde{\theta},\tilde{v}}. \tag{18}$$

A basic question, studied in detail in [12, Section 7], is to understand how the two wave packet decompositions (16) and (18) relate to one another. For instance, suppose the significant contributions to f come from a subcollection $\mathbb{W}$ of the scale r wave packets; which scale ρ wave packets contribute significantly to f? To make this question precise, we introduce the following definition.

Definition 7.1. The function $f \colon B^{n-1} \to \mathbb{C}$ is said to be concentrated on wave packets from $\mathbb{W}$ if

$$\big\|\sum_{(\theta,v) \notin \mathbb{W}} f_{\theta,v}\big\|_\infty = \mathrm{RapDec}(r)\|f\|_2.$$

With this definition, the following lemma provides a relationship between wave packet concentration properties at distinct scales.

Lemma 7.2 ([12])**.** *If f is concentrated on scale r wave packets $\mathbb{W} \subseteq \mathbb{T}[r]$, then $\tilde{f}$ is concentrated on a set of wave packets $\widetilde{\mathbb{W}} \subseteq \mathbb{T}[\rho]$ with the following property: for every $(\tilde{\theta}, \tilde{v}) \in \widetilde{\mathbb{W}}$ there exists a wave packet $(\theta, v) \in \mathbb{W}$ such that*

i) $\mathrm{dist}_H\big(T_{\tilde{\theta},\tilde{v}} + y, T_{\theta,v} \cap B(y, \rho)\big) \lesssim r^{1/2+\delta}$;

ii) $\angle(G(\xi_\theta), G(\xi_{\tilde{\theta}})) \lesssim \rho^{-1/2}$.

Here dist_H denotes the Hausdorff distance.

The lemma tells us that every small scale wave packet $(\tilde{\theta}, \tilde{v}) \in \widetilde{\mathbb{W}}$ has a 'parent' large scale wave packet $(\theta, v) \in \mathbb{W}$ such that $T_{\tilde{\theta},\tilde{v}}$ both lies close to $T_{\theta,v}$ and points in a similar direction to $T_{\theta,v}$. This behaviour is represented in Figure 2.
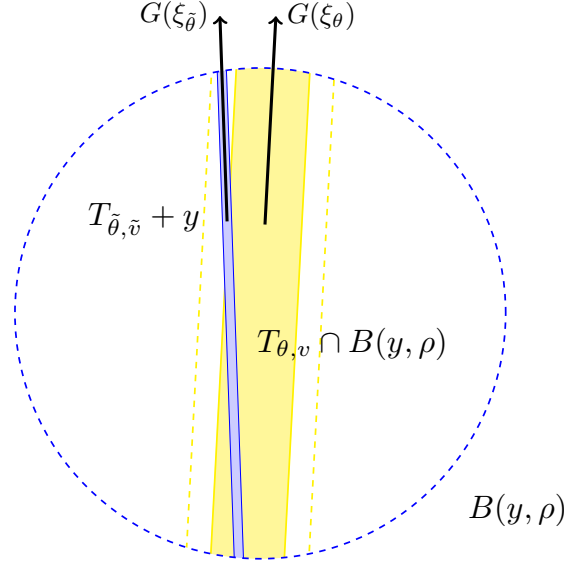
FIGURE 2. For every $(\tilde{\theta}, \tilde{v}) \in \widetilde{\mathbb{W}}$ there exists a 'parent' wave packet $(\theta, v) \in \mathbb{W}$ such that: i) $T_{\tilde{\theta}, \tilde{v}} + y$ (denoted here in blue) is contained a fixed dilate of $T_{\theta, v} \cap B(y, \rho)$ (denoted here in yellow) and ii) the angle between the directions is $O(\rho^{-1/2})$.

8. Tangential wave packets

We begin by giving the precise definition of what it means for a tube $T_{\theta, v}$ to be tangent to $\mathbf{Z}$; throughout this section $\mathbf{Z} \subset \mathbb{R}^n$ will denote an m-dimensional transverse complete intersection and $0 < \delta \ll \delta_m \ll 1$ are fixed small parameters, where δ is as in the previous section.

Definition 8.1. Letting $r \gg 1$ and $y \in B_R$, a (translated) tube $T_{\theta, v} + y$ for $(\theta, v) \in \mathbb{T}[r]$ is said to be $r^{-1/2+\delta_m}$-tangent to $\mathbf{Z}$ in $B(y, r)$ if:

i) $T_{\theta, v} + y \subseteq N_{r^{1/2+\delta_m}} \mathbf{Z} \cap B(y, r)$;

ii) For any $x \in T_{\theta, v} + y$ and $z \in \mathbf{Z} \cap B(y, r)$ with $|z - x| \lesssim r^{1/2+\delta_m}$ one has

$$\angle(G(\theta), T_z \mathbf{Z}) \lesssim r^{-1/2+\delta_m}.$$

Throughout this section, we consider a function g which is concentrated on tangential wave packets in the sense that

$$g = \sum_{(\theta, v) \in \mathbb{T}_{\mathbf{Z}}[r]} g_{\theta, v} + \mathrm{RapDec}(r)\|g\|_2 \tag{19}$$

where

$$\mathbb{T}_{\mathbf{Z}}[r] := \left\{ (\theta, v) \in \mathbb{T}[r] \, : \, T_{\theta, v} \text{ is } r^{-1/2+\delta_m}\text{-tangent to } \mathbf{Z} \text{ in } B(0, r) \right\}.$$

An important ingredient in the proof of Theorem 2.1 will be to understand what can be said about Eg under this tangency hypothesis. Recall from the discussion in Section 4 that there are three useful estimates at our disposal:

- Vanishing property of the k-broad norms,
- Transverse equidistribution estimates,

- Bounds arising from the polynomial Wolff axioms.

The purpose of this section is to provide the precise details of all three of these estimates. The first two were observed and used by Guth [12] to prove restriction estimates in high dimensions. The polynomial Wolff axioms were also applied earlier by Guth [11] in the special case of 2-surfaces in $\mathbb{R}^3$ to study the restriction problem in 3-dimensions.

8.1. **Vanishing property of the k-broad norms.** The key advantage of working with k-broad norms rather than classical L^p inequalities is that the former satisfy the following property.

Lemma 8.2. *Let $r \gg 1$, let $1 \leqslant m < k \leqslant n$, and let $\mathbf{Z}$ be m-dimensional. Suppose that g is concentrated on wave packets from $\mathbb{T}_\mathbf{Z}[r]$. Then*

$$\|Eg\|_{\mathrm{BL}^p_{k,A}(B_r)} = \mathrm{RapDec}(r)\|g\|_2.$$

The lemma follows fairly directly from the definition of the k-broad norms and the basic properties of the wave packet decomposition. The simple argument can be readily extracted from the beginning of the proof of Proposition 8.1 in [12].

8.2. **Comparing tangency properties at different scales.** The description of the transverse equidistribution estimates is a little involved and will require some preliminary definitions. In Section 7.2 we compared wave packet concentration properties at different spatial scales; we now pursue this investigation further in the tangential scenario.

As above, suppose g is concentrated on wave packets from $\mathbb{T}_\mathbf{Z}[r]$. Once again, let $r^{1/2} \leqslant \rho \leqslant r$ be a choice of smaller spatial scale and consider some ρ-ball $B(y, \rho)$ with centre $y \in B(0, r)$. Lemma 7.2 can be used to analyse the tangency properties of the scale ρ wave packets defined over the ball $B(y, \rho)$. To see this, first write

$$\tilde{g}(\xi) := e^{i(\langle y', \xi \rangle + y_n|\xi|^2)}g(\xi),$$

as in Section 7.2. By Lemma 7.2, the function $\tilde{g}$ is concentrated on scale ρ wave packets which each admit a 'parent' wave packet in $\mathbb{T}_\mathbf{Z}[r]$. The scale ρ wave packets therefore inherit tangency properties from their parents. It turns out that the angle condition inherited by the scale ρ wave packets is very strong, but the containment property is too weak to ensure that the scale ρ wave packets are tangent to $\mathbf{Z}$ itself. However, as shown in [12, Section 7], the function $\tilde{g}$ *is* concentrated on scale ρ wave packets $T_{\tilde{\theta}, \tilde{v}}$ which are tangent to various *translates* of $\mathbf{Z}$. A schematic of this behaviour is provided in Figures 3 and 4 below.

To make the preceding discussion more precise, given $b \in \mathbb{R}^n$ let $\mathbb{T}_b[\rho]$ be the subcollection of $\mathbb{T}[\rho]$ consisting of those wave packets which are $\rho^{-1/2+\delta_m}$-tangent to $\mathbf{Z} + b - y$ in $B(0, \rho)$. At least heuristically, there is a finite set of translates $\mathfrak{B} \subseteq B(0, r^{1/2+\delta_m})$ such that the $\{\mathbb{T}_b[\rho] : b \in \mathfrak{B}\}$ are pairwise disjoint and

$$\tilde{g} = \sum_{b \in \mathfrak{B}} \tilde{g}_b + \mathrm{RapDec}(r)\|g\|_2, \quad \text{where} \quad \tilde{g}_b := \sum_{(\tilde{\theta}, \tilde{v}) \in \mathbb{T}_b[\rho]} \tilde{g}_{\tilde{\theta}, \tilde{v}}. \tag{20}$$

By the spatial concentration property of the wave packets, it follows that

$$E\tilde{g}_b(\tilde{x}) = \mathbf{1}_{N_{\rho^{1/2+\delta_m}}(\mathbf{Z}+b)}(x)Eg(x) + \mathrm{RapDec}(r)\|g\|_2$$

whenever $x = \tilde{x} + y$ for some $\tilde{x} \in B(0, \rho)$. The decomposition in (20) therefore breaks $E\tilde{g}$ into pieces with the property that each piece is concentrated on a $\rho^{1/2+\delta_m}$-neighbourhood of some translate of $\mathbf{Z}$.

Finding the set of translates $\mathfrak{B}$ involves some technicalities and the precise statements are perhaps not quite as clean as the above discussion suggests. A rigorous
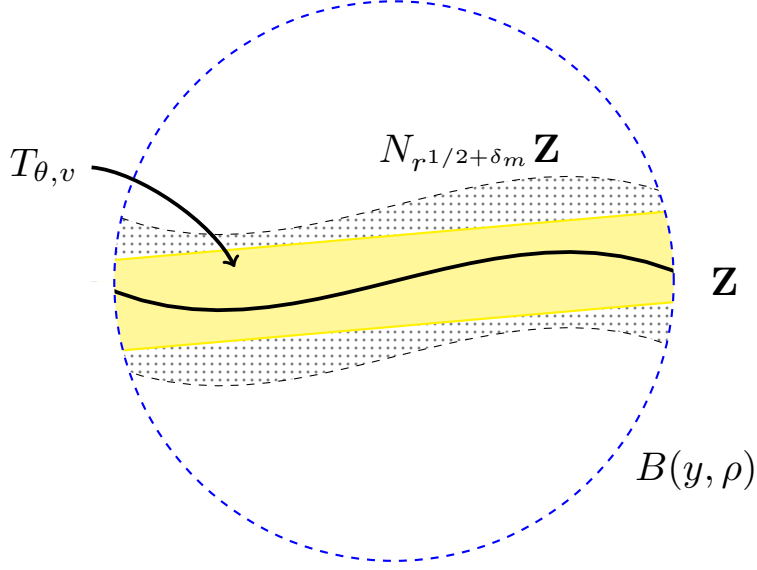
FIGURE 3. The large scale tube $T_{\theta,v}$ is $r^{-1/2+\delta_m}$-tangent to $\mathbf{Z}$ in $B(0,r)$. Here we consider its intersection with $B(y,\rho)$.

version of (20) is given by the following proposition, which is implicit in [12] and is described more explicitly in [13].

Proposition 8.3. *Let $B(y,\rho) \cap N_{\rho^{1/2+\delta_m}}\mathbf{Z} \neq \emptyset$ and let g be concentrated on wave packets from $\mathbb{T}_{\mathbf{Z}}[r]$. Then there is a set of translates $\mathfrak{B} \subset B(0,r^{1/2+\delta_m})$ such that*

$$\|Eg\|^p_{\mathrm{BL}^p_{k,A}(B(y,\rho))} \lesssim \log^2 r \sum_{b \in \mathfrak{B}} \|E\tilde{g}_b\|^p_{\mathrm{BL}^p_{k,A}(B(0,\rho)\cap N_{\rho^{1/2+\delta_m}}(\mathbf{Z}-y+b))} + \mathrm{RapDec}(r)\|g\|_2^2$$

and

$$\sum_{b \in \mathfrak{B}} \|\tilde{g}_b\|_2^2 \lesssim \|g\|_2^2.$$

The lemma can be proved by independently selecting the translates b at random, although this argument involves some technicalities. See the proof of [12, Proposition 8.1] or [13, Lemma 10.5] for further details.

8.3. **Transverse equidistribution estimates.** If $h\colon B^{n-1} \to \mathbb{C}$ is concentrated on wave packets from $\mathbb{T}_{\mathbf{Z}}[r]$, then this property constrains the support of h (since points of the support of h roughly correspond to directions of the wave packets). This in turn influences the behaviour of Eh via the uncertainty principle. In particular, it transpires that Eh is essentially constant at scale $r^{1/2}$ in directions transverse to the variety $\mathbf{Z}$. This phenomenon is encapsulated in the *transverse equidistribution estimate* of [12, Section 6] which roughly states that

$$\frac{1}{|N_{\rho^{1/2}}\mathbf{Z} \cap B_{r^{1/2}}|} \int_{N_{\rho^{1/2}}\mathbf{Z}\cap B_{r^{1/2}}} |Eh|^2 \lesssim \frac{1}{|B_{r^{1/2}}|} \int_{B_{r^{1/2}}} |Eh|^2 \qquad (21)$$

for any $r^{1/2}$-ball $B_{r^{1/2}}$ and $1 \leqslant \rho \leqslant r$. An informative case to have in mind is given by taking $\mathbf{Z}$ to be a plane in the co-ordinate hyperplane perpendicular to e_n; in this situation, a rigorous version of the above inequality can be readily verified along the lines discussed above. For the general case, the reader is referred to Sections 2 and 6
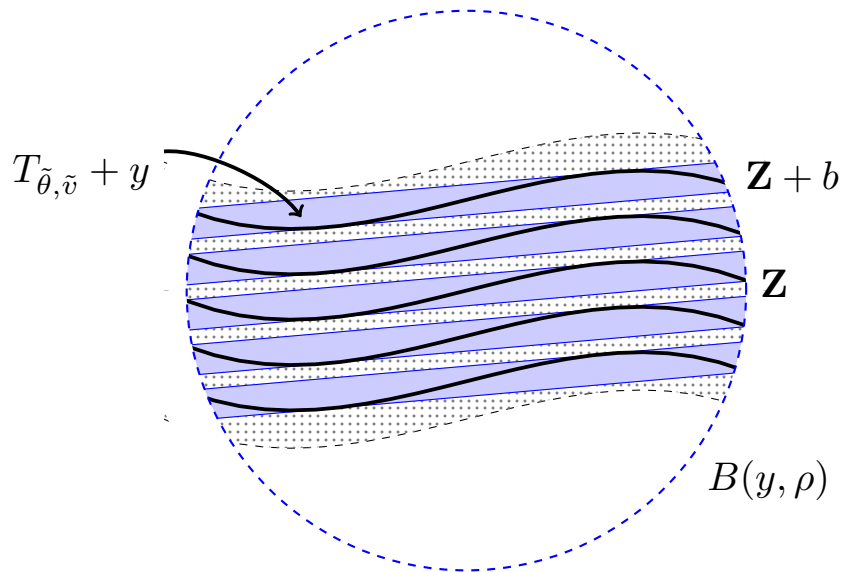
FIGURE 4. The scale ρ wave packets are partitioned into collections $\mathbb{T}_b[\rho]$. For each $(\tilde{\theta}, \tilde{v}) \in \mathbb{T}_b[\rho]$ the corresponding tube $T_{\tilde{\theta}, \tilde{v}} + y$ is tangent to the translate $\mathbf{Z} + b$ in $B(y, \rho)$.

of [12] for a more detailed discussion of the transverse equidistribution phenomenon, which plays a fundamental role in [12] and also here.

