GAFA Geometric And Functional Analysis

KAKEYA SETS OF CURVES

L. Wisewell

1 Introduction

In this paper we investigate an analogue for curves of the famous Kakeya conjecture about straight lines. The simplest version of the latter asks whether a set in $\mathbb{R}^n$ that includes a unit line segment in every direction must necessarily have dimension n. The analogue we have in mind replaces the line segments by curved arcs from a specified family. (This is a quite different problem from that considered by Minicozzi and Sogge [MS] who looked at geodesics in curved space.) The families of curves we are interested in arise from Hörmander's conjecture in harmonic analysis, which deals with oscillatory integral operators of the form

$$T_N f(x) := \int_{\mathbb{R}^{n-1}} e^{iN\varphi(x,y)} a(x,y) f(y)\, \mathrm{d}y\,. \tag{1}$$

Here $x \in \mathbb{R}^n$, $y \in \mathbb{R}^{n-1}$, a is some smooth cut-off, and the phase function φ is assumed to be smooth on the support of a and to have the following properties:

$$\text{The matrix } \tfrac{\partial^2 \varphi}{\partial x \partial y}(x,y) \text{ has full rank } n-1\,. \tag{2}$$

For all $\theta \in S^{n-1}$ the map $y \mapsto \theta \cdot \tfrac{\partial \varphi}{\partial x}(x,y)$ has only

$$\text{non-degenerate critical points}\,. \tag{3}$$

In [B2] it is observed that by making appropriate changes of variable, any phase satisfying these criteria can be expressed in the form

$$\varphi(x,y) = y^{\mathrm{t}} x' + x_n y^{\mathrm{t}} A y + O\big(|x_n||y|^3 + |x|^2 |y|^2\big) \tag{4}$$

with A an invertible $(n-1) \times (n-1)$ matrix and t denoting transpose. Hörmander showed that both Restriction and Bochner–Riesz problems can be formulated as special cases of operators T_N, which prompted him to ask the following:

QUESTION 1 (Hörmander [H]).   Is it true that for every φ satisfying the above properties, the operator T_N has the bound

$$\|T_N f\|_s \lesssim N^{-n/s}\|f\|_r \tag{5}$$

for $\frac{1}{s} < \frac{n-1}{2n}$ and $\frac{1}{s} \leq \frac{n-1}{n+1}\frac{1}{r'}$?

Hörmander himself proved this for $n = 2$ [H], and in higher dimensions it has been proved for $s \geq \frac{2(n+1)}{n-1}$ by Stein [St]. The known and conjectured regions are shown in Figure 1.
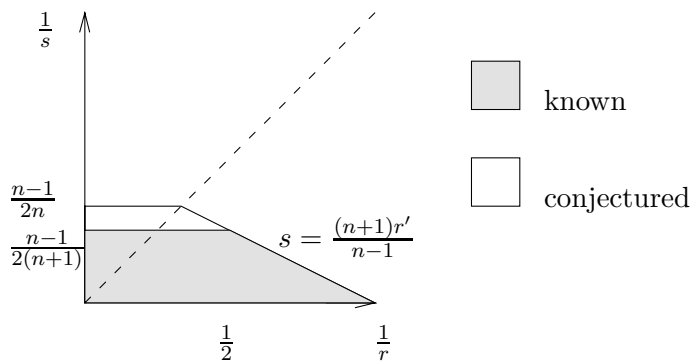

Figure 1: Exponents for Hörmander's conjecture

It was a great surprise in 1991 when Bourgain [B2] disproved Hörmander's conjecture. Roughly, he showed that in dimension three for most phases the best exponent s is strictly greater than $\frac{2n}{n-1} = 3$, and that there exist phases where the known value $\frac{2(n+1)}{n-1} = 4$ is the best. More precisely:

Theorem 1 (Worst case [B2]).   In dimension three there is a phase function, namely

$$\varphi(x,y) = x_1 y_1 + x_2 y_2 + x_3 y_1 y_2 + \tfrac{1}{2}x_3^2 y_1^2$$

for which (5) fails for all $s < 4$, even with $r = \infty$.

Theorem 2 (Generic failure [B2]).   In dimension three, if φ has the property that

$$\left.\frac{\partial^2}{\partial y^2}\left(\frac{\partial^2 \varphi}{\partial x_3^2}\right)\right|_{x=0,y=0} \quad \text{is not a multiple of} \quad \left.\frac{\partial^2}{\partial y^2}\left(\frac{\partial \varphi}{\partial x_3}\right)\right|_{x=0,y=0} \tag{6}$$

then the inequality (5) cannot hold even for $r = \infty$ unless $s \geq 118/39 > 3$.

His method was to link the oscillatory integral problems to Kakeya-type problems about curves. Diagrammatically we have the following chain of implications:

$$
\begin{array}{ccccc}
\text{Oscillatory} & & \text{Kakeya maximal} & & \text{Kakeya sets} \\
\text{integral estimates} & \Longrightarrow & \text{function estimates} & \Longrightarrow & \text{have large dimension.}
\end{array}
$$

$$(7)$$

(Compare this with Fefferman's famous counterexample to the disc conjecture [Fe], which used the fact that Kakeya sets can have measure zero to disprove a Bochner–Riesz type estimate.)

Both Restriction and Bochner–Riesz give rise to straight line problems – strictly speaking, Restriction implies the Kakeya conjecture while Bochner–Riesz implies the Nikodym conjecture, which is like the Kakeya conjecture but with the roles of positions and directions exchanged. In the straight-line case these are equivalent and so attention has been focused entirely on the former, but for any fixed class of curves they are different, as will be seen in section 3 where we consider quadratic curves. The relationship between the two is explored in [C2], [T] and [CW].

Since 1991 there has been much progress on the straight-line Kakeya problem, with contributions from Bourgain, Wolff, Katz, Łaba, Tao and Schlag. Our aim in this paper is to apply some of these new techniques to the curved case. As we shall see, we can prove positive results for certain families of curves, which may indicate that their corresponding phase functions allow reasonably good non-trivial bounds for the operators T_N.

We begin by giving precise definitions and brief proofs of the implications above. Then in section 2 we prove the so-called "trivial bound" (a maximal function estimate implying that the sets have dimension at least $\frac{n+1}{2}$), which holds for a very broad class of curves. From section 3 onwards we restrict to quadratic curves, and demonstrate (in Theorem 10) that these are still general enough to exhibit the pathological behaviour discovered by Bourgain. We then tackle the maximal function problem by means of geometry, proving a result (Theorem 12) corresponding to the lower bound $\frac{n+2}{2}$ for the dimension of Nikodym sets of parabolas satisfying a certain algebraic condition. Finally we look at arithmetic methods, and obtain (in Theorems 19, 20 and 28) lower bounds of the form $\alpha n + \beta$ with $\alpha > 1/2$ for the dimension of various sets of curves, including a bound for the Nikodym sets of the previous section which equals the best currently known for straight lines.

1.1 The relevance of Kakeya with curves.

Given a phase function φ and cutoff $a \in C_c^\infty$ as in (1), define curves and curved tubes as follows:

NOTATION. Let $y, \omega \in \mathbb{B}^{n-1}$ and let $\delta > 0$ be a thickness. Define

$$
\Gamma_y(\omega) := \big\{ x \in \mathbb{R}^n : \nabla_y\, \varphi(x,y) = \omega\,,\ (x,y) \in \mathrm{supp}(a) \big\}
$$

$$T_y^\delta(\omega) := \left\{ x \in \mathbb{R}^n : |\nabla_y \varphi(x, y) - \omega| < \delta, \ (x, y) \in \mathrm{supp}(a) \right\}$$

to be the curve "centre" ω in "direction" y and the corresponding δ-tube.