It is of particular interest to apply these observations to $h := \tilde{g}_b$, where $\tilde{g}_b$ is one of the functions introduced in the previous subsection. Indeed, by the discussion in Section 8.2, the operator $|E\tilde{g}_b|$ is concentrated in $N_{\rho^{1/2+\delta_m}}(\mathbf{Z} - y + b)$ and so expressions of the form of the left-hand side of (21) naturally arise in this context.

Estimates for L^2 quantities involving $E\tilde{g}_b$ can be related to L^2 estimates for the input function $\tilde{g}_b$ via Plancherel's theorem or, more precisely, the energy identity (11). The following consequence of transverse equidistribution will be useful, which is established in Section 7 of [12].

Lemma 8.4 (Guth [12]). *Let* $1 \leqslant \rho' \leqslant \rho \leqslant r$ *and* $|b| \lesssim r^{1/2+\delta_m}$. *Let* $\mathbf{Z}$ *be* m-*dimensional and let* g *be concentrated on wave packets from* $\mathbb{T}_{\mathbf{Z}}[r]$. *Then*

$$\max_{\theta:(\rho')^{-1/2}-\mathrm{cap}} \|\tilde{g}_b\|_{L^2(\theta)}^2 \lesssim_{\overline{\deg}\,\mathbf{Z}} r^{O(\delta_m)} \left(\frac{r}{\rho}\right)^{-\frac{n-m}{2}} \max_{\theta:(\rho')^{-1/2}-\mathrm{cap}} \|g\|_{L^2(\theta)}^2, \qquad (22)$$

where $\tilde{g}_b$ *is defined with respect to scale* ρ *wave packets as in* (20).

Note that the factor gained in (22) is the ratio of the volumes of the sets of integration in (21). The inequality (22) is explicitly stated in [12, Lemma 7.6] for the case $\rho' = 1$; the version for general $1 \leqslant \rho' \leqslant \rho$ can be deduced via similar arguments (see also the equation (8.26) from [12]).

8.4. **Applying the polynomial Wolff axioms.** Theorem 4.1 can be expressed in terms of wave packets.

Proposition 8.5 ([18]). *Let $\delta > 0$, $c, r \geqslant 1$ and $\mathbf{Z} \subseteq \mathbb{R}^n$ be m-dimensional. If $\mathbb{W} \subseteq \mathbb{T}[r]$ is such that $T_{\theta,v} \subseteq N_{cr^{1/2}}\mathbf{Z}$ for all $(\theta, v) \in \mathbb{W}$, then*

$$\#\Big\{ \theta \,:\, (\theta, v) \in \mathbb{W} \text{ for some } v \in r^{1/2}\mathbb{Z}^{n-1} \Big\} \lesssim_{\overline{\deg}\mathbf{Z}} c^{n-m}r^{\frac{m-1}{2}+\delta}.$$

From this geometric bound, we deduce an estimate involving the averaged norm

$$\|f\|_{L^2_{\mathrm{avg}}(\theta)} := \Big(\frac{1}{|\theta|} \int_\theta |f(\xi)|^2 \mathrm{d}\xi \Big)^{1/2},$$

which is a higher dimensional generalisation of an inequality that featured prominently in [11].

Lemma 8.6. *Let $\delta > 0$, $c, r \geqslant 1$ and $\mathbf{Z} \subseteq \mathbb{R}^n$ be m-dimensional. If g is concentrated on wave packets $(\theta, v) \in \mathbb{T}[r]$ satisfying $T_{\theta,v} \subseteq N_{cr^{1/2}}\mathbf{Z}$, then*

$$\|g\|^2_{L^2(B^{n-1})} \lesssim_{\overline{\deg}\mathbf{Z}} c^{n-m}r^{-\frac{n-m}{2}+\delta} \max_{\theta:r^{-1/2}-\mathrm{cap}} \|g\|^2_{L^2_{\mathrm{avg}}(\theta)}.$$

Proof. By the concentration hypothesis one may write

$$g = \sum_{(\theta,v)\in\mathbb{W}} g_{\theta,v} + \mathrm{RapDec}(r)\|g\|_2$$

where $\mathbb{W}$ are scale r wave packets satisfying $T_{\theta,v} \subseteq N_{cr^{1/2}}\mathbf{Z}$. Given an $r^{-1/2}$-cap θ, define

$$\mathbb{T}_{\mathbf{Z}}(\theta) := \Big\{ v \in r^{1/2}\mathbb{Z}^{n-1} : (\theta, v) \in \mathbb{W} \Big\}$$

and let $\Theta_{\mathbf{Z}}$ denote the collection of all $r^{-1/2}$-caps θ for which $\mathbb{T}_{\mathbf{Z}}(\theta) \neq \emptyset$. Thus, by the orthogonality and support properties of the wave packets,

$$\|g\|^2_2 \sim \sum_{\theta\in\Theta_{\mathbf{Z}}} \Big\| \sum_{v\in\mathbb{T}_{\mathbf{Z}}(\theta)} g_{\theta,v} \Big\|^2_2 \lesssim \sum_{\theta\in\Theta_{\mathbf{Z}}} \|g\|^2_{L^2(\theta)}.$$

To prove the lemma it therefore suffices to show that

$$\#\Theta_{\mathbf{Z}} \lesssim_{\overline{\deg}\mathbf{Z}} c^{n-m}r^{\frac{m-1}{2}+\delta},$$

but this immediately follows from Proposition 8.5. $\square$

9. FINDING POLYNOMIAL STRUCTURE

The purpose of this section is to reformulate the core of the (inductive) proofs in [11, 12] as a recursive process. The argument will in fact be presented as two separate algorithms:

- [alg 1] is the more involved of the two and is presented in the current section. It effects a dimensional reduction, essentially passing from an m-dimensional to an $(m-1)$-dimensional situation.
- [alg 2] is described in Section 10 below. It consists of repeated application of the first algorithm to reduce to a minimal dimensional case.

Comparing the present analysis with the original induction arguments of Guth, [alg 1] corresponds to the induction on the radius in the proof of Proposition 8.1 of [12], whilst [alg 2] corresponds to the induction on dimension.

The first algorithm. Throughout this section let $p \geqslant 2$, $0 < \varepsilon \ll 1$ be fixed and

$$\varepsilon^C \leqslant \delta \ll \delta_n \ll \delta_{n-1} \ll \cdots \ll \delta_1 \ll \delta_0 \ll \varepsilon \qquad (23)$$

be a family of small parameters. Taking, for instance, $\delta_0 := \varepsilon^{10}$, $\delta_j := \delta_{j-1}^{10}$ for $1 \leqslant j \leqslant n$ and $\delta := \delta_n^{10}$ suffices. These parameters play a rather technical role[10] and are chosen so as to satisfy the requirements of the forthcoming proof.

<u>Input</u>. [alg 1] will take as its input:
- An r-ball $B_r \subset \mathbb{R}^n$ for some choice of large scale $r \gg 1$.
- A transverse complete intersection $\mathbf{Z}$ of dimension $m \geqslant 2$.
- A function $f \in B^{n-1} \to \mathbb{C}$ concentrated on wave packets which are $r^{-1/2+\delta_m}$-tangent to $\mathbf{Z}$ in B_r.
- An admissible large integer $A \in \mathbb{N}$.

Remark 9.1. The integer A corresponds to the A parameter featured in the definition of the broad norm. It is chosen large enough to facilitate repeated application of Lemma 5.2 and Lemma 5.3. These lemmas will be used no more than δ^{-2} times and so it suffices to take $A \geqslant 2^{\delta^{-2}}$: see the discussion following (28) below.

. The description of the output of the algorithm is, unfortunately, far more involved.

<u>Output</u>. [alg 1] will output a finite sequence of sets $(\mathscr{E}_j)_{j=0}^J$, which are constructed via a recursive process. Each $\mathscr{E}_j$ is referred to as an *ensemble* and contains all the relevant information coming from the jth step of the algorithm. In particular, the ensemble $\mathscr{E}_j$ consists of:
- A word $\mathfrak{h}_j$ of length j in the alphabet $\{\mathtt{a}, \mathtt{c}\}$, is referred to as a *history*. The rationale behind this notation is that $\mathtt{a}$ is an abbreviation of 'algebraic' and $\mathtt{c}$ 'cellular'. The words $\mathfrak{h}_j$ are recursively defined by successively adjoining a single letter. Each $\mathfrak{h}_j$ records how the cells $O_j \in \mathcal{O}_j$ were constructed via repeated application of the polynomial partitioning theorem and, in particular, whether the algebraic or cellular case held in successive stages of the process.
- A choice of spatial scale $\rho_j \geqslant 1$. The ρ_j will in fact be completely determined by the initial scale r and the history $\mathfrak{h}_j$. In particular, define an auxiliary exponent $\tilde{\delta}_{m-1}$ by

$$(1 - \tilde{\delta}_{m-1})(1/2 + \delta_{m-1}) = (1/2 + \delta_m), \qquad (24)$$

noting that $\delta_{m-1}/2 \leqslant \tilde{\delta}_{m-1} \leqslant 2\delta_{m-1}$. Let $\sigma_k \colon [1, \infty) \to [0, \infty)$ be given by

$$\sigma_k(\rho) := \begin{cases} \dfrac{\rho}{2} & \text{if the } k\text{th letter of } \mathfrak{h}_j \text{ is } \mathtt{c} \\ \rho^{1-\tilde{\delta}_{m-1}} & \text{if the } k\text{th letter of } \mathfrak{h}_j \text{ is } \mathtt{a} \end{cases}$$

for each $1 \leqslant k \leqslant j$. With these definitions, take

$$\rho_j := \sigma_j \circ \cdots \circ \sigma_1(r);$$

this sequence of scales is represented pictorially by the tree in Figure 5.

Note that each σ_k is a decreasing function and therefore

$$\rho_j \leqslant r^{(1-\tilde{\delta}_{m-1})^{\#\mathtt{a}(j)}} \quad \text{and} \quad \rho_j \leqslant \frac{r}{2^{\#\mathtt{c}(j)}} \qquad (25)$$

where $\#\mathtt{a}(j)$ and $\#\mathtt{c}(j)$ denote the number of occurrences of $\mathtt{a}$ and $\mathtt{c}$ in the history $\mathfrak{h}_j$, respectively.

---

[10]They are essentially used to compensate for certain $r^{\bar{C}\delta_m}$-losses arising from the transverse equidistribution lemma.
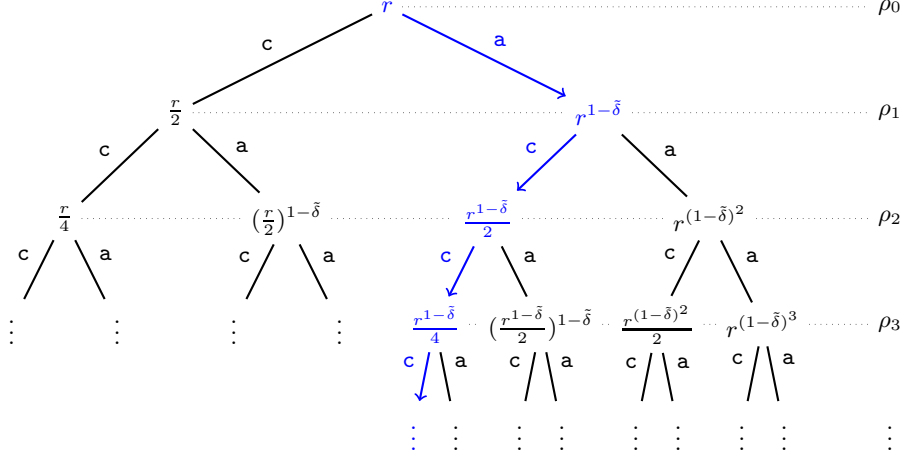
FIGURE 5. The chain of scales ρ_j. Here we drop the subscript by writing $\tilde{\delta} := \tilde{\delta}_{m-1}$. The blue path marked by arrows corresponds to the situation where the word $\mathfrak{h}_j$ of length j begins with the letters $\texttt{accc}\ldots$ for $4 \leqslant j \leqslant J$. In this case the sequence ρ_j is given by $\rho_0 = r$, $\rho_1 = r^{1-\tilde{\delta}}$, $\rho_2 = r^{1-\tilde{\delta}}/2$, $\rho_3 = r^{1-\tilde{\delta}}/4$, ...

Remark 9.2. It is perhaps useful to give some justification for the introduction of the auxiliary exponent $\tilde{\delta}_{m-1}$. In the algebraic case one passes from some scale ρ_j to a new scale $\rho_{j+1} := \rho_j^{1-\tilde{\delta}_{m-1}}$; this is encoded in the definition of $\sigma_k(\rho)$ above. In the last step of the current algorithm the analysis passes to a lower dimensional variety (or we end up in a trivial small scale case). For this step, one wishes to analyse tangency properties of the wave packets at the new scale ρ_{j+1} with respect to an $(m-1)$-dimensional variety. Looking at the second condition in Definition 8.1, this involves analysis at the scale $\rho_{j+1}^{1/2+\delta_{m-1}}$. The formula (24) ensures that $\rho_{j+1}^{1/2+\delta_{m-1}} = \rho_j^{1/2+\delta_m}$; this allows certain tangency properties to be inherited at the new scale.

- A family of subsets $\mathcal{O}_j$ of $\mathbb{R}^n$ which will be referred to as *cells*. Each cell $O_j \in \mathcal{O}_j$ will have diameter at most ρ_j.
- A collection of functions $(f_{O_j})_{O_j \in \mathcal{O}_j}$. Each f_{O_j} is concentrated on wave packets in $\mathbb{T}[\rho_j]$ which are $\rho_j^{-1/2+\delta_m}$-tangent to some translate of $\mathbf{Z}$ on (a ball of radius ρ_j containing) O_j.
- A large integer $d \in \mathbb{N}$ which depends only on the admissible parameters and $\overline{\deg} \mathbf{Z}$.

Moreover, the components of the ensemble are defined so as to ensure that, for certain coefficients[11]

$$C_{j,\delta}^{\mathrm{I}}(d,r),\; C_{j,\delta}^{\mathrm{II}}(d),\; C_{j,\delta}^{\mathrm{III}}(d,r) \lesssim_{d,\delta} r^{\delta_0} d^{\#\mathtt{c}(j)\delta} \tag{26}$$

and $A_j := 2^{-\#\mathtt{a}(j)} A \in \mathbb{N}$, the following properties hold:

---

[11]The quantity $d^{\#\mathtt{c}(j)\delta}$ may be large (and non-admissible). Nevertheless, these $d^{\#\mathtt{c}(j)\delta}$ losses will be compensated for by other gains in the argument: see Remark 10.1 below.

Property I. Most of the mass of $\|Ef\|^p_{\mathrm{BL}^p_{k,A}(B_r)}$ is concentrated over the $O_j \in \mathcal{O}_j$:

$$\|Ef\|^p_{\mathrm{BL}^p_{k,A}(B_r)} \leqslant C^{\mathrm{I}}_{j,\delta}(d,r) \sum_{O_j \in \mathcal{O}_j} \|Ef_{O_j}\|^p_{\mathrm{BL}^p_{k,A_j}(O_j)} + \mathrm{err}(j) \qquad (\mathrm{I})_j$$

where $\mathrm{err}(j) := jr^{-N}\|f\|^p_2$ for some fixed $N \in \mathbb{N}$ is a harmless 'error' term.

Property II. The functions f_{O_j} satisfy

$$\sum_{O_j \in \mathcal{O}_j} \|f_{O_j}\|^2_2 \leqslant C^{\mathrm{II}}_{j,\delta}(d)d^{\#\mathbf{c}(j)}\|f\|^2_2. \qquad (\mathrm{II})_j$$

Property III. Furthermore, each individual f_{O_j} satisfies

$$\|f_{O_j}\|^2_2 \leqslant C^{\mathrm{III}}_{j,\delta}(d,r)\Big(\frac{r}{\rho_j}\Big)^{-\frac{n-m}{2}} d^{-\#\mathbf{c}(j)(m-1)}\|f\|^2_2 \qquad (\mathrm{III})_j$$

and

$$\max_{\theta:\rho^{-1/2}-\mathrm{cap}} \|f_{O_j}\|^2_{L^2_{\mathrm{avg}}(\theta)} \leqslant C^{\mathrm{III}}_{j,\delta}(d,r)\Big(\frac{r}{\rho_j}\Big)^{-\frac{n-m}{2}} \max_{\theta:\rho^{-1/2}-\mathrm{cap}} \|f\|^2_{L^2_{\mathrm{avg}}(\theta)} \qquad (\mathrm{III}_{\mathrm{loc}})_j$$

for all $1 \leqslant \rho \leqslant \rho_j$.

The factors $C^{\mathrm{I}}_{j,\delta}(d,r)$, $C^{\mathrm{II}}_{j,\delta}(d)$ and $C^{\mathrm{III}}_{j,\delta}(d,r)$ play a minor technical role in the analysis but, nevertheless, it is useful to work with explicit formulæ for these coefficients. In particular, they are defined by

$$C^{\mathrm{I}}_{j,\delta}(d,r) := d^{\#\mathbf{c}(j)\delta}(\log r)^{2\#\mathbf{a}(j)(1+\delta)},$$
$$C^{\mathrm{II}}_{j,\delta}(d) := d^{\#\mathbf{c}(j)\delta+n\#\mathbf{a}(j)(1+\delta)},$$
$$C^{\mathrm{III}}_{j,\delta}(d,r) := d^{\#\mathbf{c}(j)\delta+\#\mathbf{a}(j)\delta}r^{\bar{C}\#\mathbf{a}(j)\delta_m},$$

where $\bar{C}$ is some suitably chosen large constant.

The initial step. The initial ensemble $\mathscr{E}_0$ is defined by taking:

- $\mathfrak{h} := \emptyset$ to be the empty word;
- $\rho_0 := r$;
- $\mathcal{O}_0$ the collection consisting of a single cell $O_0 := N_{r^{1/2+\delta_m}}\mathbf{Z} \cap B_r$;
- $f_{O_0} := f$.

At this point it is convenient also to fix $d \in \mathbb{N}$ to be some large integer, to be determined later, which depends only on admissible parameters and $\overline{\deg}\,\mathbf{Z}$.

With these definitions, Property I holds due to the hypothesis on f and the spatial concentration property of the wave packets, whilst Properties II and III both hold vacuously.

The recursive step. Assume the ensembles $\mathscr{E}_0, \dots, \mathscr{E}_j$ have all been constructed for some $j \in \mathbb{N}_0$ and that they all satisfy the desired properties.

`Stopping conditions`. The algorithm has two stopping conditions which are labelled `[tiny]` and `[tang]`.

Stop:`[tiny]` The algorithm terminates if $\rho_j \leqslant r^{\tilde{\delta}_{m-1}}$.

Stop:`[tang]` Let $C_{\mathtt{tang}}$ and $C_{\mathtt{alg}}$ be fixed admissible constants, chosen large enough to satisfy the forthcoming requirements of the proof, and $\tilde{\rho} := \rho_j^{1-\tilde{\delta}_m}$. The algorithm terminates if the inequalities

$$\sum_{O_j \in \mathcal{O}_j} \|Ef_{O_j}\|^p_{\mathrm{BL}^p_{k,A_j}(O_j)} \leqslant C_{\mathtt{tang}} \sum_{S \in \mathcal{S}} \|Ef_S\|^p_{\mathrm{BL}^p_{k,A_j/2}(B_{\tilde{\rho}}[S])}$$

and

$$\sum_{S \in \mathcal{S}} \|f_S\|_2^2 \leqslant C_{\mathtt{tang}} r^{n\tilde{\delta}_m} \sum_{O_j \in \mathcal{O}_j} \|f_{O_j}\|_2^2; \tag{27}$$

$$\max_{\substack{S \in \mathcal{S} \\ \theta : \rho^{-1/2}-\mathrm{cap}}} \|f_S\|_{L^2_{\mathrm{avg}}(\theta)}^2 \leqslant C_{\mathtt{tang}} \max_{\substack{O_j \in \mathcal{O}_j \\ \theta : \rho^{-1/2}-\mathrm{cap}}} \|f_{O_j}\|_{L^2_{\mathrm{avg}}(\theta)}^2$$

hold for all $1 \leqslant \rho \leqslant \tilde{\rho}$ for some choice of:

- $\mathcal{S}$ a collection of transverse complete intersections in $\mathbb{R}^n$ all of equal dimension $m-1$ and degree at most $C_{\mathrm{alg}} d$;
- $B_{\tilde{\rho}}[S]$ an assignment of a $\tilde{\rho}$-ball to each $S \in \mathcal{S}$;
- f_S an assignment of a function to each $S \in \mathcal{S}$ which is concentrated on wave packets $\tilde{\rho}^{-1/2+\delta_{m-1}}$-tangent to S on $B_{\tilde{\rho}}$ in the sense of Definition 8.1.

The stopping condition [tang] is somewhat involved, but it can be roughly interpreted as forcing the algorithm to terminate if one can pass to a lower dimensional situation.

If either of the above conditions hold, then the stopping time is defined to be $J := j$. Recalling (25), the stopping condition [tiny] implies that the algorithm must terminate after finitely many steps and, moreover,

$$\#\mathtt{a}(J) \lesssim \delta_{m-1}^{-1} \log(\delta_{m-1}^{-1}) \quad \text{and} \quad \#\mathtt{c}(J) \lesssim \log r. \tag{28}$$

These estimates can be combined with the explicit formulæ for $C_{j,\delta}^{\mathrm{I}}(d,r)$, $C_{j,\delta}^{\mathrm{II}}(d)$ and $C_{j,\delta}^{\mathrm{III}}(d,r)$ to show that the bound (26) always holds, provided δ_m is chosen to be sufficiently small relative to δ_{m-1}. Furthermore, by choosing $A \geqslant 2^{\delta^{-2}}$, say, one can ensure that the A_j defined above are indeed integers.

Recursive step. Suppose that neither stopping condition [tiny] nor [tang] is met. One proceeds to construct the ensemble $\mathscr{E}_{j+1}$ as follows.

Given $O_j \in \mathcal{O}_j$, apply the polynomial partitioning theorem, Theorem 6.3, with degree d to

$$\|Ef_{O_j}\|_{\mathrm{BL}_{k,A_j}^p(O_j \cap N_{\rho_j^{1/2+\delta_m}}(\mathbf{Z}+x_{O_j}))}^p = \|Ef_{O_j}\|_{\mathrm{BL}_{k,A_j}^p(O_j)}^p + \mathrm{RapDec}(r)\|f\|_2^p,$$

where $x_{O_j} \in \mathbb{R}^n$ is a choice of translate such that f_{O_j} is concentrated on wave packets $\rho_j^{-1/2+\delta_m}$-tangent to $\mathbf{Z}+x_{O_j}$ in (a ρ_j-ball containing) O_j. For each $O_j \in \mathcal{O}_j$ either the cellular or the algebraic case holds, as defined in Theorem 6.3. Let $\mathcal{O}_{j,\mathrm{cell}}$ denote the subcollection of $\mathcal{O}_j$ consisting of all cells for which the cellular case holds and $\mathcal{O}_{j,\mathrm{alg}} := \mathcal{O}_j \setminus \mathcal{O}_{j,\mathrm{cell}}$. Thus, by $(\mathrm{I})_j$, one may bound $\|Ef\|_{\mathrm{BL}_{k,A}^p(B_r)}^p$ by

$$C_{j,\delta}^{\mathrm{I}}(d,r) \Big[\sum_{O_j \in \mathcal{O}_{j,\mathrm{cell}}} \|Ef_{O_j}\|_{\mathrm{BL}_{k,A_j}^p(O_j)}^p + \sum_{O_j \in \mathcal{O}_{j,\mathrm{alg}}} \|Ef_{O_j}\|_{\mathrm{BL}_{k,A_j}^p(O_j)}^p \Big] + \mathrm{err}(j);$$

the analysis splits into two cases depending on which term in this sum dominates.

▶ **Cellular-dominant case.** Suppose that the inequality

$$\sum_{O_j \in \mathcal{O}_{j,\mathrm{alg}}} \|Ef_{O_j}\|_{\mathrm{BL}_{k,A_j}^p(O_j)}^p \leqslant \sum_{O_j \in \mathcal{O}_{j,\mathrm{cell}}} \|Ef_{O_j}\|_{\mathrm{BL}_{k,A_j}^p(O_j)}^p \tag{29}$$

holds so that

$$\|Ef\|_{\mathrm{BL}_{k,A}^p(B_r)}^p \leqslant 2C_{j,\delta}^{\mathrm{I}}(d,r) \sum_{O_j \in \mathcal{O}_{j,\mathrm{cell}}} \|Ef_{O_j}\|_{\mathrm{BL}_{k,A_j}^p(O_j)}^p + \mathrm{err}(j). \tag{30}$$

Definition of $\mathscr{E}_{j+1}$. Define $\mathfrak{h}_{j+1}$ by adjoining the letter c to the word $\mathfrak{h}_j$. Thus, it follows from the definitions that

$$\rho_{j+1} = \frac{\rho_j}{2}, \quad \#\mathsf{c}(j+1) = \#\mathsf{c}(j) + 1 \quad \text{and} \quad \#\mathsf{a}(j+1) = \#\mathsf{a}(j). \quad (31)$$

The next generation of cells $\mathcal{O}_{j+1}$ will arise from the cellular decomposition of Theorem 6.3. Fix $O_j \in \mathcal{O}_{j,\mathrm{cell}}$ so that there exists some polynomial $P\colon \mathbb{R}^n \to \mathbb{R}$ of degree $O(d)$ with the following properties:

  i) $\#\mathrm{cell}(P) \sim d^m$ and each $O \in \mathrm{cell}(P)$ has diameter at most ρ_{j+1}.
  ii) One may pass to a refinement of $\mathrm{cell}(P)$ such that if

$$\mathcal{O}_{j+1}(O_j) := \left\{ O \setminus N_{\rho_j^{1/2+\delta_m}} Z(P) \,:\, O \in \mathrm{cell}(P) \right\} \quad (32)$$

denotes the corresponding collection of $\rho_j^{1/2+\delta_m}$-shrunken cells, then

$$\|Ef_{O_j}\|^p_{\mathrm{BL}^p_{k,A_j}(O_j)} \lesssim d^m \|Ef_{O_j}\|^p_{\mathrm{BL}^p_{k,A_j}(O_{j+1})} \qquad \text{for all } O_{j+1} \in \mathcal{O}_{j+1}(O_j).$$

Given $O_{j+1} \in \mathcal{O}_{j+1}(O_j)$, define

$$f_{O_{j+1}} := \sum_{\substack{(\theta,v)\in\mathbb{T}[\rho_j] \\ T_{\theta,v}\cap O_{j+1}\neq\emptyset}} (f_{O_j})_{\theta,v}.$$

It is a simple consequence of the fundamental theorem of algebra (or Bézout's theorem) that any tube $T_{\theta,v}$ for $(\theta,v) \in \mathbb{T}[\rho_j]$ can enter at most $O(d)$ cells $O_{j+1} \in \mathcal{O}_{j+1}(O_j)$ (it is for this reason that one works with the collection of *shrunken* cells as defined in (32)). Consequently, by the basic orthogonality between the wave packets,

$$\sum_{O_{j+1}\in\mathcal{O}_{j+1}(O_j)} \|f_{O_{j+1}}\|_2^2 \lesssim d\|f_{O_j}\|_2^2. \quad (33)$$

By the pigeonhole principle, one may therefore pass to a refinement of $\mathcal{O}_{j+1}(O_j)$ such that

$$\|f_{O_{j+1}}\|_2^2 \lesssim d^{-(m-1)}\|f_{O_j}\|_2^2 \qquad \text{for all } O_{j+1} \in \mathcal{O}_{j+1}(O_j). \quad (34)$$

Finally, define

$$\mathcal{O}_{j+1} := \bigcup_{O_j\in\mathcal{O}_{j,\mathrm{cell}}} \mathcal{O}_{j+1}(O_j).$$

This completes the construction of $\mathscr{E}_{j+1}$ and it remains to check that the new ensemble satisfies the desired properties.[12] In view of this, it is useful to note that

$$C^N_{j,\delta}(d,r) = d^{-\delta}C^N_{j+1,\delta}(d,r) \quad \text{for } N \in \{\mathrm{I},\mathrm{II},\mathrm{III}\} \quad \text{and} \quad A_j = A_{j+1}, \quad (35)$$

which follows immediately from (31) and the definition of the $C^N_{j,\delta}(d,r)$ and A_j.

---

[12]There is a slight technical issue here as the $f_{O_{j+1}}$ are required to satisfy the tangency hypothesis at scale ρ_{j+1}; this is not quite directly inherited from the parent f_{O_j} functions since they only satisfy a tangency hypothesis at scale ρ_j. Although the scales differ by only a factor of 2, the construction is applied repeatedly as part of the recursive process and therefore such factors can build up and potentially threaten the argument. One may deal with this problem by performing a further decomposition of the cells O_{j+1} and functions $f_{O_{j+1}}$ using Proposition 8.3: the details are omitted since the argument is similar (but significantly simpler) to that used to treat the algebraic case below. See also Lemma 10.2 of [13].

Property I. Fixing $O_j \in \mathcal{O}_{j,\mathrm{cell}}$, observe that $\#\mathcal{O}_{j+1}(O_j) \sim d^m$ and

$$\|Ef_{O_j}\|_{\mathrm{BL}^p_{k,A_j}(O_j)}^p \lesssim d^m \|Ef_{O_j}\|_{\mathrm{BL}^p_{k,A_j}(O_{j+1})}^p \qquad \text{for all } O_{j+1} \in \mathcal{O}_{j+1}(O_j)$$

by the properties i) and ii) from the polynomial partitioning theorem and the fact that $\mathcal{O}_{j+1}(O_j)$ is obtained by twice refining a set of cardinality comparable to that of $\mathrm{cell}(P)$. Thus,

$$\|Ef_{O_j}\|_{\mathrm{BL}^p_{k,A_j}(O_j)}^p \lesssim \sum_{O_{j+1} \in \mathcal{O}_{j+1}(O_j)} \|Ef_{O_j}\|_{\mathrm{BL}^p_{k,A_j}(O_{j+1})}^p$$

and, recalling (30) and (35), one deduces that

$$\|Ef\|_{\mathrm{BL}^p_{k,A}(B_r)}^p \leqslant Cd^{-\delta} C^{\mathrm{I}}_{j+1,\delta}(d,r) \sum_{O_{j+1} \in \mathcal{O}_{j+1}} \|Ef_{O_j}\|_{\mathrm{BL}^p_{k,A_{j+1}}(O_{j+1})}^p + \mathrm{err}(j).$$

By the definition of $f_{O_{j+1}}$ and the spatial concentration property of the wave packets, it follows that

$$Ef_{O_j}(x) = Ef_{O_{j+1}}(x) + \mathrm{RapDec}(r)\|f\|_2 \qquad \text{for all } x \in O_{j+1}.$$

This inequality relies on the fact that $\rho_{j+1} \gtrsim r^\delta$, which is valid since it is assumed that the stopping condition [tiny] fails. If r is sufficiently large, then one concludes that

$$\|Ef\|_{\mathrm{BL}^p_{k,A}(B_r)}^p \leqslant Cd^{-\delta} C^{\mathrm{I}}_{j+1,\delta}(d,r) \sum_{O_{j+1} \in \mathcal{O}_{j+1}} \|Ef_{O_{j+1}}\|_{\mathrm{BL}^p_{k,A_{j+1}}(O_{j+1})}^p + \mathrm{err}(j+1).$$

Thus, provided d is chosen large enough so as to ensure that the additional $d^{-\delta}$ factor absorbs the unwanted constant C, one deduces $(\mathrm{I})_{j+1}$. This should be compared with Solymosi and Tao's approach to polynomial partitioning [30].