Here $\mathbb{B}^{n-1}$ denotes a ball in $\mathbb{R}^{n-1}$ of some constant radius: for quadratic curves the unit ball will do, but more generally we will need to choose the constant to depend on φ, although of course larger sets of directions can then be handled simply by taking unions of small enough balls.

Using the rank condition (2) and the implicit function theorem we see that $\Gamma_y(\omega)$ is indeed a smooth curve. The descriptions "centre" and "direction" are to aid intuition; in some cases the meaning of the variables may in fact be the other way round. On one hand, the Restriction problem for the paraboloid corresponds to the phase $\varphi(x, y) = y^\mathrm{t} x' + x_n y^\mathrm{t} y$, so that $\Gamma_y(\omega)$ is a straight line centred at $\left(\begin{smallmatrix} \omega \\ 0 \end{smallmatrix} \right)$ in direction $\left(\begin{smallmatrix} y \\ 1 \end{smallmatrix} \right)$. But on the other hand, Bochner–Riesz for the paraboloid has the phase $\varphi(x, y) = \frac{1}{x_n} y^\mathrm{t} x' + \frac{1}{x_n} y^\mathrm{t} y$, and $\Gamma_y(\omega)$ is still a straight line, but with centre $\left(\begin{smallmatrix} y \\ 0 \end{smallmatrix} \right)$ and direction $\left(\begin{smallmatrix} \omega \\ 1 \end{smallmatrix} \right)$.

Because of the smoothness of φ, the tubes have the following "doubling property": Suppose that $|y - \bar{y}| < \delta$ and $|\omega - \bar{\omega}| < \delta$. Then $T_y^{C\delta}(\omega) \supset T_{\bar{y}}^\delta(\bar{\omega})$ for some constant C depending only on φ. This will often allow us to consider only finite δ-separated collections of tubes.

By analogy with the straight-line case, we define the following sets:

DEFINITION (Curved Kakeya set). A set $E \subset \mathbb{R}^n$ is a curved Kakeya set (associated to φ) if for all $y \in \mathbb{B}^{n-1}$ there exists an $\omega \in \mathbb{B}^{n-1}$ such that $\Gamma_y(\omega) \subset E$.

So for the Restriction phase above, this is the usual definition of a Kakeya set, while for the Bochner–Riesz phase this is a set that includes a line segment in some direction through every one of a large set of points, which might be termed a *Nikodym set* although this is not quite the same definition as is usually given in, say, [F].

As we shall see shortly, curved Kakeya sets need not have full dimension, so rather than a conjecture we have a question:

QUESTION 2. *Given a phase function, what is the minimum possible dimension for its corresponding Kakeya sets? For which curves must the dimension be exactly n?*

We can ask this about either the usual Hausdorff notion of dimension, or more weakly about the *Minkowski dimension*. This is always greater than or equal to the Hausdorff and is simpler to use: a set E has (upper) Minkowski dimension at least d if and only if its δ-neighbourhood $\mathrm{nbd}_\delta(E)$ has Lebesgue measure satisfying $|\mathrm{nbd}_\delta(E)| \gtrsim \delta^{n-d}$. This is the notion we shall use in section 5 when applying the arithmetic techniques.

More difficult questions about overlap of tubes can be posed in terms of maximal operators.

DEFINITION (Curved maximal operator). The curved Kakeya maximal function (associated to φ and of eccentricity $1/\delta$) is the operator that takes a function f on $\mathbb{R}^n$ to the function $\mathcal{K}_\delta\, f$ on $\mathbb{B}^{n-1}$ given by

$$\mathcal{K}_\delta\, f(y) := \sup_{\omega \in \mathbb{B}^{n-1}} \frac{1}{|T_y(\omega)|} \int_{T_y(\omega)} |f(x)|\, \mathrm{d}x\,.$$

In the straight-line case it is conjectured that this should have $L^n \to L^n$ operator norm at most $\delta^{-\varepsilon}$. This and the bounds which follow by interpolation are shown in Figure 2. However as we shall see shortly, for many
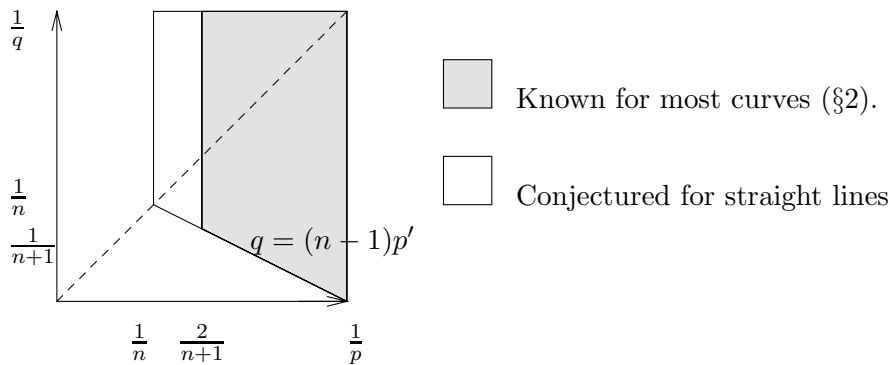

Figure 2: Region where the $L^p \to L^q$ norms of the Kakeya maximal operator should be at most $\delta^{1-n/p-\varepsilon}$

families of curves the $L^n \to L^n$ bound is false, so we have another question.

QUESTION 3. *Given a phase function, how can we find the best bound for the corresponding maximal functions? For which phases is an $L^n \to L^n$ bound of order $\delta^{-\varepsilon}$ possible?*

These estimates imply lower bounds for the dimension of the sets in the following way.

PROPOSITION 3 (Maximal implies dimension). *Assume that for some φ an estimate $\|\mathcal{K}_\delta\, f\|_{q,\infty} \leq C\delta^{-\alpha}\|f\|_{p,1}$ holds. Then the corresponding Kakeya sets have Hausdorff dimension at least $n - p\alpha$.*

So a sharp $L^p \to L^q$ bound implies that the sets have dimension at least p. The implication is easy if we use Minkowski dimension: simply let f be the characteristic function of the δ-neighbourhood of the set, and the

required estimate follows. The proof for the Hausdorff dimension is similar but requires a dyadic pigeonholing argument, following exactly that given by Wolff for the straight-line case in [Wo2, Lemma 1.6], with an implicit function argument to obtain suitable parameterisations of the curves.

So we have the second of the implications (7). We now turn to the first which relates the above to the oscillatory integrals T_N. The proof will show why the curves we have defined are natural.

PROPOSITION 4. *For some phase φ satisfying Hörmander's criteria (2) and (3), suppose that $\|T_N f\|_s \lesssim N^{-n/s}\|f\|_r$. Then the corresponding curved Kakeya maximal function is of restricted weak type (p,q) with norm at most $\delta^{-2(n/p-1)}$, where $p = (s/2)'$ and $q = (r/2)'$,*

To prove this, and also to prove the estimates for the maximal functions in later sections, it is helpful to linearise the maximal function so that instead of an L^p bound we can prove a "covering lemma" similar to those in [C1].