Property II. By the construction,

$$\sum_{O_{j+1} \in \mathcal{O}_{j+1}} \|f_{O_{j+1}}\|_2^2 = \sum_{O_j \in \mathcal{O}_{j,\mathrm{cell}}} \sum_{O_{j+1} \in \mathcal{O}_{j+1}(O_j)} \|f_{O_{j+1}}\|_2^2$$
$$\lesssim d \sum_{O_j \in \mathcal{O}_j} \|f_{O_j}\|_2^2,$$

where the inequality follows from a term-wise application of (33). Thus, $(\mathrm{II})_j$ and (35) imply that

$$\sum_{O_{j+1} \in \mathcal{O}_{j+1}} \|f_{O_{j+1}}\|_2^2 \lesssim d^{-\delta} C^{\mathrm{II}}_{j+1,\delta}(d) d^{\#\mathtt{c}(j+1)} \|f\|_2^2$$

and, provided d is chosen sufficiently large, one deduces $(\mathrm{II})_{j+1}$.

Property III. Fix $O_j \in \mathcal{O}_{j,\mathrm{cell}}$, $O_{j+1} \in \mathcal{O}_{j+1}(O_j)$ and recall from (34) that

$$\|f_{O_{j+1}}\|_2^2 \lesssim d^{-(m-1)} \|f_{O_j}\|_2^2. \tag{36}$$

Thus, $(\mathrm{III})_j$ and (35) imply that

$$\|f_{O_{j+1}}\|_2^2 \lesssim d^{-\delta} C^{\mathrm{III}}_{j+1,\delta}(d,r) \Big(\frac{r}{\rho_j}\Big)^{-\frac{n-m}{2}} d^{-(\#\mathtt{c}(j)+1)(m-1)} \|f\|_2^2.$$

Since $\rho_j \sim \rho_{j+1}$ and $\#\mathtt{c}(j)+1 = \#\mathtt{c}(j+1)$, provided d is chosen sufficiently large, one deduces $(\mathrm{III})_{j+1}$.

The local inequality $(\mathrm{III}_{\mathrm{loc}})_{j+1}$ follows in a similar manner but with one key difference: the inequality (36) is no longer available due to the localisation in the

L^2-norms. Instead, one uses simple orthogonality between the wave packets to prove that

$$\max_{\theta:\rho^{-1/2}-\mathrm{cap}} \|f_{O_{j+1}}\|^2_{L^2_{\mathrm{avg}}(\theta)} \lesssim \max_{\theta:\rho^{-1/2}-\mathrm{cap}} \|f_{O_j}\|^2_{L^2_{\mathrm{avg}}(\theta)}$$

for $1 \leqslant \rho \leqslant \rho_{j+1} \leqslant \rho_j$.

▶ **Algebraic-dominant case.** Suppose that the hypothesis (29) of the cellular-dominant case fails so that

$$\sum_{O_j \in \mathcal{O}_{j,\mathrm{cell}}} \|Ef_{O_j}\|^p_{\mathrm{BL}^p_{k,A_j}(O_j)} < \sum_{O_j \in \mathcal{O}_{j,\mathrm{alg}}} \|Ef_{O_j}\|^p_{\mathrm{BL}^p_{k,A_j}(O_j)} \qquad (37)$$

and, consequently,

$$\|Ef\|^p_{\mathrm{BL}^p_{k,A}(B_r)} \leqslant 2C^{\mathrm{I}}_{j,\delta}(d,r) \sum_{O_j \in \mathcal{O}_{j,\mathrm{alg}}} \|Ef_{O_j}\|^p_{\mathrm{BL}^p_{k,A_j}(O_j)} + \mathrm{err}(j). \qquad (38)$$

Each cell in $\mathcal{O}_{j,\mathrm{alg}}$ satisfies the condition of the algebraic case of Theorem 6.3; this information is used to construct the $(j+1)$-generation ensemble.

Definition of $\mathscr{E}_{j+1}$. Define $\mathfrak{h}_{j+1}$ by adjoining the letter $\mathtt{a}$ to the word $\mathfrak{h}_j$. Thus, it follows from the definitions that

$$\rho_{j+1} = \rho_j^{1-\tilde{\delta}_{m-1}}, \quad \#\mathtt{c}(j+1) = \#\mathtt{c}(j) \quad \text{and} \quad \#\mathtt{a}(j+1) = \#\mathtt{a}(j) + 1.$$

The next generation of cells is constructed from the varieties which arise from the algebraic case in Theorem 6.3. Fix $O_j \in \mathcal{O}_{j,\mathrm{alg}}$ so that there exists a transverse complete intersection $\mathbf{Y}$ of dimension $m-1$ and $\overline{\deg}\,\mathbf{Y} \leqslant C_{\mathrm{alg}}d$ such that

$$\|Ef_{O_j}\|^p_{\mathrm{BL}^p_{k,A_j}(O_j)} \lesssim \|Ef_{O_j}\|^p_{\mathrm{BL}^p_{k,A_j}(O_j \cap N_{\rho_j^{1/2+\delta_m}}\mathbf{Y})}.$$

Let $\mathcal{B}(O_j)$ be a cover of $O_j \cap N_{\rho_j^{1/2+\delta_m}}\mathbf{Y}$ by finitely-overlapping balls of radius ρ_{j+1}. For each $B \in \mathcal{B}(O_j)$ let $\mathbb{T}_B$ denote the collection of all $(\theta, v) \in \mathbb{T}[\rho_j]$ for which $T_{\theta,v} \cap B \cap N_{\rho_j^{1/2+\delta_m}}\mathbf{Y} \neq \emptyset$. This set is partitioned into subsets $\mathbb{T}_{B,\mathrm{tang}}$ and $\mathbb{T}_{B,\mathrm{trans}}$ consisting of wave packets in $\mathbb{T}_B$ which are tangential and transverse to $\mathbf{Y}$ on B, respectively, by defining

$$\mathbb{T}_{B,\mathrm{tang}} := \left\{ (\theta, v) \in \mathbb{T}_B \,:\, T_{\theta,v} \text{ is } \rho_{j+1}^{-1/2+\delta_{m-1}}\text{-tangent to } \mathbf{Y} \text{ on } B \right\}$$

and $\mathbb{T}_{B,\mathrm{trans}} := \mathbb{T}_B \setminus \mathbb{T}_{B,\mathrm{tang}}$. This setup is slightly inconsistent with the definition of tangent from Definition 8.1 (since the wave packets in $\mathbb{T}_B$ are at the large scale ρ_j rather than ρ_{j+1}) and therefore some clarification is necessary.

Definition 9.3. In this context, the tangency condition means that the following conditions hold:

  i) $T_{\theta,v} \cap 2B \subseteq N_{2\rho_{j+1}^{1/2+\delta_{m-1}}}\mathbf{Y} = N_{2\rho_j^{1/2+\delta_m}}\mathbf{Y}$;

  ii) If $x \in T_{\theta,v}$ and $y \in \mathbf{Y} \cap 2B$ satisfy $|y - x| \lesssim \rho_{j+1}^{1/2+\delta_{m-1}} = \rho_j^{1/2+\delta_m}$, then

$$\angle(G(\theta), T_y\mathbf{Y}) \lesssim \rho_{j+1}^{-1/2+\delta_{m-1}}.$$

By the basic concentration property of the wave packets, one may decompose the function Ef_{O_j} on B as

$$Ef_{O_j}(x) = Ef_{B,\mathrm{trans}}(x) + Ef_{B,\mathrm{tang}}(x) + \mathrm{RapDec}(r)\|f\|_2 \qquad \text{for all } x \in B$$

where

$$f_{B,\mathrm{tang}} := \sum_{(\theta,v) \in \mathbb{T}_{B,\mathrm{tang}}} (f_{O_j})_{\theta,v} \qquad \text{and} \qquad f_{B,\mathrm{trans}} := \sum_{(\theta,v) \in \mathbb{T}_{B,\mathrm{trans}}} (f_{O_j})_{\theta,v}.$$

The functions $f_{B,\mathrm{tang}}$ are in fact concentrated on scale ρ_{j+1} wave packets which are $\rho_{j+1}^{-1/2+\delta_{m-1}}$-tangent to $\mathbf{Y}$ in B in precisely the sense of Definition 8.1. This can be seen by a direct application of Lemma 7.2. In addition, by the local version of the basic orthogonality between wave packets,

$$\max_{\theta:\rho^{-1/2}-\mathrm{cap}} \|f_{B,\mathrm{tang}}\|^2_{L^2_{\mathrm{avg}}(\theta)} \lesssim \max_{\theta:\rho^{-1/2}-\mathrm{cap}} \|f_{O_j}\|^2_{L^2_{\mathrm{avg}}(\theta)}$$

whenever $B \in \mathcal{B}(O_j)$ and $1 \leqslant \rho \leqslant \rho_j$. Provided that the constant C_{tang} is suitably chosen, these observations imply that the functions $f_{B,\mathrm{tang}}$ satisfy the conditions stated in (27) from the definition of the stopping condition [tang].

By hypothesis, [tang] fails and, consequently, one may deduce that

$$\sum_{O_j \in \mathcal{O}_{j,\mathrm{alg}}} \|Ef_{O_j}\|^p_{\mathrm{BL}^p_{k,A_j}(O_j)} \lesssim \sum_{O_j \in \mathcal{O}_{j,\mathrm{alg}}} \sum_{B \in \mathcal{B}(O_j)} \|Ef_{B,\mathrm{trans}}\|^p_{\mathrm{BL}^p_{k,A_{j+1}}(B)}, \qquad (39)$$

where this inequality holds up to the inclusion of a rapidly decaying error term on the right-hand side. Indeed, by the triangle inequality for broad norms (Lemma 5.2) and since $A_{j+1} = A_j/2$, one may dominate the left-hand side of (39) by

$$\sum_{O_j \in \mathcal{O}_{j,\mathrm{alg}}} \sum_{B \in \mathcal{B}(O_j)} \left[\|Ef_{B,\mathrm{tang}}\|^p_{\mathrm{BL}^p_{k,A_{j+1}}(B)} + \|Ef_{B,\mathrm{trans}}\|^p_{\mathrm{BL}^p_{k,A_{j+1}}(B)} \right] + \mathrm{RapDec}(r)\|f\|^p_2.$$

By the preceding observations, the failure of the stopping condition [tang] forces

$$\sum_{O_j \in \mathcal{O}_{j,\mathrm{alg}}} \sum_{B \in \mathcal{B}(O_j)} \|Ef_{B,\mathrm{tang}}\|^p_{\mathrm{BL}^p_{k,A_{j+1}}(B)} < \frac{1}{C_{\mathrm{tang}}} \sum_{O_j \in \mathcal{O}_j} \|Ef_{O_j}\|^p_{\mathrm{BL}^p_{k,A_j}(O_j)}$$

(since it has been shown that all other conditions for [tang] are met). Recalling (37), for a suitable choice of constant C_{tang}, this implies (39).

The functions $f_{B,\mathrm{trans}}$ and sets B are further decomposed so as to ensure favourable tangency properties with respect to translates of the variety $\mathbf{Z}$ at the new scale ρ_{j+1}. Let $\mathbf{Z}_B := \mathbf{Z} + x_{O_j} - x_B$ where x_B denotes the centre of $B \in \mathcal{B}(O_j)$. Proposition 8.3 implies that for each $B \in \mathcal{B}(O_j)$ there exists a finite set of translates $\mathfrak{B} \subseteq B(0, \rho_j^{1/2+\delta_m})$ such that

$$\|Ef_{B,\mathrm{trans}}\|^p_{\mathrm{BL}^p_{k,A_{j+1}}(B)} \lesssim \log^2 r \sum_{B \in \mathfrak{B}} \|E\tilde{f}_{B,\mathrm{trans},b}\|^p_{\mathrm{BL}^p_{k,A_{j+1}}(B(0,\rho_{j+1}) \cap N_{\rho_{j+1}^{1/2+\delta_m}}(\mathbf{Z}_B + b))}$$
$$(40)$$

holds up to the inclusion of a rapidly decaying error term, whilst

$$\sum_{b \in \mathfrak{B}} \|\tilde{f}_{B,\mathrm{trans},b}\|^2_2 \lesssim \|f_{B,\mathrm{trans}}\|^2_2. \qquad (41)$$

Here the functions $\tilde{f}_{B,\mathrm{trans},b}$ are defined as in Section 7.2. In particular, each $\tilde{f}_{B,\mathrm{trans},b}$ is concentrated on wave packets which are $\rho_{j+1}^{-1/2+\delta_m}$-tangent to $\mathbf{Z}_B + b$ in $B(0, \rho_{j+1})$. Finally, define

$$\mathcal{O}_{j+1}(O_j) := \left\{ B \cap N_{\rho_{j+1}^{1/2+\delta_m}}(\mathbf{Z} + x_{O_j} + b) \, : \, B \in \mathcal{B}(O_j) \text{ and } b \in \mathfrak{B} \right\}$$

and for any $O_{j+1} = B \cap N_{\rho_{j+1}^{1/2+\delta_m}}(\mathbf{Z} + x_{O_j} + b) \in \mathcal{O}_{j+1}(O_j)$ let $f_{O_{j+1}}$ satisfy $\tilde{f}_{O_{j+1}} := \tilde{f}_{B,\mathrm{trans},b}$; once again, we are using the definition of the map $g \mapsto \tilde{g}$ from Section 7.2. Thus, each $f_{O_{j+1}}$ is concentrated on wave packets which are $\rho_{j+1}^{-1/2+\delta_m}$-tangent to $\mathbf{Z} + x_{O_{j+1}}$ in $B \in \mathcal{B}(O_j)$, where $x_{O_{j+1}} = x_{O_j} + b$. The collection of cells $\mathcal{O}_{j+1}$ is then given by

$$\mathcal{O}_{j+1} := \bigcup_{O_j \in \mathcal{O}_{j,\mathrm{alg}}} \mathcal{O}_{j+1}(O_j).$$

It remains to verify that the ensemble $\mathscr{E}_{j+1}$ satisfies the desired properties. In view of this, it is useful to note that

$$
\begin{aligned}
C_{j,\delta}^{\mathrm{I}}(d,r) &= (\log r)^{-2(1+\delta)} C_{j+1,\delta}^{\mathrm{I}}(d,r), \\
C_{j,\delta}^{\mathrm{II}}(d) &= d^{-n(1+\delta)} C_{j+1,\delta}^{\mathrm{II}}(d), \\
C_{j,\delta}^{\mathrm{III}}(d,r) &= r^{-\bar{C}\delta_m} d^{-\delta} C_{j+1,\delta}^{\mathrm{III}}(d,r),
\end{aligned}
\tag{42}
$$

which can be verified directly from the definitions.

Property I. By combining (39) and (40) together with the various definitions one obtains

$$
\sum_{O_j \in \mathcal{O}_{j,\mathrm{alg}}} \|Ef_{O_j}\|_{\mathrm{BL}_{k,A_j}^p(O_j)}^p \lesssim \log^2 r \sum_{O_{j+1} \in \mathcal{O}_{j+1}} \|Ef_{O_{j+1}}\|_{\mathrm{BL}_{k,A_{j+1}}^p(O_{j+1})}^p,
$$

where this inequality holds up to the inclusion of a rapidly decaying error term on the right-hand side. Recalling (38) and (42), it follows that

$$
\|Ef\|_{\mathrm{BL}_{k,A}^p(B_r)}^p \leqslant \frac{C \cdot C_{j+1,\delta}^{\mathrm{I}}(d,r)}{(\log r)^{2\delta}} \sum_{O_{j+1} \in \mathcal{O}_{j+1}} \|Ef_{O_{j+1}}\|_{\mathrm{BL}_{k,A_{j+1}}^p(O_{j+1})}^p + \mathrm{err}(j+1).
$$

Provided r is chosen to be sufficiently large, one may absorb the unwanted constant C by the additional $(\log r)^{-2\delta}$ factor and thereby deduce $(\mathrm{I})_{j+1}$.