DEFINITION (Linearised operator). Decompose $\mathbb{B}^{n-1}$ into disjoint δ-cubes Q_j for $j \in \mathbb{B}^{n-1} \cap \delta\mathbb{Z}^{n-1}$. To each index j associate a tube $T_j = T_{y_j}^\delta(\omega_j)$ where $y_j \in Q_j$ and $\omega_j \in \mathbb{B}^{n-1}$. Define a linearisation of $\mathcal{K}_\delta$ by

$$\mathcal{LK}_\delta f(y) := \sum_j \mathbb{1}_{Q_j}(y) \frac{1}{|T_j|} \int_{T_j} f(x)\, \mathrm{d}x\,.$$

Now taking the adjoint of this operator puts the problem in the following useful form:

LEMMA 5 (Covering lemma). *Let $\{T_j\}_{j=1}^M$ be $1 \times \delta$-tubes with centres ω_j and directions y_j (where both of these are in $\mathbb{B}^{n-1}$). Then the estimate*

$$\left\| \sum_{j=1}^M \mathbb{1}_{T_j} \right\|_{p'} \leq A(\delta)(\delta^{n-1}M)^{1/q'}$$

holds for all choices of $y_j \in Q_j$ with arbitrary ω_j if and only if the (curved) Kakeya maximal function is of weak type (p, q) with constant $A(\delta)$.

This is easy to prove, and the details are given in [W2].

Now we use the linearisation to sketch a proof of Proposition 4, which will show the reason for the definition of the curves. This proof is similar to that given by Wolff in [Wo2, pp. 153–154] for the Restriction problem, but incorporating ideas found in Bourgain's "generic failure" proof for curves [B2, pp. 326–327].

Proof of Proposition 4. Suppose that we are given tubes T_j with directions $y_j \in Q_j$ and arbitrary centres ω_j as above. Set

$$f(y) = \sum_{j=1}^{M} \varepsilon_j e^{-iN\omega_j \cdot (y-y_j)} \mathbb{1}_{Q_j}(y)$$

where the ε_j are random signs. Then $T_N f$ is a sum of integrals $\int_{Q_j} e^{iN(\varphi(x,y)-\omega_j \cdot (y-y_j))} a(x,y) \, dy$, and it is easy to see that if we choose $N \sim \delta^{-2}$ then the phase is roughly constant for $x \in T_j$ so that the integral is at least $\delta^{n-1} \mathbb{1}_{T_j}(x)$. Applying Khinchin's inequality and the assumed bound for $\|T_N\|_{r \to s}$ gives the covering lemma required. $\square$

An immediate corollary is that the optimal $s = \frac{2n}{n-1}$ would imply the best estimate $\|\mathcal{K}_\delta\|_{n \to n} \lesssim \delta^{-\varepsilon}$ for the curved Kakeya maximal function and full dimension for the sets. However, away from the optimal exponent this correspondence becomes very poor: Stein's result $s = \frac{2(n+1)}{n-1}$ merely implies that the sets have dimension at least 1.

2 "Trivial" Results for Most Curves

Before we go on to discuss the dimension of curved Kakeya and Nikodym sets we should first mention that they can indeed have measure zero. This is proved in [W1] by adapting a result due to Sawyer, and in fact applies to more general problems of surfaces lying in sets of measure zero.

Also we point out that, as with straight lines, the problem is entirely understood in dimension 2 since Hörmander's conjecture is true in the plane, and so the implications (7) give the best possible bounds for the maximal functions and set dimensions.

The next simplest result is the $\frac{n+1}{2}$ bound. For the maximal function with straight lines this was first proved in 1986 by Christ, Duoandikoetxea and Rubio de Francia [ChDR] using Fourier transform methods. Since then, more geometric proofs have been given. One of these is a two-slice version of the arithmetic methods which we look at in section 5 for the set dimension problem. Here we use the "bush argument" of Bourgain [B1] to obtain the stronger maximal function estimate. In a sense this is analogous to the two-slice method by point-line duality, since the main idea is an estimate for the size of the intersection of two different tubes (compare with Lemma 15). So we begin by looking at the way curved tubes can intersect, and prove the $L^{\frac{n+1}{2}}$ bound for a very broad class of families of curves. We then show geometrically why the non-degeneracy criterion (3)

is crucial, by providing a counterexample to the maximal function result if it is not assumed. We also mention the slightly curious fact that even without non-degeneracy, the set dimension result still holds.

At this point we restrict our attention to phases of a slightly simpler form than (4), namely those in which the higher-order terms depend only on x_n and not on x'. For convenience in later sections we write these as

$$\varphi(x,y) = y^{\mathrm{t}} M(x_n) x' + \tilde{\varphi}(x_n, y), \qquad (8)$$

where $M : \mathbb{R} \to GL(n-1, \mathbb{R})$ is a matrix-valued function. The curves corresponding to this can be parametrised by the height x_n as follows:

$$\Gamma_y(\omega) = \left\{ \begin{pmatrix} M(x_n)^{-1} \left[\omega - \nabla_y \tilde{\varphi}(x_n, y) \right] \\ x_n \end{pmatrix} : (x,y) \in \mathrm{supp}(a) \right\}, \qquad (9)$$

where the matrix inverse exists by the rank condition (2). This notation is introduced because the phases we want to look at in sections 3–5 which give rise to parabolic curves are more conveniently expressed in the form (8) than as in (4) where we had $M(x_n) \equiv I$.

We also now specify the radius of the ball $\mathbb{B}^{n-1}$. Given a phase φ of the form (8), write $\psi = \frac{\partial}{\partial x_n} \nabla_y \tilde{\varphi}$. Hörmander's criterion (3) tells us that the matrix $\frac{\partial}{\partial y} \psi$ has non-zero determinant throughout the support of the cutoff a, so let k be the minimum absolute value of its eigenvalues on this support. Then by the definition of the derivative, find $\rho > 0$ such that

$$\frac{|\psi(x_n, y) - \psi(z_n, \bar{y}) - D\psi(x_n, \bar{y})(y - \bar{y})|}{|y - \bar{y}|} < \frac{k}{2} \qquad (10)$$

whenever $|y - \bar{y}| \le \rho$. This constant ρ depends only on φ, and $\mathbb{B}^{n-1}$ will be taken to mean the ball of this radius from now on. (In the case of quadratic phases which we consider in the next sections, the above fraction is zero and so this issue does not arise.)

We now state and prove the crucial estimate for the size of the intersection of two tubes. Call two tubes $T_y(\omega)$, $T_{\bar{y}}(\bar{\omega})$ *d-separated* if $|y - \bar{y}| \ge d$.