Property II. Fix $O_j \in \mathcal{O}_{j,\mathrm{alg}}$ and note that

$$
\begin{aligned}
\sum_{O_{j+1} \in \mathcal{O}_{j+1}(O_j)} \|f_{O_{j+1}}\|_2^2 &= \sum_{B \in \mathcal{B}(O_j)} \sum_{b \in \mathfrak{B}} \|f_{B,\mathrm{trans},b}\|_2^2 \\
&\lesssim \sum_{B \in \mathcal{B}(O_j)} \|f_{B,\mathrm{trans}}\|_2^2
\end{aligned}
\tag{43}
$$

by the definition of $f_{O_{j+1}}$ and (41). To estimate the latter sum one exploits the transversal property of the wave packets of the $f_{B,\mathrm{trans}}$. The key observation is the following algebraic-geometric result of Guth, which appears in Lemma 5.7 of [12] and can be roughly thought of as a continuum version of the fundamental theorem of algebra (or Bézout's theorem).

Lemma 9.4 ([12]). *Suppose T is an infinite cylinder in $\mathbb{R}^n$ of radius ρ and central axis ℓ and $\mathbf{Y}$ is a transverse complete intersection. For $\alpha > 0$ let*

$$
\mathbf{Y}_{>\alpha} := \{y \in \mathbf{Y} : \angle(T_y\mathbf{Y}, \ell) > \alpha\}.
$$

The set $\mathbf{Y}_{>\alpha} \cap T$ is contained in a union of $O\big((\overline{\deg}\,\mathbf{Y})^n\big)$ balls of radius $\rho\alpha^{-1}$.

By choosing the implicit constants correctly in Definition 9.3, a wave packet $(\theta,v) \in \mathbb{T}_B$ belongs to $\mathbb{T}_{B,\mathrm{trans}}$ if and only if the angle condition ii) fails to be satisfied. Indeed, if ii) holds, then since $T_{\theta,v} \cap B \cap N_{\rho_j^{1/2+\delta_m}}\mathbf{Y} \neq \emptyset$ by the definition of $\mathbb{T}_B$, the containment property i) automatically follows and therefore $(\theta,v) \in \mathbb{T}_{B,\mathrm{tang}}$ (see, for instance, [13, Proposition 9.2] for details of an argument of this type). Thus, given any $(\theta,v) \in \bigcup_{B \in \mathcal{B}(O_j)} \mathbb{T}_{B,\mathrm{trans}}$, it follows from the definitions that

$$
\angle(G(\theta), T_y\mathbf{Y}) \gtrsim \rho_{j+1}^{-1/2+\delta_{m-1}}
\tag{44}
$$

for some $y \in \mathbf{Y} \cap 2B$ with $|y-x| \lesssim \rho_{j+1}^{1/2+\delta_{m-1}}$ for some $x \in T_{\theta,v}$. This implies that

$$
T \cap B \cap \mathbf{Y}_{\gtrsim \rho_{j+1}^{-1/2+\delta_{m-1}}} \neq \emptyset
$$

where T is the infinite cylinder that shares the core line of $T_{\theta,v}$ but has radius $\sim \rho_{j+1}^{1/2+\delta_{m-1}}$. Observe that

$$\underbrace{\rho_{j+1}^{1/2+\delta_{m-1}}}_{\substack{\sim\text{Radius} \\ \text{of } T}} \big[\underbrace{\rho_{j+1}^{-1/2+\delta_{m-1}}}_{\substack{\sim\text{Angle} \\ \text{from } (44)}}\big]^{-1} = \underbrace{\rho_{j+1}}_{\substack{\text{Radius} \\ \text{of } B \in \mathcal{B}(O_j)}} .$$

Thus, by Lemma 9.4, any $(\theta,v) \in \bigcup_{B \in \mathcal{B}(O_j)} \mathbb{T}_{B,\text{trans}}$ lies in at most $O(d^n)$ of the sets $\mathbb{T}_{B,\text{trans}}$ and, consequently, by the basic orthogonality between the wave packets,

$$\sum_{B \in \mathcal{B}(O_j)} \|f_{B,\text{trans}}\|_2^2 \lesssim d^n \|f_{O_j}\|_2^2.$$

Combining this inequality with (43) and summing over all $O_j \in \mathcal{O}_{j,\text{alg}}$,

$$\sum_{O_{j+1} \in \mathcal{O}_{j+1}} \|f_{O_{j+1}}\|_2^2 \lesssim d^n \sum_{O_j \in \mathcal{O}_j} \|f_{O_j}\|_2^2.$$

Applying $(\text{II})_j$ and (42), one concludes that

$$\sum_{O_{j+1} \in \mathcal{O}_{j+1}} \|f_{O_{j+1}}\|_2^2 \lesssim d^{-n\delta} C_{j+1,\delta}^{\text{II}}(d,r) \|f\|_2^2.$$

Thus, provided d is chosen sufficiently large, one deduces $(\text{II})_{j+1}$.

<u>Property III</u>. Fix $O_j \in \mathcal{O}_{j,\text{alg}}$ and $O_{j+1} \in \mathcal{O}_{j+1}(O_j)$ and suppose that $\tilde{f}_{O_{j+1}} = \tilde{f}_{B,\text{trans},b}$. Recall that the function $f_{B,\text{trans}}$ is concentrated on scale ρ_j wave packets which are $\rho_j^{-1/2+\delta_m}$-tangent to some translate of $\mathbf{Z}$ on some ρ_j-ball. It therefore follows from the transverse equidistribution estimate (22) of Lemma 8.4 with $\rho := 1$ that

$$\|\tilde{f}_{B,\text{trans},b}\|_2^2 \lesssim_{\overline{\deg}\,\mathbf{Z}} r^{\bar{C}\delta_m} \Big(\frac{\rho_j}{\rho_{j-1}}\Big)^{-\frac{n-m}{2}} \|f_{B,\text{trans}}\|_2^2.$$

On the other hand, by the basic orthogonality between the wave packets,

$$\|f_{B,\text{trans}}\|_2^2 \lesssim \|f_{O_j}\|_2^2.$$

Applying $(\text{III})_j$ and (42), one concludes that

$$\|f_{O_{j+1}}\|_2^2 \lesssim_{\overline{\deg}\,\mathbf{Z}} d^{-\delta} C_{j+1,\delta}^{\text{III}}(d,r) \Big(\frac{r}{\rho_{j+1}}\Big)^{-\frac{n-m}{2}} d^{-\#\mathsf{c}(j+1)(m-1)} \|f\|_2^2.$$

Thus, provided d is chosen sufficiently large, one deduces $(\text{III})_{j+1}$. The local version, $(\text{III}_{\text{loc}})_{j+1}$, follows in a similar manner, using the local transverse equidistribution estimate (22) for general values of $1 \leqslant \rho \leqslant \rho_{j+1}$.

10. Proof of Theorem 2.1

10.1. **The second algorithm.** Theorem 2.1 is established by repeated application of the algorithm [alg 1] from the previous section. This process forms part of a second algorithm which is referred to as [alg 2] and is described presently.

Throughout this section, let p_ℓ denote Lebesgue exponents, to be fixed later, defined for $k \leqslant \ell \leqslant n$ and satisfying

$$p_k \geqslant p_{k+1} \geqslant \ldots \geqslant p_n =: p \geqslant 2.$$

The numbers $0 \leqslant \alpha_\ell, \beta_\ell \leqslant 1$ for $k \leqslant \ell \leqslant n$ are then defined in terms of the p_ℓ by

$$\frac{1}{p_\ell} =: \frac{1-\alpha_{\ell-1}}{2} + \frac{\alpha_{\ell-1}}{p_{\ell-1}} \quad \text{and} \quad \beta_\ell := \prod_{i=\ell}^{n-1} \alpha_i \qquad \text{for } k+1 \leqslant \ell \leqslant n-1$$

and $\alpha_n :=: \beta_n := 1$. Also fix $\varepsilon > 0$ and define the small parameters δ_ℓ as in the previous section so that the inequalities in (23) hold.

There are two stages to [alg 2], which can roughly be described as follows:

- **The recursive stage**: Ef is repeatedly decomposed into pieces with favourable tangency properties with respect to varieties of progressively lower dimension.
- **The final stage**: Ef is further decomposed into very small scale pieces.

To begin, the recursive stage of [alg 2] is described.

<u>Input</u>. Fix $R \gg 1$ and let $f \colon B^{n-1} \to \mathbb{C}$ be smooth and bounded and, without loss of generality, assume that f satisfies the *non-degeneracy hypothesis*

$$\|Ef\|_{\mathrm{BL}^p_{k,A}(B_R)} \geqslant C_{\mathrm{hyp}} R^\varepsilon \|f\|_2 \tag{45}$$

where C_{hyp} and $A \in \mathbb{N}$ are constants which are chosen sufficiently large to satisfy the forthcoming requirements of the proof.

<u>Output</u>. The $(n + 1 - \ell)$th step of the recursion will produce:

- An $(n + 1 - \ell)$-tuple of:
    - scales $\vec{r}_\ell = (r_n, \ldots, r_\ell)$ satisfying $R = r_n > r_{n-1} > \cdots > r_\ell$;
    - large and (in general) non-admissible parameters $\vec{D}_\ell = (D_n, \ldots, D_\ell)$;
    - integers $\vec{A} = (A_n, \ldots, A_\ell)$ satisfying $A = A_n > A_{n-1} > \cdots > A_\ell$.

  Each of these $(n + 1 - \ell)$-tuples is formed by adjoining a component to the corresponding $(n - \ell)$-tuple from the previous stage.
- A family $\vec{\mathcal{S}}_\ell$ of $(n + 1 - \ell)$-tuples of transverse complete intersections $\vec{S}_\ell = (S_n, \ldots, S_\ell)$ satisfying $\dim S_i = i$ and $\overline{\deg} S_i = O(1)$ for $\ell \leqslant i \leqslant n$.
- An assignment of a function $f_{\vec{S}_\ell}$ and a ball $B_{r_\ell}[\vec{S}_\ell]$ to each $\vec{S}_\ell \in \vec{\mathcal{S}}_\ell$ with the property that $f_{\vec{S}_\ell}$ is concentrated on scale r_ℓ wave packets which are $r_\ell^{-1/2+\delta_\ell}$-tangent to S_ℓ in $B_{r_\ell}[\vec{S}_\ell]$ (here S_ℓ is the final component of $\vec{S}_\ell$). For notational convenience, the dependence on $\vec{S}_\ell$ will often be suppressed in the $B_{r_\ell}[\vec{S}_\ell]$ notation by simply writing B_{r_ℓ}.

This data is chosen so that the following properties hold:

Notation. Throughout this section a large number of harmless $R^{C\delta_0}$ factors appear in the inequalities, where C is a constant depending on p and n. By choosing δ_0 sufficiently small relative to ε, at the end of the argument one may dominate any $R^{C\delta_0}$ by R^ε, say, which constitutes an acceptable loss in the inequality. Thus, for notational convenience, given $A, B \geqslant 0$ let $A \lesssim B$ or $B \gtrsim A$ denote $A \lesssim R^{C\delta_0} B$.

<u>Property 1</u>. The inequality

$$\|Ef\|_{\mathrm{BL}^p_{k,A}(B_R)} \lesssim M(\vec{r}_\ell, \vec{D}_\ell) \|f\|_2^{1-\beta_\ell} \Big(\sum_{\vec{S}_\ell \in \vec{\mathcal{S}}_\ell} \|Ef_{\vec{S}_\ell}\|_{\mathrm{BL}^{p_\ell}_{k,A_\ell}(B_{r_\ell})}^{p_\ell} \Big)^{\frac{\beta_\ell}{p_\ell}} \tag{46}$$

holds for

$$M(\vec{r}_\ell, \vec{D}_\ell) := \Big(\prod_{i=\ell}^{n-1} D_i \Big)^{(n-\ell)\delta} \Big(\prod_{i=\ell}^{n-1} r_i^{\frac{1}{2}(\beta_{i+1}-\beta_i)} D_i^{\frac{1}{2}(\beta_{i+1}-\beta_\ell)} \Big).$$

<u>Property 2</u>. For $\ell \leqslant n - 1$ the inequality

$$\sum_{\vec{S}_\ell \in \vec{\mathcal{S}}_\ell} \|f_{\vec{S}_\ell}\|_2^2 \lesssim D_\ell^{1+\delta} \sum_{\vec{S}_{\ell+1} \in \vec{\mathcal{S}}_{\ell+1}} \|f_{\vec{S}_{\ell+1}}\|_2^2$$

holds.

Property 3. For $\ell \leqslant n-1$ the inequalities

$$\max_{\vec{S}_\ell \in \vec{\mathcal{S}}_\ell} \|f_{\vec{S}_\ell}\|_2^2 \lessapprox \left(\frac{r_{\ell+1}}{r_\ell}\right)^{-\frac{n-\ell-1}{2}} D_\ell^{-\ell+\delta} \max_{\vec{S}_{\ell+1} \in \vec{\mathcal{S}}_{\ell+1}} \|f_{\vec{S}_{\ell+1}}\|_2^2$$

and

$$\max_{\substack{\vec{S}_\ell \in \vec{\mathcal{S}}_\ell \\ \theta:\rho^{-1/2}-\mathrm{cap}}} \|f_{\vec{S}_\ell}\|_{L^2_{\mathrm{avg}}(\theta)}^2 \lessapprox \left(\frac{r_{\ell+1}}{r_\ell}\right)^{-\frac{n-\ell-1}{2}} D_\ell^\delta \max_{\substack{\vec{S}_{\ell+1} \in \vec{\mathcal{S}}_{\ell+1} \\ \theta:\rho^{-1/2}-\mathrm{cap}}} \|f_{\vec{S}_{\ell+1}}\|_{L^2_{\mathrm{avg}}(\theta)}^2$$

hold for $1 \leqslant \rho \leqslant r_\ell$.

First step. Vacuously, the function f is concentrated on scale R wave packets which are $R^{-1/2+\delta_n}$-tangent to the n-dimensional variety $\mathbb{R}^n$ on B_R. Thus, one may define

- $r_n := R$; $D_n := 1$ and $A_n := A$.
- $\mathcal{S}_n$ is the collection consisting of a single 1-tuple $\vec{S}_n = (S_n)$ where $S_n := \mathbb{R}^n$.
- $f_{\vec{S}_n} := f$ and $B_{r_n}[\vec{S}_n] := B_R$.

With these definitions, all the desired properties vacuously hold.

$(n+2-\ell)$th step. Let $\ell \geqslant 1$ and suppose that the recursive algorithm has ran through $n+1-\ell$ steps. Since each function $f_{\vec{S}_\ell}$ is concentrated on wave packets $r_\ell^{-1/2+\delta_\ell}$-tangent to S_ℓ on $B_{r_\ell}[S_\ell]$, one may apply [alg 1] to bound the k-broad norm $\|Ef_{\vec{S}_\ell}\|_{\mathrm{BL}^{p_\ell}_{k,A_\ell}(B_{r_\ell})}$. One of two things can happen: either [alg 1] terminates due to the stopping condition [tiny] or it terminates due to the stopping condition [tang]. The current recursive process terminates if the contributions from terms of the former type dominate:

Stopping condition. The recursive stage of [alg 2] has a single stopping condition, which is denoted by [tiny-dom].