LEMMA 6.   *Assuming (3), there is a constant C depending only on φ such that if two d-separated δ-tubes corresponding to curves of the form (9) meet, then the diameter of their intersection is at most $C\delta/d$.*

Proof. Curves of the form (9) have tangents given by

$$\begin{pmatrix} -M(x_n)^{-1} M'(x_n) M(x_n)^{-1} \left[\omega - \nabla_y \tilde{\varphi}(x_n, y) \right] - M(x_n)^{-1} \frac{\partial}{\partial x_n} \nabla_y \tilde{\varphi}(x_n, y) \\ 1 \end{pmatrix}.$$

If two different curves $\Gamma_y(\omega)$ and $\Gamma_{\bar{y}}(\bar{\omega})$ meet at height $x_n = t_0$, then

$$M(t_0)^{-1} \left[\omega - \nabla_y \tilde{\varphi}(t_0, y) \right] = M(t_0)^{-1} \left[\bar{\omega} - \nabla_y \tilde{\varphi}(t_0, \bar{y}) \right]$$

and so the difference between their tangents is simply

$$\left| M(t_0)^{-1}[\psi(t_0, y) - \psi(t_0, \bar{y})] \right|$$
$$\gtrsim \left| D\psi(t_0, \bar{y})(y - \bar{y}) \right| - \left| \psi(t_0, y) - \psi(t_0, \bar{y}) - D\psi(t_0, \bar{y})(y - \bar{y}) \right|$$
$$> \tfrac{k}{2}|y - \bar{y}| \quad \text{by (10)}.$$

So the tangents are at an angle comparable to $|y - \bar{y}|$ and so the diameter of the intersection is at most $\frac{\delta}{\delta + |y - \bar{y}|}$, giving the result claimed. $\square$

This allows us to prove the $L^{\frac{n+1}{2}}$ bound using the bush argument.

Theorem 7.   *Assuming (3), the curved Kakeya maximal function $\mathcal{K}_\delta$ corresponding to curves of the form (9) satisfies*

$$\| \mathcal{K}_\delta f \|_q \le C_\varepsilon \delta^{-(n/p - 1 + \varepsilon)} \| f \|_p \tag{11}$$

for $1 \le p \le \frac{n+1}{2}$ and $1 \le q \le (n-1)p'$.

Proof. It is enough to prove a restricted weak type estimate at the endpoint, since this implies strong type at the cost of an additional log [CHS, p. 48]. The proof follows exactly the bush argument for the straight-line case, a suitable version of which is given in [Wo2] or [W2]. $\square$

So the so-called "trivial bound" holds for *all* curves of the form (9), and in particular, it is true for the "worst case" example of Bourgain. That example had curves given by

$$\begin{pmatrix} \omega_1 - x_3 y_2 - x_3^2 y_1 \\ \omega_2 - x_3 y_1 \\ x_3 \end{pmatrix}.$$

If we choose $\omega_1 = 0$, $\omega_2 = -y_2$ then we see that each curve lies in the surface $x_1 = x_2 x_3$. So the Kakeya set has dimension two, and for this φ the "trivial" bound is in fact best possible.

This suggests that $\frac{n+1}{2}$ for the set dimension and maximal function ought to correspond to the exponent $s = \frac{2(n+1)}{n-1}$ in Hörmander's conjecture, since this is the result that is known to be true for all phases and cannot be improved for Bourgain's example. However, the implication proved in Proposition 4 is weaker; one feels that the factor of 2 in the power of δ we obtained should not be there.

The proof of the "trivial" bound also reveals the reason for the non-degeneracy criterion (3), since if this does not hold, then the curves can essentially share a tangent, which makes the intersection of the tubes larger than the estimate given in Lemma 6. This is illustrated in Figure 3. As one would expect, this behaviour means that the $L^{\frac{n+1}{2}}$ estimate for the maximal function fails.

(a) Proper Intersection   (b) Tangential Intersection

Figure 3: The intersection of two curved tubes

PROPOSITION 8. *If the non-degeneracy criterion (3) fails, then (11) also fails for $p = \frac{n+1}{2}$.*

Proof. Suppose that a degenerate critical point occurs at (t_0, z_0). With $\psi(x_n, y) := \frac{\partial}{\partial x_n} \nabla_y \tilde{\varphi}(x_n, y)$ as on page 1326, this means that $\det \frac{\partial}{\partial y} \psi(t, y)$ evaluated at (t_0, z_0) is zero. Let U be the subspace

$$U = \left\{ u \in \mathbb{R}^{n-1} : \tfrac{\partial}{\partial y} \psi(t_0, z_0) u = 0 \right\}$$

and let $r \geq 1$ denote the dimension of this subspace. As always, the curves have the parametrisation (9), and so for each y we choose $\omega = \nabla_y \tilde{\varphi}(t_0, y)$ to make all of the curves meet at the bad point. Consider those directions y such that $y - z_0 \in U$. As before, the difference in tangent of the curves Γ_y and Γ_{z_0} at their intersection is $M(t_0)^{-1}[\psi(t_0, y) - \psi(t_0, z_0)]$. By the definition of the derivative we then have

$$\psi(t_0, y) - \psi(t_0, z_0) = \tfrac{\partial}{\partial y} \psi(t_0, z_0)(y - z_0) + O\big(|y - z_0|^2\big)$$
$$= O\big(|y - z_0|^2\big) \quad \text{since } y - z_0 \in U,$$

which is at most δ provided that $|y - z_0| \lesssim \sqrt{\delta}$. Pick a maximal $\sqrt{\delta}$-separated subset $\{y_j\}_{j=1}^M$ of these y. Then $M \sim \sqrt{\delta}^{-r}$, and each tube $T_j := T_y^\delta(\nabla_y \tilde{\varphi}(t_0, y))$ meets $T_0 := T_{z_0}^\delta(\nabla_y \tilde{\varphi}(t_0, z_0))$ at an angle of at most δ. So there is a cylinder of radius δ and length some small constant c which is included in all M of the tubes. Hence

$$\left\| \sum_{j=0}^M \mathbb{1}_{T_j} \right\|_{\frac{n+1}{n-1}} \geq M \big(\delta^{n-1} c \big)^{\frac{n-1}{n+1}}$$
$$\gtrsim \delta^{-\frac{n-1}{n+1} - \frac{r}{2(n+1)}} \big(\delta^{n-1} M \big)^{\frac{n}{n+1}}.$$

Hence, by the covering lemma (Lemma 5), we find that the $L^{\frac{n+1}{2}} \to L^{n+1}$ norm of $\mathcal{K}_\delta$ is at least $\delta^{-\frac{n-1}{n+1} - \frac{r}{2(n+1)}}$ which is greater than the estimate $\delta^{-\frac{n-1}{n+1}}$ obtained in the non-degenerate case. $\qquad\square$

Rather curiously, the non-degeneracy criterion is not required for the set dimension problem.

PROPOSITION 9. *For any phase ϕ of the form (8) satisfying (2) (but not necessarily (3)), the corresponding curved Kakeya sets have Hausdorff and Minkowski dimension at least $\frac{n+1}{2}$.*

Proof. This is intuitively clear, since a Kakeya set of degenerate curves includes a set of non-degenerate ones by simply removing slices around the "bad" heights. Shifting and scaling part of what remains so that it lies in the region $x_n \in [-1, 1]$ gives a set of curves that falls within the scope of Theorem 7. So this subset, and hence the whole set, of the original curves has Minkowski and Hausdorff dimension at least $\frac{n+1}{2}$. To prove this fact directly one merely needs to note that the conclusion of Lemma 6 is true when $\delta = d$ whether the intersection is tangential or not, and then follow the proof of Theorem 7. $\qquad\square$

3   Negative Results for Quadratic Curves

In the next three sections we look at the possibility of non-trivial results. Since we already know that these cannot hold of all classes of curves, from here onwards restrict our attention to simpler ones. Notice that in both of Bourgain's theorems the bad behaviour is caused by the presence of terms non-linear in x in the phase function. For this reason we now focus entirely on parabolic curves of the following form:

$$\Gamma_y(\omega) = \left\{ \begin{pmatrix} \omega - tAy - t^2By \\ t \end{pmatrix} : t \in [-1, 1] \right\}$$

where A and B are $(n-1) \times (n-1)$ real symmetric matrices. Kakeya questions about these curves arise from the phase

$$\varphi(x, y) := y^t x' + x_n \tfrac{1}{2} y^t A y + x_n^2 \tfrac{1}{2} y^t B y \qquad (12)$$

which is of the form (8), and includes the Restriction problem as the special case $B = 0$. Meanwhile, the phase

$$\varphi(x, y) = \tfrac{1}{2x_n} y^t (A + x_n B)^{-1} y - \tfrac{1}{x_n} y^t (A + x_n B)^{-1} x' \qquad (13)$$

is again of the form (8) and includes Bochner–Riesz as the case $B = 0$, but it gives rise to the same curves above but with y and ω exchanged. For this

reason it is now convenient to call the Kakeya maximal function arising from (13) a *Nikodym maximal function* and to denote it by

$$\mathcal{N}_\delta f(\omega) := \sup_{y \in \mathbb{B}^{n-1}} \frac{1}{|T_y(\omega)|} \int_{T_y(\omega)} |f(x)| \, dx \,,$$

where we now fix the curves $\Gamma_y(\omega)$ and tubes $T_y(\omega)$ to be as above. So $\Gamma_y(\omega)$ is always a parabola through the point $\binom{\omega}{0}$ whose direction is governed by y. Similarly, we define a *curved Nikodym set* to be one which includes a $\Gamma_y(\omega)$ for every $\omega \in \mathbb{B}^{n-1}$.

It is easy to check that both phases above satisfy the first of Hörmander's criteria (2); the second says we must assume that

$$\det(A + 2x_n B) \neq 0 \tag{14}$$

throughout the support of the cutoff a, which we will take to be $[-1,1]$, with the obvious deletion of a neighbourhood of $x_n = 0$ in the second case. Also, by applying linear maps to x and/or y in the phase, we see that the oscillatory integral problem is invariant under congruence of the matrices. Since they are symmetric, we may assume that one of them is diagonal, or even has only 0 and ± 1 on the diagonal. This is of limited help, but in the special case where one of A, B is positive-definite we are able to simultaneously diagonalise (that is, using change of variable we can make both A and B diagonal). This will enable us to perform certain computations that seem intractable in the general case.

At the level of the curves rather than the phases, we are free to multiply through by any invertible matrix. This allows us, if it is convenient, to replace A by I and B by $C := A^{-1}B$, so that the curves are now

$$\Gamma_y(\omega) = \left\{ \begin{pmatrix} \omega - ty - t^2 Cy \\ t \end{pmatrix} : t \in [-1,1] \right\} . \tag{15}$$

Note however that this matrix is not necessarily symmetric, nor is B assumed to be invertible. By a further transformation we may assume that C is in rational canonical form.

Note also that if B is a multiple of A (so $C = \lambda I$ say) then we can eliminate C altogether using the diffeomorphism $x_n + \lambda x_n^2 \mapsto x_n$. So the curved case only arises if C is not a multiple of I. Moreover, in the case $C = 0$, where the two phases above correspond to Restriction and Bochner–Riesz respectively, their corresponding maximal functions are related. This is because the transformation

$$(x', x_n) \mapsto \left(\tfrac{x'}{x_n}, \tfrac{1}{x_n} \right)$$

(which was first used by Carbery [C2] in showing that Restriction implies Bochner–Riesz) maps straight lines to straight lines: specifically, the line centred at $\binom{\omega}{0}$ in direction $\binom{y}{1}$ maps to the line with centre $\binom{y}{0}$ and direction $\binom{\omega}{1}$. This is why the Kakeya and Nikodym problems for straight lines are equivalent. Importantly, however, this transformation does *not* map parabolas to parabolas even if the roles of position and direction are exchanged. So there is no reason to expect $\mathcal{K}_\delta$ and $\mathcal{N}_\delta$ to satisfy the same bounds for a given matrix C, and in fact we shall see later that they do not.

Although very simple, these phases are general enough to exhibit many kinds of behaviour. Taking $C = \binom{0\ 0}{1\ 0}$ gives the "worst case" example of Theorem 1. More interestingly still, the generic failure criterion (6) of Theorem 2 has the simple form

$$C \text{ is not a multiple of } I.$$

Bourgain's proof of Theorem 2 is considerably simpler for these special curves, works in higher dimensions, and in fact gives a better bound in dimension 3, so we include the details here.

Theorem 10. *Suppose that the characteristic polynomial of C divided by its minimum polynomial consists of irreducible factors each of multiplicity at most k, where $0 \le k \le n - 2$. Then*

$$\| \mathcal{K}_\delta \|_{p \to 1} \gtrsim \delta^{\frac{1}{2n-3-k}(1-\frac{1}{p})-\frac{1}{p}}$$

for all p. If $k = 0$ and additionally $\operatorname{tr} \operatorname{adj} C = 0$, then this is strengthened to $\delta^{\frac{1}{2n-2}(1-\frac{1}{p})-\frac{1}{p}}$, while if $k = 0$ and $\operatorname{tr} \operatorname{adj} C = 0$ and $\det C = 0$ it is strengthened further to $\delta^{-1/p}$.

This applies only to the Kakeya problem with parabolas and not the Nikodym version, although if $n = 3$ then phases of the form (13) are covered by Bourgain's generic failure result (Theorem 2) after making changes of variable and expanding the $1/x_n$ as a power series to obtain the standard form (4).

Combining this with the implication of Proposition 4 gives the following partial answers to Questions 3 and 1:

COROLLARY 11. *If φ is of the form (12) with B not a multiple of A, then the desired estimate $\| \mathcal{K}_\delta \|_{p \to q} \lesssim \delta^{1-n/p}$ for the corresponding Kakeya maximal function is false for all $p > n - \frac{n-k-2}{2n-k-4}$ even with $q = \infty$, and the estimate $\| T_N \|_{r \to s} \lesssim N^{-n/s}$ is false for all $s < \frac{2n}{n-1} + \frac{2n-2k-2}{(2n-k-3)(n-1)(2n-3)}$ even with $r = 1$.*

If $k = 0$ and additionally $\operatorname{tr} \operatorname{adj} C = 0$, then these are strengthened to $p > n - \frac{n-1}{2n-3}$ and $s < \frac{2n}{n-1} + \frac{2n-2}{2(n-1)^2(2n-3)}$, while if $k = 0$ and $\operatorname{tr} \operatorname{adj} C = 0$ and $\det C = 0$ these are strengthened further to $p > n - 1$ and $s < \frac{2n}{n-1} + \frac{2}{(n-1)(2n-3)}$.

So we cannot achieve the optimal $p = n$, $s = \frac{2n}{n-1}$ unless $k = n - 2$, which means that the minimum polynomial of C is linear and so $C \parallel I$. Moreover, since a "generic" matrix has its characteristic and minimal polynomials equal, we can "usually" achieve no better than $p = n - \frac{1}{2}$ and $s = \frac{2n}{n-1} + \frac{1}{(n-1)(2n-3)}$.