Stop:[tiny-dom] Suppose that the inequality

$$\sum_{\vec{S}_\ell \in \vec{\mathcal{S}}_\ell} \|Ef_{\vec{S}_\ell}\|_{\mathrm{BL}^{p_\ell}_{k,A_\ell}(B_{r_\ell})}^{p_\ell} \leqslant 2 \sum_{\vec{S}_\ell \in \vec{\mathcal{S}}_{\ell,\mathrm{tiny}}} \|Ef_{\vec{S}_\ell}\|_{\mathrm{BL}^{p_\ell}_{k,A_\ell}(B_{r_\ell})}^{p_\ell} \tag{47}$$

holds, where the right-hand summation is restricted to those $S_\ell \in \vec{\mathcal{S}}_\ell$ for which [alg 1] terminates owing to the stopping condition [tiny]. Then [alg 2] terminates.

Assume that the condition [tiny-dom] is not met. Necessarily,

$$\sum_{\vec{S}_\ell \in \vec{\mathcal{S}}_\ell} \|Ef_{\vec{S}_\ell}\|_{\mathrm{BL}^{p_\ell}_{k,A_\ell}(B_{r_\ell})}^{p_\ell} \leqslant 2 \sum_{\vec{S}_\ell \in \vec{\mathcal{S}}_{\ell,\mathrm{tang}}} \|Ef_{\vec{S}_\ell}\|_{\mathrm{BL}^{p_\ell}_{k,A_\ell}(B_{r_\ell})}^{p_\ell}, \tag{48}$$

where the right-hand summation is restricted to those $S_\ell \in \vec{\mathcal{S}}_\ell$ for which [alg 1] does not terminate owing to [tiny] and therefore terminates owing to [tang]. Consequently, for each $\vec{S}_\ell \in \vec{\mathcal{S}}_{\ell,\mathrm{tang}}$ the inequalities

$$\|Ef_{\vec{S}_\ell}\|_{\mathrm{BL}^{p_\ell}_{k,A_\ell}(B_{r_\ell})}^{p_\ell} \lessapprox D_{\ell-1}^\delta \sum_{S_{\ell-1} \in \mathcal{S}_{\ell-1}[\vec{S}_\ell]} \|Ef_{\vec{S}_{\ell-1}}\|_{\mathrm{BL}^{p_\ell}_{k,2A_{\ell-1}}(B_{r_{\ell-1}})}^{p_\ell}, \tag{49}$$

and

$$\sum_{S_{\ell-1}\in\mathcal{S}_{\ell-1}[\vec{S}_\ell]}\|f_{\vec{S}_{\ell-1}}\|_2^2 \lesssim D_{\ell-1}^{1+\delta}\|f_{\vec{S}_\ell}\|_2^2;\tag{50}$$

$$\max_{S_{\ell-1}\in\mathcal{S}_{\ell-1}[\vec{S}_\ell]}\|f_{\vec{S}_{\ell-1}}\|_2^2 \lesssim \Big(\frac{r_\ell}{r_{\ell-1}}\Big)^{-\frac{n-\ell}{2}}D_{\ell-1}^{-(\ell-1)+\delta}\|f_{\vec{S}_\ell}\|_2^2;\tag{51}$$

$$\max_{\substack{S_{\ell-1}\in\mathcal{S}_{\ell-1}[\vec{S}_\ell]\\ \theta:\rho^{-1/2}-\text{cap}}}\|f_{\vec{S}_{\ell-1}}\|_{L^2_{\text{avg}}(\theta_\ell)}^2 \lesssim \Big(\frac{r_\ell}{r_{\ell-1}}\Big)^{-\frac{n-\ell}{2}}D_{\ell-1}^{\delta}\max_{\theta:\rho^{-1/2}-\text{cap}}\|f_{\vec{S}_\ell}\|_{L^2_{\text{avg}}(\theta)}^2\tag{52}$$

hold for $1\leqslant\rho\leqslant r_{\ell-1}$ for some choice of:

- Scale $R^\delta < r_{\ell-1} < r_\ell$, an (in general) non-admissible number $D_{\ell-1}$ and some large integer $A_{\ell-1}$ satisfying $A_{\ell-1}\sim A_\ell$;
- Family $\mathcal{S}_{\ell-1}[\vec{S}_{\ell-1}]$ of $(\ell-1)$-dimensional transverse complete intersections of degree $O(1)$;
- Assignment $B_{r_{\ell-1}}[\vec{S}_{\ell-1}]$ of an $r_{\ell-1}$-ball to every $S_{\ell-1}\in\mathcal{S}_{\ell-1}[\vec{S}_\ell]$;
- Assignment $f_{\vec{S}_{\ell-1}}=(f_{\vec{S}_\ell})_{S_{\ell-1}}$ of a function to every $S_{\ell-1}\in\mathcal{S}_{\ell-1}[\vec{S}_\ell]$ which is concentrated on wave packets which are $r_{\ell-1}^{-1/2+\delta_{\ell-1}}$-tangent to $S_{\ell-1}$ on $B_{r_{\ell-1}}[S_{\ell-1}]$.

Each inequality (49), (50), (51) and (52) is obtained by combining the definition of the stopping condition `[tang]` with Properties I, II and both the global and local variants of Property III from `[alg 1]`, respectively.[13] Indeed, using the notation from `[alg 1]`, we take

$$r:=r_\ell, \quad r_{\ell-1}:=\rho_J^{1-\tilde{\delta}_{\ell-1}}, \quad \text{and} \quad D_{\ell-1}:=d^{\#c(J)}.$$

Note that the $R^{O(\delta_0)}D_{\ell-1}^\delta$ factors arise in the above inequalities owing to (26).

The $r_{\ell-1}$, $D_{\ell-1}$ and $A_{\ell-1}$ can depend on the choice of $\vec{S}_\ell$, but this dependence can be essentially removed by pigeonholing. Indeed, recalling that $\#_c(J)=O(\log R)$, one may find a subset of the $\mathcal{S}_{\ell,\text{tang}}$ over which the $D_{\ell-1}$ all have a common value and, moreover, the inequality (47) still holds except that the constant $\frac{1}{2}$ is now replaced with, say, R^{δ_0}. A brief inspection of `[alg 1]` shows that, once we have pigeonholed in the parameter N above, both $r_{\ell-1}$ and $A_{\ell-1}$ immediately inherit the desired uniformity.

Letting $\vec{\mathcal{S}}_{\ell-1}$ denote the structured set

$$\vec{\mathcal{S}}_{\ell-1}:=\Big\{ (\vec{S}_\ell,S_{\ell-1}) : \vec{S}_\ell\in\vec{\mathcal{S}}_{\ell,\text{tang}} \text{ and } S_{\ell-1}\in\mathcal{S}_{\ell-1}[\vec{S}_\ell] \Big\},$$

where $\vec{\mathcal{S}}_{\ell,\text{tang}}$ is understood to be the refined collection described in the previous paragraph, it remains to verify the desired properties for the newly constructed data. Property 2 follows immediately from (50) and Property 3 from (51) and (52), so it remains only to verify Property 1.

By combining the inequality (46) from the previous stage of the algorithm with (48) and (49), one deduces that

$$\|Ef\|_{\text{BL}_{k,A}^p(B_R)}\lesssim D_{\ell-1}^\delta M(\vec{r}_\ell,\vec{D}_\ell)\|f\|_2^{1-\beta_\ell}\Big(\sum_{\vec{S}_{\ell-1}\in\vec{\mathcal{S}}_{\ell-1}}\|Ef_{\vec{S}_{\ell-1}}\|_{\text{BL}_{k,2A_{\ell-1}}^{p_\ell}(B_{r_{\ell-1}})}^{p_\ell}\Big)^{\frac{\beta_\ell}{p_\ell}}.$$

---

[13]Here the 'error terms' $\text{err}(j):=jr^{-N}\|f\|_2^p$ in Property I of `[alg 1]` can be ignored owing to the non-degeneracy hypothesis (45).

Writing

$$\Big(\sum_{\vec{S}_{\ell-1} \in \vec{\mathcal{S}}_{\ell-1}} \|Ef_{\vec{S}_{\ell-1}}\|_{\mathrm{BL}^{p_\ell}_{k,2A_{\ell-1}}(B_{r_{\ell-1}})}^{p_\ell} \Big)^{\frac{1}{p_\ell}} = \Big\| \|Ef_{\vec{S}_{\ell-1}}\|_{\mathrm{BL}^{p_\ell}_{k,2A_{\ell-1}}(B_{r_{\ell-1}})} \Big\|_{\ell^{p_\ell}(\vec{\mathcal{S}}_{\ell-1})},$$

one may apply the logarithmic convexity inequality from Lemma 5.3 to dominate this expression by

$$\Big\| \|Ef_{\vec{S}_{\ell-1}}\|_{\mathrm{BL}^2_{k,A_{\ell-1}}(B_{r_{\ell-1}})} \Big\|_{\ell^2(\vec{\mathcal{S}}_{\ell-1})}^{1-\alpha_{\ell-1}} \Big\| \|Ef_{\vec{S}_{\ell-1}}\|_{\mathrm{BL}^{p_{\ell-1}}_{k,A_{\ell-1}}(B_{r_{\ell-1}})} \Big\|_{\ell^{p_{\ell-1}}(\vec{\mathcal{S}}_{\ell-1})}^{\alpha_{\ell-1}}.$$

By the standard L^2 estimate (12) applied to broad norms,

$$\Big\| \|Ef_{\vec{S}_{\ell-1}}\|_{\mathrm{BL}^2_{k,A_{\ell-1}}(B_{r_{\ell-1}})} \Big\|_{\ell^2(\vec{\mathcal{S}}_{\ell-1})} \lesssim r_{\ell-1}^{1/2} \Big(\sum_{\vec{S}_{\ell-1} \in \vec{\mathcal{S}}_{\ell-1}} \|f_{\vec{S}_{\ell-1}}\|_2^2 \Big)^{1/2}$$

and, by Property 2 for the the families of functions $(f_{\vec{S}_i})_{\vec{S}_i \in \vec{\mathcal{S}}_i}$ for $\ell-1 \leqslant i \leqslant n-1$, it follows that

$$\Big\| \|Ef_{\vec{S}_{\ell-1}}\|_{\mathrm{BL}^2_{k,A_{\ell-1}}(B_{r_{\ell-1}})} \Big\|_{\ell^2(\vec{\mathcal{S}}_{\ell-1})} \lessapprox \Big(r_{\ell-1} \prod_{i=\ell-1}^{n-1} D_i^{1+\delta} \Big)^{1/2} \|f\|_2.$$

One may readily verify that

$$D_{\ell-1}^\delta \cdot M(\vec{r}_\ell, \vec{D}_\ell) \cdot \Big(r_{\ell-1} \prod_{i=\ell-1}^{n-1} D_i^{1+\delta} \Big)^{\frac{1}{2}(1-\alpha_{\ell-1})\beta_\ell} \leqslant M(\vec{r}_{\ell-1}, \vec{D}_{\ell-1})$$

and so combining the above estimates yields

$$\|Ef\|_{\mathrm{BL}^p_{k,A}(B_R)} \lessapprox M(\vec{r}_{\ell-1}, \vec{D}_{\ell-1}) \|f\|_2^{1-\beta_{\ell-1}} \Big\| \|Ef_{\vec{S}_{\ell-1}}\|_{\mathrm{BL}^{p_{\ell-1}}_{k,A_{\ell-1}}(B_{r_{\ell-1}})} \Big\|_{\ell^{p_{\ell-1}}(\vec{\mathcal{S}}_{\ell-1})}^{\beta_{\ell-1}},$$

which is Property 1 in this case.

The final stage. If the algorithm has not stopped by the kth step, then it necessarily terminates at the kth step. Indeed, otherwise (46) would hold for $\ell = k-1$ and functions $f_{\vec{S}_{k-1}}$ concentrated on wave packets which are tangent to some transverse complete intersection of dimension $k-1$. By the vanishing property of the k-broad norms as described in Lemma 8.2, one would then have

$$\|Ef_{\vec{S}_{k-1}}\|_{\mathrm{BL}^{p_{k-1}}_{k,A_{k-1}}(B_{r_{k-1}})} = \mathrm{RapDec}(R) \|f_{\vec{S}_{k-1}}\|_2$$

and it would easily follow from (46) that $\|Ef\|_{\mathrm{BL}^p_{k,A}(B_R)} = \mathrm{RapDec}(R)\|f\|_2$. If R is sufficiently large, then this would contradict the non-degeneracy hypothesis (45).

Suppose the recursive process terminates at step m, so that $m \geqslant k$. For each $\vec{S}_m \in \vec{\mathcal{S}}_{m,\mathtt{tiny}}$ let $\mathcal{O}[\vec{S}_m]$ denote the final collection of cells output by [alg 1] (that is, the collection denoted by $\mathcal{O}_J$ in the notation of Section 9) when applied to estimate the broad norm $\|Ef_{\vec{S}_m}\|_{\mathrm{BL}^{p_m}_{k,A_m}(B_{r_m})}$. Each $O \in \mathcal{O}[\vec{S}_m]$ has diameter at most R^{δ_0} by the definition of the stopping condition [tiny]. By Properties I, II and III of [alg 1] one has

$$\|Ef_{\vec{S}_m}\|_{\mathrm{BL}^{p_m}_{k,A_m}(B_{r_m})}^{p_m} \lessapprox D_{m-1}^\delta \sum_{O \in \mathcal{O}[\vec{S}_m]} \|Ef_O\|_{\mathrm{BL}^{p_m}_{k,A_{m-1}}(O)}^{p_m},$$

for some $A_{m-1} \sim A_m$ where the functions f_O satisfy

$$\sum_{O \in \mathcal{O}[\vec{S}_m]} \|f_O\|_2^2 \lessapprox D_{m-1}^{1+\delta} \|f_{\vec{S}_m}\|_2^2 \tag{53}$$

and

$$\max_{O \in \mathcal{O}[\vec{S}_m]} \|f_O\|_2^2 \lessapprox \left(\frac{r_m}{r_{m-1}}\right)^{-\frac{n-m}{2}} D_{m-1}^{-(m-1)+\delta} \|f_{\vec{S}_m}\|_2^2 \tag{54}$$

for D_{m-1} a large non-admissible parameter. In particular, $D_{m-1} := d^{\#\mathbf{c}(J)}$ where J is the stopping time for this final application of [alg 1]. Once again, by pigeon-holing, one may pass to a subcollection of $\mathcal{S}_{m,\mathtt{tiny}}$ and thereby assume that the D_{m-1} (and also the A_{m-1}) all share a common value.

If $\mathcal{O}$ denotes the union of the $\mathcal{O}[\vec{S}_m]$ over all $\vec{S}_m$ belonging to subcollection of $\mathcal{S}_{m,\mathtt{tiny}}$ described above, then

$$\|Ef\|_{\mathrm{BL}_{k,A}^p(B_R)} \lessapprox D_{m-1}^{\delta} M(\vec{r}_m, \vec{D}_m) \|f\|_2^{1-\beta_m} \left(\sum_{O \in \mathcal{O}} \|Ef_O\|_{\mathrm{BL}_{k,A_{m-1}}^{p_m}(O)}^{p_m}\right)^{\frac{\beta_m}{p_m}}. \tag{55}$$

This concludes the description of [alg 2].