Note that in three dimensions we are dealing with 2×2 matrices and so we always have $k = 0$ as long as $C \nparallel I$. So we cannot exceed the bound $p = \frac{5}{2} = \frac{n+2}{2}$, which for straight lines is due to Wolff [Wo1]. If additionally $\operatorname{tr} C = 0$ we cannot exceed $p = 7/3$, which was obtained for straight lines first by Bourgain [B1] and then by Schlag [S]. Their results have since been improved for straight lines, but the above theorem for curves suggests that $5/2$ and $7/3$ are "natural barriers" in the problem. The case $k = 0$ and $\operatorname{tr} \operatorname{adj} C = 0$ and $\det C = 0$ in dimension three corresponds precisely to Bourgain's "worst case" example of Theorem 1, and in fact there are analogues in higher dimensions as we shall see later.

The gain in three dimensions here compared with Theorem 2 is because of the absence of higher order terms. Their absence is also needed to make the proof work in dimensions 4 and above, since Bourgain's method of handling the general case in [B2] uses that the order of the terms neglected is equal to the dimension n.

Proof of Theorem 10. It is enough to show that we can choose suitable $\omega = \omega(y)$ to produce a set of curves that is too small. We will use a linear function: $\omega = Wy$. We claim that if we can make the determinant of the map $y \mapsto x' := Wy - ty - t^2 Cy$ of order $|t|^m$ for small t, then $\| \mathcal{K}_\delta \|_{p \to 1} \gtrsim \delta^{1/m - 1/p(1+1/n)}$ for all p.

Fix $t \in [-\delta^{1/m}, \delta^{1/m}]$ so that the determinant is at most δ. Then if y ranges over the ball in $\mathbb{R}^{n-1}$ of radius 1, we find that x' ranges over a set of measure at most δ, and we are interested in the size of the δ-neighbourhood of this. Now since the eigenvalues of the map are bounded, no side of the set can be larger than $|y| < 1$, but having all sides this large would exceed the maximum permitted volume. So the worst case has $n - 2$ sides of length 1 and one thin side of length δ so that the volume does not exceed that permitted by the determinant.

Hence the largest possible neighbourhood is of measure δ. Now allowing x_n to vary over the interval $[-\delta^{1/m}, \delta^{1/m}]$ gives us that the union E of these tubes of length $\delta^{1/m}$ has measure at most $\delta^{1+1/m}$. Observing that $\mathcal{K}_\delta \, \mathbb{1}_E(y) \geq \delta^{1/m}$ for all $y \in \mathbb{B}^{n-1}$ proves the claim.

So we must consider when the above condition on the determinant is satisfied for some $m \geq n$. Clearly if $C \parallel I$ then it cannot be, since the determinant is just the characteristic polynomial of W evaluated at $(t+\lambda t^2)$, but in all other cases we can simply write down a suitable W. By a change of variable we may assume that C is in rational canonical form; that is, $C = C_{p_0} \oplus \cdots \oplus C_{p_k}$ where k is as above, each C_{p_i} is the companion matrix of the polynomial p_i, p_k is the minimum polynomial of C, and p_i divides p_{i+1} for $i = 0, \ldots, k-1$.

We show that for an $l \times l$ companion matrix

$$
C = \begin{pmatrix}
c_1 & 1 & 0 & 0 & \ldots & 0 \\
c_2 & 0 & 1 & 0 & \ldots & 0 \\
c_3 & 0 & 0 & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
c_{l-1} & 0 & 0 & \ldots & 0 & 1 \\
c_l & 0 & 0 & \ldots & 0 & 0
\end{pmatrix}
$$

we can achieve $\det(W - tI - t^2 C) \lesssim t^{2l-1}$. Choose W to be zero except in the first column, whose elements are as follows:

$$
w_{1,1} = 0 \qquad\qquad w_{2,1} = -1 \qquad\qquad w_{i,1} = c_{i-2} \text{ for } i \geq 3.
$$

If we expand $\det(W - 2tI - 2t^2 C)$ down the first column we obtain

$$
(-1)^l \left[(c_1 t^2 + t)t^{l-1} - (c_2 t^2 + 1)t^l + \sum_{i=3}^{l} (-1)i - 1(c_i t^2 - c_{i-2})t^{l+i-2} \right]
$$

$$
= (-1)^l \left[c_1 t^{l+1} - c_2 t^{l+2} + \sum_{i=3}^{l} (-1)^{i-1} c_i t^{l+i} - \sum_{i=1}^{l-2} (-1)^{i-1} c_i t^{l+i} \right]
$$

$$
= c_{l-1} t^{2l-1} - c_l t^{2l} .
$$

This proves the result since there are $k + 1$ blocks each of order $l_i \times l_i$ with $\sum_i l_i = n - 1$, and so we take $m = 2(n-1) - k - 1$. If $k = 0$ then we have just one block of order $(n-1) \times (n-1)$. The conditions $\operatorname{tr} \operatorname{adj} C = 0$ and $\det C = 0$ correspond to $c_{n-2} = 0$ and $c_{n-1} = 0$, which allow us to take $m = 2(n-1)$ or $m = \infty$ respectively, giving the improvements stated. $\square$

4 Geometric Methods

In this section we will prove a result for the curved Kakeya maximal function of a particular class of curves which implies that the corresponding sets have dimension at least $\frac{n+2}{2}$. In the straight-line case this result is due to Wolff [Wo1] and uses geometric techniques, however, here we shall adapt a more recent proof due to Katz [K]. In both, the main geometric object is the *hairbrush*, a configuration of tubes which all pass through some central fixed one. Wolff's idea was that such configurations can be handled by grouping the tubes into planes all containing the central tube, and then by applying the known dimension 2 result for the Kakeya maximal function in each plane. Curves, however, cannot easily be grouped in this way, which is why we turned to Katz's work. His proof seems more elementary, in that it isolates the geometry showing that the main fact is that a triangle lies in a plane, and the remainder of the argument is a simple (but clever) splitting up of the linearised maximal function into bounded pieces.

Our result is the following:

Theorem 12. *The Nikodym maximal function $\mathcal{N}_\delta$ satisfies the bound*

$$\| \mathcal{N}_\delta \|_{n \to n} \lesssim \delta^{-\frac{n-2}{2n}}$$

provided that the curves under consideration are parabolas of the form (15) with $C^2 = 0$.

By Proposition 3 this implies that the Nikodym sets of these curves have Hausdorff and Minkowski dimension at least $\frac{n+2}{2}$.

The condition on the matrix C arises naturally in the proof as we shall see in Proposition 14. This class of curves seems to be particularly amenable to the proof methods that have been used in the straight-line case, since further results for these curves will be obtained by the arithmetic methods in section 5. However, the class is not equivalent to straight lines since, as we will see in section 6, the Kakeya conjecture fails completely for all these curves. Unfortunately it seems difficult to give the criterion $C^2 = 0$ any geometric interpretation.

We shall actually prove Theorem 12 for the linearised version of the Nikodym maximal function $\mathcal{LN}_\delta$ – recall its definition from page 1324: We have divided $\mathbb{R}^{n-1}$ into δ-cubes Q_j where j runs over $\mathbb{B}^{n-1} \cap \delta\mathbb{Z}^{n-1}$. To each index j we have an associated curved tube $T_j = T_{y_j}(\omega_j)$ where $\omega_j \in Q_j$ and y_j is arbitrary. Then

$$\mathcal{LN}_\delta f(\omega) = \sum_j \mathbb{1}_{Q_j}(\omega) \frac{1}{|T_j|} \int_{T_j} f(x) \, \mathrm{d}x \,.$$

Of course, we must seek bounds that are independent of the choice of the tubes. We shall also need to define related functions where the index set is specified:

$$\mathcal{L}\mathcal{N}_{\mathbb{A}}f(\omega) = \sum_{j \in \mathbb{A}} \mathbb{1}_{Q_j}(\omega)\frac{1}{|T_j|}\int_{T_j} f(x)\,\mathrm{d}x\,.$$

As in Wolff's approach, the main geometric object considered is the *hairbrush*:

DEFINITION. Let $\mathbb{A}$ be a finite set of indices $j \in \delta\mathbb{Z}^{n-1}$. A *hairbrush* is a set $\mathbb{H} \subseteq \mathbb{A}$ such that there exists some curved $1 \times \delta$ tube T that intersects all T_i with $i \in \mathbb{H}$.

Note that the central tube T can be any curved tube of the family, not necessarily one of those associated to some j.

Much of the geometry of the situation is encoded in the behaviour of these hairbrushes, in the form of the following lemma:

LEMMA 13 (Hairbrush lemma). *If the curves are parabolas with $C^2 = 0$, then for all hairbrushes $\mathbb{H}$ we have $\|\mathcal{L}\mathcal{N}_{\mathbb{H}}\|_{n \to n} \leq C(\log 1/\delta)^{\alpha}$ for some constant α*

The proof of this will involve surfaces, and will show why we are able to handle only a restricted class of curves. But given the lemma, we can prove the theorem just as in the straight-line case, by splitting up the operator into many sums.

Proof of Theorem 12. As usual it is enough to prove a weak type estimate. By the covering lemma (Lemma 5) the theorem is true if and only if

$$\Big\| \sum_{j \in \mathbb{A}} \mathbb{1}_{T_j} \Big\|_{n'} \lesssim \delta^{-\frac{n-2}{2n}}(\delta^{n-1}\#\mathbb{A})^{1/n'}$$

$$\Longleftrightarrow \int_{\mathbb{R}^n} \Big(\sum_{j \in \mathbb{A}} \mathbb{1}_{T_j}(x)\Big)^{\frac{n}{n-1}} \mathrm{d}x \lesssim \delta^{-\frac{n-2}{2(n-1)}}(\delta^{n-1}\#\mathbb{A})$$

$$\Longleftrightarrow \sum_{j \in \mathbb{A}} \frac{1}{|T_j|} \int_{T_j} \Big(\sum_{i \in \mathbb{A}} \mathbb{1}_{T_i}(x)\Big)^{\frac{1}{n-1}} \mathrm{d}x \lesssim \delta^{-\frac{n-2}{2(n-1)}}\#\mathbb{A}\,.$$

Denote the quantity appearing in the first sum by $M_{\mathbb{A}}$, that is

$$M_{\mathbb{A}}(j) := \frac{1}{|T_j|}\int_{T_j} \Big(\sum_{i \in \mathbb{A}} \mathbb{1}_{T_i}(x)\Big)^{\frac{1}{n-1}} \mathrm{d}x.$$

We would like to subdivide this quantity into dyadic scales, by considering those i that are at distance between 2^{-k} and $2^{-(k+1)}$ from j. Note that by elementary properties of sequences of positive reals, the sum over k can

then be pulled out of the integral. But what remains then depends only on pairs i, j with $|i - j| \sim 2^{-k}$. Let P be a cube in $\mathbb{R}^{n-1}$ of side 10×2^{-k}. Note that this cube is larger than the cubes Q_j since $\delta < 2^{-k}$. It then suffices, for every choice of P, to obtain the estimate

$$\sum_{j \in \mathbb{A} \cap P} M_k(j) := \sum_{j \in \mathbb{A} \cap P} \frac{1}{|T_j|} \int_{T_j} \Big(\sum_{\substack{i \in \mathbb{A} \cap P \\ |i-j| \sim 2^{-k}}} \mathbb{1}_{T_i}(x) \Big)^{\frac{1}{n-1}} \, \mathrm{d}x$$

$$\lesssim \delta^{-\frac{n-2}{2(n-1)}} \#(\mathbb{A} \cap P)$$

and then sum over P and k. Both sums have only logarithmically many terms.

The next stage is to find as many large hairbrushes in $\mathbb{A} \cap P$ as possible, where *large* means of cardinality at least N, to be chosen later. So, if there exists some curved tube T (of the form $T_y(\omega)$ but not necessarily one of the T_j) such that there are at least N elements $i \in \mathbb{A} \cap P$ with $T \cap T_i \neq \emptyset$, then call these elements $\mathbb{H}_1$. Then look for another large hairbrush in the remaining elements $\mathbb{A} \cap P \setminus \mathbb{H}_1$. Eventually there are no more hairbrushes, so call the remaining bad elements $\mathbb{B}$. This constructs hairbrushes $\mathbb{H}_1, \ldots, \mathbb{H}_m$ each of cardinality at least N, and a bad set $\mathbb{B} := \mathbb{A} \setminus (\mathbb{H}_1 \cup \cdots \cup \mathbb{H}_m)$. Let $\mathbb{H} := \mathbb{H}_1 \cup \cdots \cup \mathbb{H}_m$.

Since the hairbrushes are disjoint sets of indices (although the tubes they correspond to may well not be), and $\mathbb{A} \cap P$ has at most $2^{-k(n-1)} \delta^{-(n-1)}$ elements, it follows that $m \leq 2^{-k(n-1)} \delta^{-(n-1)}/N$.

Now split the sum into four pieces

$$\sum_{j \in \mathbb{A} \cap P} M_k(j) \leq \sum_{j \in \mathbb{H}} M_{k,\mathbb{H}}(j) + \sum_{j \in \mathbb{B}} M_{k,\mathbb{H}}(j) + \sum_{j \in \mathbb{H}} M_{k,\mathbb{B}}(j) + \sum_{j \in \mathbb{B}} M_{k,\mathbb{B}}(j)$$

where

$$M_{k,\mathbb{H}}(j) := \frac{1}{|T_j|} \int_{T_j} \Big(\sum_{\substack{i \in \mathbb{H} \\ |i-j| \sim 2^{-k}}} \mathbb{1}_{T_i}(x) \Big)^{\frac{1}{n-1}} \, \mathrm{d}x$$

$$M_{k,\mathbb{B}}(j) := \frac{1}{|T_j|} \int_{T_j} \Big(\sum_{\substack{i \in \mathbb{B} \\ |i-j| \sim 2^{-k}}} \mathbb{1}_{T_i}(x) \Big)^{\frac{1}{n-1}} \, \mathrm{d}x.$$