10.2. **Applying [alg 2] to prove k-broad estimates.** Having arrived at the final decomposition of the broad norm given by (55), the task is now to apply the properties guaranteed by the algorithm in order to prove the desired estimates. In particular, one wishes to show that the quantity

$$M(\vec{r}_m, \vec{D}_m) \left(\sum_{O \in \mathcal{O}} \|Ef_O\|_{\mathrm{BL}_{k,A_{m-1}}^{p_m}(O)}^{p_m}\right)^{\frac{\beta_m}{p_m}}$$

featured in (55) can be effectively bounded, provided that the exponents $p_k, \ldots, p_n$ are suitably chosen. Since each $O \in \mathcal{O}$ has diameter at most R^{δ_0}, trivially one may bound

$$\|Ef_O\|_{\mathrm{BL}_{k,A_{m-1}}^{p_m}(O)} \lessapprox \|f_O\|_2$$

and, thus, it follows that

$$\left(\sum_{O \in \mathcal{O}} \|Ef_O\|_{\mathrm{BL}_{k,A_{m-1}}^{p_m}(O)}^{p_m}\right)^{\frac{\beta_m}{p_m}} \lessapprox \left(\sum_{O \in \mathcal{O}} \|f_O\|_2^2\right)^{\frac{\beta_m}{p_m}} \max_{O \in \mathcal{O}} \|f_O\|_2^{2(\frac{1}{2}-\frac{1}{p_m})\beta_m}.$$

The definition of the β_m ensures that

$$\left(\frac{1}{2} - \frac{1}{p_m}\right)\beta_m = \frac{1}{2} - \frac{1}{p_n}$$

whilst (53) and repeated application of Property 2 from [alg 2] imply that

$$\sum_{O \in \mathcal{O}} \|f_O\|_2^2 \lessapprox \left(\prod_{i=m-1}^{n-1} D_i^{1+\delta}\right) \|f\|_2^2.$$

Combining this estimate with (55) and the definition of $M(\vec{r}_m, \vec{D}_m)$, one concludes that

$$\|Ef\|_{\mathrm{BL}_{k,A}^p(B_R)} \lessapprox \prod_{i=m-1}^{n-1} r_i^{\frac{\beta_{i+1}-\beta_i}{2}} D_i^{\frac{\beta_{i+1}}{2}-(\frac{1}{2}-\frac{1}{p_n})+O(\delta)} \|f\|_2^{\frac{2}{p_n}} \max_{O \in \mathcal{O}} \|f_O\|_2^{1-\frac{2}{p_n}} \tag{56}$$

where $r_{m-1} := 1$. The problem is now to bound the maximum appearing on the right-hand side of this expression.

By (54) and repeated application of Property 3 of `[alg 2]`, it follows that for any $m \leqslant \ell \leqslant n$ the inequality

$$\max_{O \in \mathcal{O}} \|f_O\|_2^2 \lesssim \prod_{i=m-1}^{\ell-1} \left(\frac{r_{i+1}}{r_i}\right)^{-\frac{n-i-1}{2}} D_i^{-i+\delta} \max_{\vec{S}_\ell \in \vec{\mathcal{S}}_\ell} \|f_{\vec{S}_\ell}\|_2^2$$

$$= r_\ell^{-\frac{n-\ell}{2}} \prod_{i=m-1}^{\ell-1} r_i^{-1/2} D_i^{-i+\delta} \max_{\vec{S}_\ell \in \vec{\mathcal{S}}_\ell} \|f_{\vec{S}_\ell}\|_2^2 \qquad (57)$$

holds. This bound will be exploited in different ways.

10.3. **Guth's estimate revisited.** As a warm up exercise for the more involved computation to follow, here Guth's k-broad estimate from [12] is recovered using the above inequalities. In particular, taking $\ell = n$, the inequality (57) simplifies to give:

Key estimate.

$$\max_{O \in \mathcal{O}} \|f_O\|_2^2 \lesssim \prod_{i=m-1}^{n-1} r_i^{-1/2} D_i^{-i+\delta} \|f\|_2^2.$$

Combining this with (56), one concludes that

$$\|Ef\|_{\mathrm{BL}_{k,A}^p(B_R)} \lesssim \prod_{i=m-1}^{n-1} r_i^{X_i} D_i^{Y_i + O(\delta)} \|f\|_2 \qquad (58)$$

where

$$X_i := \frac{\beta_{i+1} - \beta_i}{2} - \frac{1}{2}\left(\frac{1}{2} - \frac{1}{p_n}\right); \qquad Y_i := \frac{\beta_{i+1}}{2} - (i+1)\left(\frac{1}{2} - \frac{1}{p_n}\right).$$

In order to ensure that there is an acceptable dependence on R in (58), the parameters must be chosen so as to ensure that $X_i, Y_i \leqslant 0$ for $m \leqslant i \leqslant n-1$ and $Y_{m-1} \leqslant 0$.

Remark 10.1. Ostensibly, the above conditions on the Y_i do not take into account the additional $O(\delta)$-powers of the D_i in (58). By perturbing the exponents which result under these conditions and choosing δ sufficiently small depending on the choice of perturbation, the $O(\delta)$-powers may nevertheless be safely handled. This perturbative argument yields an open range of k-broad estimates, which can be trivially extended to a closed range via interpolation through logarithmic convexity (the interpolation argument relies on the fact that one is permitted an R^ε-loss in the constants in the k-broad inequalities).

Recalling from the definitions that

$$\beta_i = \left(\frac{1}{2} - \frac{1}{p_n}\right)\left(\frac{1}{2} - \frac{1}{p_i}\right)^{-1},$$

the condition $X_i \leqslant 0$ is equivalent to

$$\left(\frac{1}{2} - \frac{1}{p_{i+1}}\right)^{-1} - \left(\frac{1}{2} - \frac{1}{p_i}\right)^{-1} \leqslant 1 \qquad (59)$$

whilst the condition $Y_{i-1} \leqslant 0$ is equivalent to

$$\left(\frac{1}{2} - \frac{1}{p_i}\right)^{-1} - 2i \leqslant 0. \qquad (60)$$

Choose $p_m := \frac{2m}{m-1}$ so that the exponent satisfies $(\frac{1}{2} - \frac{1}{p_m})^{-1} = 2m$ and therefore (60) is saturated in the $i = m$ case. The remaining p_i are then chosen so as to satisfy

$$\left(\frac{1}{2} - \frac{1}{p_i}\right)^{-1} = m + i$$

so that (59) is saturated for every value of i. With this choice, (60) automatically holds for all the remaining indices $m + 1 \leqslant i \leqslant n$. The worst situation occurs when $m = k$, in which case one deduces that the inequality

$$\|Ef\|_{\mathrm{BL}^p_{k,A}(B_R)} \lesssim R^\varepsilon \|f\|_2 \tag{61}$$

holds for all $p \geqslant 2 + \frac{4}{n+k-2}$; this exponent agrees with that featured in [12, Proposition 8.1] and, indeed, the above argument is simply a reformulation of the proof appearing in [12].

10.4. **Improvement using the polynomial Wolff axioms.** To prove Theorem 2.1, the argument of the previous subsection is augmented with the bounds coming from the polynomial Wolff axiom theorem. This follows the strategy of [11], which established the $n = 3$ case of the theorem. The goal is to improve the range of p at the expense of weakening the L^2-type estimate (61) to an L^∞-type estimate

$$\|Ef\|_{\mathrm{BL}^p_{k,A}(B_R)} \lesssim R^\varepsilon \|f\|_\infty.$$

One key observation is that the choice of exponents in the previous subsection does not saturate the constraint (60) coming from the D_i exponents for $m \leqslant i \leqslant n - 1$. This provides some leeway, and the polynomial Wolff axiom theorem allows one to trade an acceptable loss in the D_i exponents for a gain in the r_i exponents, and thereby leads to an improvement in the p range.

Fix $m \leqslant \ell \leqslant n$ and apply Lemma 8.6 to deduce that

$$\max_{\vec{S}_\ell \in \vec{\mathcal{S}}_\ell} \|f_{\vec{S}_\ell}\|_2^2 \lessapprox r_\ell^{-\frac{n-\ell}{2}} \max_{\substack{\vec{S}_\ell \in \vec{\mathcal{S}}_\ell \\ \theta_\ell : r_\ell^{-1/2} - \mathrm{cap}}} \|f_{\vec{S}_\ell}\|_{L^2_{\mathrm{avg}}(\theta_\ell)}^2. \tag{62}$$

Let $\ell \leqslant i \leqslant n - 1$ and note that, by Property 3 of [alg 2],

$$\max_{\substack{\vec{S}_i \in \vec{\mathcal{S}}_i \\ \theta_i : r_i^{-1/2} - \mathrm{cap}}} \|f_{\vec{S}_i}\|_{L^2_{\mathrm{avg}}(\theta_i)}^2 \lessapprox \left(\frac{r_{i+1}}{r_i}\right)^{-\frac{n-i-1}{2}} D_i^\delta \max_{\substack{\vec{S}_{i+1} \in \vec{\mathcal{S}}_{i+1} \\ \theta_{i+1} : r_{i+1}^{-1/2} - \mathrm{cap}}} \|f_{\vec{S}_{i+1}}\|_{L^2_{\mathrm{avg}}(\theta_{i+1})}^2 \tag{63}$$

Combining (62) with $n - \ell$ applications of (63), we obtain

$$\max_{\vec{S}_\ell \in \vec{\mathcal{S}}_\ell} \|f_{\vec{S}_\ell}\|_2^2 \lessapprox r_\ell^{-\frac{n-\ell}{2}} \prod_{i=\ell}^{n-1} \left(\frac{r_{i+1}}{r_i}\right)^{-\frac{n-i-1}{2}} D_i^\delta \max_{\theta : R^{-1/2} - \mathrm{cap}} \|f\|_{L^2_{\mathrm{avg}}(\theta)}^2$$

$$\leqslant \left(\prod_{i=\ell}^{n-1} r_i^{-1/2}\right)\left(\prod_{i=\ell}^{n-1} D_i^\delta\right) \|f\|_\infty^2.$$

Substituting this estimate into (57), one concludes that

$$\max_{O \in \mathcal{O}} \|f_O\|_2^2 \lessapprox r_\ell^{-\frac{n-\ell}{2}} \left(\prod_{i=m}^{n-1} r_i^{-1/2} D_i^\delta\right)\left(\prod_{i=m-1}^{\ell-1} D_i^{-i}\right) \|f\|_\infty^2$$

for all $m \leqslant \ell \leqslant n$. Finally, these $n - m + 1$ different estimates are combined into a single inequality by taking a weighted geometric mean, yielding:

Key estimate. *Let* $0 \leqslant \gamma_m, \ldots, \gamma_n \leqslant 1$ *satisfy* $\sum_{j=m}^{n} \gamma_j = 1$. *Then*

$$\max_{O \in \mathcal{O}} \|f_O\|_2^2 \lesssim \prod_{i=m-1}^{n-1} r_i^{-\frac{1+(n-i)\gamma_i}{2}} D_i^{-i(1-\sum_{j=m}^{i} \gamma_j)+O(\delta)} \|f\|_\infty^2.$$

Thus, the parameters γ_j allow a loss in the D_i exponents to be traded for a gain in the r_i exponents.

The key estimate may be combined with the inequality (56) from Section 10.2 to yield the bound

$$\|Ef\|_{\mathrm{BL}_{k,A}^{p_n}(B_R)} \lesssim \prod_{i=m-1}^{n-1} r_i^{X_i} D_i^{Y_i+O(\delta)} \|f\|_\infty$$

where

$$X_i := \frac{\beta_{i+1} - \beta_i}{2} - \frac{1+(n-i)\gamma_i}{2}\left(\frac{1}{2} - \frac{1}{p_n}\right);$$

$$Y_i := \frac{\beta_{i+1}}{2} - \left(1 + i\left(1 - \sum_{j=m}^{i} \gamma_j\right)\right)\left(\frac{1}{2} - \frac{1}{p_n}\right).$$

As in Section 10.3, one chooses the various exponents so as to ensure $X_i, Y_i \leqslant 0$ for all $m \leqslant i \leqslant n - 1$ and $Y_{m-1} \leqslant 0$. Owing to the extra degrees of freedom offered by the γ_j parameters, in this case one may in fact saturate all the conditions: that is, the parameters may be chosen so as to ensure that $X_i = Y_i = 0$. Indeed, the condition $X_i = 0$ is equivalent to

$$\left(\frac{1}{2} - \frac{1}{p_{i+1}}\right)^{-1} - \left(\frac{1}{2} - \frac{1}{p_i}\right)^{-1} = 1 + (n - i)\gamma_i \tag{64}$$

whilst the condition $Y_{i-1} = 0$ is equivalent to

$$\left(\frac{1}{2} - \frac{1}{p_i}\right)^{-1} = 2i - 2(i - 1)\sum_{j=m}^{i-1} \gamma_j \tag{65}$$

Once again, choose $p_m := \frac{2m}{m-1}$ so that (65) holds in the $i = m$ case. The remaining p_i are then defined in terms of the γ_j by the equation

$$\left(\frac{1}{2} - \frac{1}{p_i}\right)^{-1} = m + i + \sum_{j=m}^{i-1} (n - j)\gamma_j \tag{66}$$

so that each of the $n - m$ constraints in (64) is met.

It remains to solve for the $n - m + 1$ variables $\gamma_m, \ldots, \gamma_n$; note that there are $n - m + 1$ remaining constraints (in particular, there are $n - m$ constraints left over from (65) together with the condition that the γ_j must sum to 1) and so the number of equations in our system equals the number of variables. By comparing the right-hand sides of (65) and (66), it follows that

$$\sum_{j=m}^{i-1} (n - j + 2i - 2)\gamma_j = i - m \qquad \text{for } m + 1 \leqslant i \leqslant n, \tag{67}$$

from which we read off that $\gamma_m = (n + m)^{-1}$. To solve this linear system, let κ_i denote the left-hand side of the equation in the above display and observe that

$$\kappa_{i+1} - 2\kappa_i + \kappa_{i-1} = (n + i)\gamma_i - (n + i - 3)\gamma_{i-1} \qquad \text{for } m + 1 \leqslant i \leqslant n - 1,$$

where $\kappa_m := 0$. On the other hand, by considering the right-hand side of (67), it is clear that $\kappa_{i+1} - 2\kappa_i + \kappa_{i-1} = 0$. Combining these observations gives a recursive relation for the γ_j and from this one deduces that

$$\gamma_j = \frac{1}{n+m} \prod_{i=m}^{j-1} \frac{n+i-2}{n+i+1} = \frac{(n+m-1)(n+m-2)}{(n+j)(n+j-1)(n+j-2)}$$

for $m+1 \leqslant j \leqslant n-1$. The remaining parameter γ_n is then given by[14]

$$\gamma_n = 1 - \sum_{j=m}^{n-1} \gamma_j,$$

so that the γ_j sum to 1.

It remains to check that these parameter values give the correct value of p_n, corresponding to the exponent $p_n(k)$ stated in Theorem 2.1. It follows from (65) that

$$\left(\frac{1}{2} - \frac{1}{p_n}\right)^{-1} = 2n - 2(n-1) \sum_{j=m}^{n-1} \frac{(n+m-1)(n+m-2)}{(n+j)(n+j-1)(n+j-2)}. \tag{68}$$

The expression on the right-hand side can be simplified by first writing the denominator in each summand as

$$\frac{1}{(n+j)(n+j-1)(n+j-2)} = \frac{1}{2}\left(\frac{1}{n+j-2} - \frac{1}{n+j-1}\right) - \frac{1}{2}\left(\frac{1}{n+j-1} - \frac{1}{n+j}\right)$$

and then using the resulting telescoping property of the sum. This yields the identity

$$\sum_{j=m}^{n-1} \frac{(n+m-1)(n+m-2)}{(n+j)(n+j-1)(n+j-2)} = \frac{1}{2}\left(1 - \frac{(n+m-1)(n+m-2)}{(2n-1)2(n-1)}\right). \tag{69}$$

Plugging this into (68) and performing some simple algebraic manipulations, one concludes that

$$p_n = 2 + \frac{8(2n-1)}{n(5n+2m-9) + m(m-3) + 4} \leqslant p_n(k)$$

for $m \geqslant k$, which completes the proof. $\qquad \square$

11. Final remarks

Remark 11.1. One direction by which the argument could be improved would be to develop a more efficient mechanism for converting k-broad estimates into linear estimates than Proposition 2.2. One such mechanism does indeed already exist and is described in the work of Bourgain–Guth (see the fourth section of [6] or [22, 38] for an alternative presentation of this method). In particular, Bourgain–Guth [6] use Kakeya-type estimates to prove a stronger version of Proposition 2.2 in which the constraint $p \geqslant 2 + \frac{4}{2n-k}$ is slightly relaxed. Demeter [9] used this approach (combined with recent advances on the Kakeya conjecture [15, 46]) to give the previous best range for the restriction problem in $\mathbb{R}^4$ (namely, $p > 2 + \frac{66642}{83303}$). In fact, using Theorem 2.1 (and, in particular, the 3-broad estimate in four dimensions with $p_4(3) = 2 + \frac{7}{9}$) one can slightly improve Demeter's result to $p > 2 + \frac{1407}{1759}$ via the same method. For other low dimensions the use of the more efficient Bourgain–Guth mechanism is limited due to the lack of understanding of the Kakeya problem in this regime. In high dimensions, however, stronger Kakeya maximal and X-ray transform estimates are available owing to the sum-difference approach to Kakeya,

---

[14]To ensure this is a valid solution, one must verify that $\gamma_n \geqslant 0$ (so that $0 \leqslant \gamma_j \leqslant 1$ for all $m \leqslant j \leqslant n$). This property follows directly from the identity (69) below.

which was pioneered by Bourgain [5] and later honed by Katz–Tao [19, 20] and Oberlin [27]. Potentially, improvements could be obtained in high dimensional cases using the more efficient Bourgain–Guth mechanism and the Kakeya-type estimates arising from sum-difference theory; however, since the computation of the various exponents is rather involved and any gain is likely to be very small, this has not been pursued here.

Remark 11.2. An alternative approach to improving the range of restriction estimates would be to attempt to establish a stronger version of Theorem 2.1. This has been achieved for $n = 3$ in the work of Wang [40] who showed that (BL_k^p) holds in the wider range $p > 3 + \frac{3}{13}$ in this case (this in turn implies the best-known result on the restriction problem in $\mathbb{R}^3$; see Figure 1). The proof of Wang's theorem relies on a careful analysis and exploitation of certain underlying geometric features of the restriction problem; it would be of interest to extend and incorporate this analysis into the study of higher dimensional situations.

Remark 11.3. It is not difficult to extend the methods of this article to treat the class of (compact pieces of) hypersurfaces with strictly positive principal curvatures, which includes the unit sphere S^{n-1}. To do this, one applies a standard argument to reduce considerations to hypersurfaces of *elliptic-type*, as defined in [25, 37] (see also [35, 11]). One may then appeal to the more general transverse equidistribution results of [13] in place of Lemma 8.4. A more involved version of the Bourgain–Guth method for passing from k-broad to linear estimates is also required, but this already essentially appears in [6] (see also [13]). For this it is useful to work with the class of elliptic-type hypersurfaces (rather than specific examples such as S^{n-1}), since this class is closed under parabolic rescaling.

On the other hand, the method breaks down when one considers general (compact pieces of) hypersurfaces of non-vanishing Gaussian curvature. For instance, for the prototypical example of a graph of a non-degenerate quadratic form, the transverse equidistribution estimate from Lemma 8.4 fails to hold in mixed signature cases (see [13] for further discussion of such phenomena).

Remark 11.4. Another possible direction in which to strengthen the results would be to establish analogous estimates for Bochner–Riesz multipliers. An obvious approach to this would be to follow the classical Carleson–Sjölin argument [8] (see also [17]) which reduces the problem to establishing certain L^p estimates for oscillatory integrals of the form

$$T^\lambda f(x) := \int_{\mathbb{R}^n} e^{i\lambda|x-y|} a(x,y) f(y) \, dy, \tag{70}$$

where a is some smooth, compactly supported amplitude. Here the key difficulty is to obtain a favourable dependence in the inequality on the parameter $\lambda \gg 1$. After fixing one of the components of y and scaling, one obtains an operator which can be thought of as a perturbed version of Ef. The problem is then to show that the arguments used to study Ef are stable under perturbation; see [21, 6, 13] for recent examples of this approach, producing the current best-known results for the Bochner–Riesz problem.

Again it is useful to work with a class of oscillatory integral operators which is closed under rescaling, rather than just the specific example arising from the Bochner–Riesz problem. Here some care is needed, however: for a natural class of variable coefficient operators which extends the family of extension operators associated to positively-curved hypersurfaces, the desired L^p estimates are *false* for the range of p featured in this article. Counterexamples of this kind first appeared in work of Bourgain [3] and were further studied in [4, 26, 41, 6] (see also [13]). For

instance, Minicozzi and Sogge [26] considered the analogue of (70) defined over a compact Riemannian manifold (M, g) given by

$$T^\lambda f(x) := \int_M e^{i\lambda \mathrm{dist}_g(x,y)} a(x,y) f(y) \, \mathrm{d}y, \tag{71}$$

where dist_g is the Riemannian distance function on M. These operators arise naturally in the study of Bochner–Riesz multipliers on compact manifolds, defined with respect to the spectral decomposition of the Laplace–Beltrami operator (see, for instance, [29, Chapter 5]). In [26] examples of (M, g) were found for which the desired L^p estimates for (71) could only hold for a relatively small range of p. Sharp inequalities for such examples were later established in the work of Guth, Iliopoulou and the first author [13]. The problematic behaviour for certain M can be attributed to the fact that analogues of the polynomial Wolff axioms can fail to hold for families of geodesic tubes relevant to the study of T^λ.

Remark 11.5. It is well-known that L^p-estimates for the extension operator imply bounds for the Kakeya maximal function. Let $\mathbf{T}$ be a collection of direction-separated R-tubes in $\mathbb{R}^n$, with angle at least $R^{-1/2}$ between each pair of tubes. If the estimate

$$\|Ef\|_{L^p(\mathbb{R}^n)} \lesssim \|f\|_{L^p(\mathbb{R}^{n-1})} \tag{72}$$

is valid for some $p > 2$, then

$$\Big\| \sum_{T \in \mathbf{T}} \mathbf{1}_T \Big\|_{L^{p/2}(\mathbb{R}^n)} \lesssim R^{n-1-\frac{2n}{p}} \Big(\sum_{T \in \mathbf{T}} |T| \Big)^{2/p}; \tag{73}$$

see, for example, [43] for a proof of this fact. New estimates for the Kakeya maximal operator with $n = 9$ are obtained by plugging in our estimates for the extension operator. For other values of n the maximal function estimates that arise in this way are strictly weaker than those previously obtained by Wolff [42] or Katz–Tao [20].

Maximal inequalities such as (73) imply lower bounds on the dimensions of Kakeya sets. Recall that a set $K \subset \mathbb{R}^n$ is *Kakeya* if it is compact and it contains a unit line segment in every direction. Let $d(n)$ denote the infimum of the Hausdorff dimensions of Kakeya sets in $\mathbb{R}^n$; explicitly,

$$d(n) := \inf\{\dim K : K \subset \mathbb{R}^n \text{ Kakeya}\}.$$

The Kakeya conjecture then asserts that $d(n) = n$. As is well-known, the inequality (73) implies that

$$d(n) \geqslant \frac{2p}{p-2} - n.$$

However, the aforementioned maximal inequality is not strong enough to improve over the existing lower bounds of Katz–Tao [20] for the Hausdorff dimension of Kakeya sets, obtained via the sum-difference method.

In terms of the asymptotic perspective espoused in this article, if (72) holds for $p = 2 + \lambda n^{-1} + O(n^{-2})$, then

$$d(n) \geqslant \frac{4-\lambda}{\lambda} n + O(1).$$

Taking λ to be the value given by Theorem 1.1, it follows that

$$\frac{4-\lambda}{\lambda} = 4 - 2\sqrt{3} = 0.535...,$$

which provides a high dimensional improvement over the classical $d(n) \geqslant \frac{n+2}{2}$ bound of Wolff [42]. Once again, this does not improve the results of Katz–Tao [20]. Nevertheless, it seems of interest that one can go beyond the $d(n) \geqslant \frac{n}{2} + O(1)$ range for the Kakeya problem using a different approach than the sum-difference method, and that oscillatory methods are becoming more effective in the Kakeya problem.

We have since obtained further bounds for the Kakeya conjecture by applying similar arguments to those of this article directly in that context [16].

References

[1] J. Bennett, A. Carbery, and T. Tao, *On the multilinear restriction and Kakeya conjectures*, Acta Math. **196** (2006), no. 2, 261–302.

[2] J. Bourgain, *Besicovitch type maximal operators and applications to Fourier analysis*, Geom. Funct. Anal. **1** (1991), no. 2, 147–187.

[3] ———, *L^p-estimates for oscillatory integrals in several variables*, Geom. Funct. Anal. **1** (1991), no. 4, 321–374.

[4] ———, *Some new estimates on oscillatory integrals*, Essays on Fourier analysis in honor of Elias M. Stein (Princeton, NJ, 1991), Princeton Math. Ser. **42**, 83–112.

[5] ———, *On the dimension of Kakeya sets and related maximal inequalities*, Geom. Funct. Anal. **9** (1999), no. 2, 256–282.

[6] J. Bourgain and L. Guth, *Bounds on oscillatory integral operators based on multilinear estimates*, Geom. Funct. Anal. **21** (2011), no. 6, 1239–1295.

[7] A. Carbery, *Restriction implies Bochner-Riesz for paraboloids*, Math. Proc. Cambridge Philos. Soc. **111** (1992), no. 3, 525–529.

[8] L. Carleson and P. Sjölin, *Oscillatory integrals and a multiplier problem for the disc*, Studia Math. **44** (1972), 287–299.

[9] C. Demeter, *On the restriction theorem for paraboloid in $\mathbb{R}^4$*, Colloq. Math. **156** (2019), 301–311.

[10] C. Fefferman, *Inequalities for strongly singular convolution operators*, Acta Math. **124** (1970), 9–36.

[11] L. Guth, *A restriction estimate using polynomial partitioning*, J. Amer. Math. Soc. **29** (2016), no. 2, 371–413.

[12] ———, *Restriction estimates using polynomial partitioning II*, Acta Math. 221 (2018), 81–142.

[13] L. Guth, J. Hickman, and M. Iliopoulou, *Sharp estimates for oscillatory integral operators via polynomial partitioning*, Acta Math., to appear, Preprint: arXiv:1710.10349.

[14] L. Guth and N.H. Katz, *On the Erdős distinct distances problem in the plane*, Ann. of Math. (2) **181** (2015), no. 1, 155–190.

[15] L. Guth and J. Zahl, *Polynomial Wolff axioms and Kakeya-type estimates in $\mathbb{R}^4$*, Proc. London Math. Soc. **117** (2018), no. 1, 192–220.

[16] J. Hickman and K.M. Rogers, *New Kakeya estimates using the polynomial Wolff axioms*, Preprint: arXiv:1901.01802.

[17] L. Hörmander, *Oscillatory integrals and multipliers on FL^p*, Ark. Mat. **11** (1973), 1–11.

[18] N.H. Katz and K.M. Rogers, *On the polynomial wolff axioms*, Geom. Funct. Anal. **28** (2018), 1706–1716.

[19] N.H. Katz and T. Tao, *Bounds on arithmetic projections, and applications to the Kakeya conjecture*, Math. Res. Lett. **6** (1999), no. 5-6, 625–630.

[20] ———, *New bounds for Kakeya problems*, J. Anal. Math. **87** (2002), 231–263, Dedicated to the memory of Thomas H. Wolff.

[21] S. Lee, *Linear and bilinear estimates for oscillatory integral operators related to restriction to hypersurfaces*, J. Funct. Anal. **241** (2006), 56–98.

[22] R. Lucà and K.M. Rogers, *Average decay of the Fourier transform of measures with applications*, J. Eur. Math. Soc. **21** (2019), 465–506.

[23] Maplesoft, a division of Waterloo Maple Inc., *Maple 17*.

[24] A. Moyua, A. Vargas, and L. Vega, *Schrödinger maximal function and restriction properties of the Fourier transform*, Internat. Math. Res. Notices (1996), no. 16, 793–815.

[25] ———, *Restriction theorems and maximal operators related to oscillatory integrals in $\mathbf{R}^3$*, Duke Math. J. **96** (1999), no. 3, 547–574.

[26] W.P. Minicozzi II and C.D. Sogge, *Negative results for Nikodym maximal functions and related oscillatory integrals in curved space*, Math. Res. Lett. **4** (1997), no. 2-3, 221–237.

[27] R. Oberlin, *Two bounds for the X-ray transform*, Math. Z. **266** (2010), no. 3, 623–644.

[28] J. Ramos, *A trilinear restriction estimate with sharp dependence on transversality*, Amer. J. Math. **140** (2018), no. 6, 1443–1469.

[29] C.D. Sogge, *Fourier integrals in classical analysis*, second edition, Cambridge Tracts in Mathematics **210**, Cambridge University Press, Cambridge, 2017.

[30] J. Solymosi and T. Tao, *An incidence theorem in higher dimensions*, Disc. Comp. Geom. **48** (2012), 255–280.

[31] E.M. Stein, *Some problems in harmonic analysis*, Harmonic analysis in Euclidean spaces (Proc. Sympos. Pure Math., Williams Coll., Williamstown, Mass., 1978), Part 1, Proc. Sympos. Pure Math., XXXV, Part, Amer. Math. Soc., Providence, R.I., 1979, pp. 3–20.

[32] A.H. Stone and J.W. Tukey, *Generalized "sandwich" theorems*, Duke Math. J. **9** (1942), 356–359.

[33] R.S. Strichartz, *Restrictions of Fourier transforms to quadratic surfaces and decay of solutions of wave equations*, Duke Math. J. **44** (1977), no. 3, 705–714.

[34] T. Tao, *The Bochner-Riesz conjecture implies the restriction conjecture*, Duke Math. J. **96** (1999), no. 2, 363–375.

[35] ———, *A sharp bilinear restrictions estimate for paraboloids*, Geom. Funct. Anal. **13** (2003), no. 6, 1359–1384.

[36] T. Tao and A. Vargas, *A bilinear approach to cone multipliers. I. Restriction estimates*, Geom. Funct. Anal. **10** (2000), no. 1, 185–215.

[37] T. Tao, A. Vargas, and L. Vega, *A bilinear approach to the restriction and Kakeya conjectures*, J. Amer. Math. Soc. **11** (1998), no. 4, 967–1000.

[38] F. Temur, *A Fourier restriction estimate for surfaces of positive curvature in $\mathbb{R}^6$*, Rev. Mat. Iberoam. **30** (2014), no. 3, 1015–1036.

[39] P.A. Tomas, *A restriction theorem for the Fourier transform*, Bull. Amer. Math. Soc. **81** (1975), 477–478.

[40] H. Wang, *A restriction estimate in $\mathbb{R}^3$ using brooms*, Preprint: arXiv:1802.04312.

[41] L. Wisewell, *Kakeya sets of curves*, Geom. Funct. Anal. **15** (2005), no. 6, 1319–1362.

[42] T. Wolff, *An improved bound for Kakeya type maximal functions*, Rev. Mat. Iberoamericana **11** (1995), no. 3, 651–674.

[43] ———, *Recent work connected with the Kakeya problem*, in *Prospects in mathematics (Princeton, NJ, 1996)*, 129–162, Amer. Math. Soc., Providence, RI.

[44] ———, *A sharp bilinear cone restriction estimate*, Ann. of Math. (2) **153** (2001), no. 3, 661–698.

[45] R. Wongkew, *Volumes of tubular neighbourhoods of real algebraic varieties*, Pacific J. Math. **159** (1993), no. 1, 177–184.

[46] J. Zahl, *A discretized Severi-type theorem with applications to harmonic analysis*, Geom. Funct. Anal. **28** (2018), no. 4, 1131–1181.

MATHEMATICAL INSTITUTE, UNIVERSITY OF ST ANDREWS, NORTH HAUGH, ST ANDREWS, FIFE, KY16 9SS, UK.

E-mail address: jeh25@st-andrews.ac.uk

INSTITUTO DE CIENCIAS MATEMÁTICAS CSIC-UAM-UC3M-UCM, CANTOBLANCO, MADRID, 28049, SPAIN.

E-mail address: keith.rogers@icmat.es