The first sum is estimated using the hairbrush lemma. For

$$\|\mathcal{L}\mathcal{N}_{\mathbb{H}} f\|_n^n = \int \Big(\sum_i \mathcal{L}\mathcal{N}_{\mathbb{H}_i} f(\omega) \Big)^n \, \mathrm{d}\omega$$

$$= \sum_i \int \big(\mathcal{L}\mathcal{N}_{\mathbb{H}_i} f(\omega) \big)^n \, \mathrm{d}\omega \text{ since each } \omega \text{ gives only one non-zero term}$$

$$\leq m\|\mathcal{L}\mathcal{N}_{\mathbb{H}_i}\|_{n\to n}^n\|f\|_n^n\,,$$

showing that $\|\mathcal{L}\mathcal{N}_{\mathbb{H}}\|_{n\to n} \leq Cm^{1/n}(\log 1/\delta)^\alpha$ by Lemma 13. Then by the covering lemma we obtain

$$\sum_{j\in\mathbb{H}} M_{k,\mathbb{H}}(j) \leq (Cm^{1/n}(\log 1/\delta)^\alpha)^{n'}\#\mathbb{H}\,.$$

For the second sum

$$\sum_{j\in\mathbb{B}} M_{k,\mathbb{H}}(j) = \sum_{j\in\mathbb{B}} \frac{1}{|T_j|} \int_{T_j} \Big(\sum_{\substack{i\in\mathbb{H}\\|i-j|\sim 2^{-k}}} \mathbb{1}_{T_i}(x) \Big)^{\frac{1}{n-1}} \mathrm{d}x$$

$$\leq \sum_{j\in\mathbb{B}} \Big(\frac{1}{|T_j|} \int_{T_j} \sum_{\substack{i\in\mathbb{H}\\|i-j|\sim 2^{-k}}} \mathbb{1}_{T_i}(x)\,\mathrm{d}x \Big)^{\frac{1}{n-1}} \qquad \text{by Jensen}$$

$$\leq (\#\mathbb{B})^{1-\frac{1}{n-1}} \Big(\sum_{j\in\mathbb{B}} \frac{1}{|T_j|} \int_{T_j} \sum_{\substack{i\in\mathbb{H}\\|i-j|\sim 2^{-k}}} \mathbb{1}_{T_i}(x)\,\mathrm{d}x \Big)^{\frac{1}{n-1}} \quad \text{by Hölder}$$

$$\leq (\#\mathbb{B})^{1-\frac{1}{n-1}} \Big(\sum_{i\in\mathbb{H}} \sum_{\substack{j\in\mathbb{B}\\|i-j|\sim 2^{-k}}} \frac{1}{|T_j|}|T_i \cap T_j| \Big)^{\frac{1}{n-1}} \text{by swapping sums}$$

$$\lesssim (\#\mathbb{B})^{1-\frac{1}{n-1}} \Big(\sum_{i\in\mathbb{H}} \sum_{\substack{j\in\mathbb{B}\\T_i\cap T_j\neq\emptyset}} \frac{1}{\delta^{n-1}}\delta^n 2^k \Big)^{\frac{1}{n-1}} \qquad \text{by Lemma 6}$$

$$\leq (\#\mathbb{B})^{1-\frac{1}{n-1}} \Big(\sum_{i\in\mathbb{H}} N\delta 2^k \Big)^{\frac{1}{n-1}} \quad \text{since no large hairbrushes in } \mathbb{B}$$

$$= (\#\mathbb{B})^{1-\frac{1}{n-1}} (\#\mathbb{H}N\delta 2^k)^{\frac{1}{n-1}}$$

$$\leq (\#\mathbb{A}\cap P)\big(N\delta 2^k\big)^{\frac{1}{n-1}}.$$

The third and fourth sums can be tackled together, since for *all $j \in \mathbb{A}\cap P$* we have

$$2M_{k,\mathbb{B}}(j) := \frac{1}{|T_j|} \int_{T_j} \Big(\sum_{\substack{i\in\mathbb{B}\\|i-j|\sim 2^{-k}}} \mathbb{1}_{T_i}(x) \Big)^{\frac{1}{n-1}} \mathrm{d}x$$

$$\leq \Big(\frac{1}{|T_j|} \int_{T_j} \sum_{\substack{i\in\mathbb{B}\\|i-j|\sim 2^{-k}}} \mathbb{1}_{T_i}(x)\,\mathrm{d}x \Big)^{\frac{1}{n-1}} \qquad \text{by Jensen}$$

$$= \left(\frac{1}{|T_j|} \sum_{\substack{i \in \mathbb{B} \\ |i-j| \sim 2^{-k}}} |T_i \cap T_j| \right)^{\frac{1}{n-1}}$$

$$\lesssim \left(\frac{1}{\delta^{n-1}} \sum_{\substack{i \in \mathbb{B} \\ T_i \cap T_j \neq \emptyset}} \delta^n 2^k \right)^{\frac{1}{n-1}} \qquad \text{by Lemma 6}$$

$$\leq \left(N \delta 2^k \right)^{\frac{1}{n-1}} \qquad \text{since no large hairbrushes in } \mathbb{B}.$$

So the last three sums all give an estimate of $\#(\mathbb{A} \cap P)(N\delta 2^k)^{\frac{1}{n-1}}$, while the first, after putting in the upper bound for m, gives $\#\mathbb{H}(2^{-k(n-1)}\delta^{-(n-1)}/N)^{\frac{1}{n-1}}$. We optimally choose $N = 2^{-k\frac{n}{2}}\delta^{-\frac{n}{2}}$ and add the four pieces to obtain

$$\sum_{j \in \mathbb{A} \cap P} M_k(j) \lesssim 2^{-k\frac{n-2}{2(n-1)}} \delta^{-\frac{n-2}{2(n-1)}} \#(\mathbb{A} \cap P)$$

which gives the result after summing over all P of side 2^{-k} and all k. $\qquad \square$

To prove the hairbrush lemma for curves, we need an analogue of the following fact about straight lines:

> Two intersecting straight lines determine a plane, and thus a
> third line intersecting these two is fixed up to one parameter,
> i.e. its direction must lie parallel to the plane, or equivalently
> the point where it meets the base plane $x_n = 0$ must lie along
> a fixed line.

So we must now study the locus of all curves meeting two given ones.

By the linearity of (15) in y and ω, we may assume that one of the given curves is $\Gamma_0(0)$. Let the other be $\Gamma_{y_0}(\omega_0)$ and assume that they meet at height t_0. The surface is the locus of those curves $\Gamma_y(\omega)$ that meet the first at s and the second at u. Note that none of these three heights are equal, since the curves are never horizontal, and we must exclude the possibility of $\Gamma_y(\omega)$ meeting the two given curves at their common point, since this would allow every curve to belong to the locus. This is made clearer by the following picture (Figure 4).

Now we have the following equations:

$$0 = \omega_0 - t_0 y_0 - t_0^2 C y_0 \tag{16}$$

$$0 = \omega - sy - s^2 C y \tag{17}$$

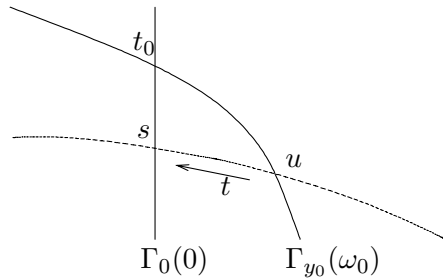$$\omega_0 - u y_0 - u^2 C y_0 = \omega - uy - u^2 C y \tag{18}$$

Figure 4: Notation for a curved triangle

Subtracting (17) from (18) we find that

$$y = \frac{1}{s-u}\big(I + (s+u)C\big)^{-1}\big(\omega_0 - uy_0 - u^2Cy_0\big)$$

which is well defined because $s \neq u$ and by (14). Substitute into (17) to find ω:

$$\omega = \frac{s}{s-u}\big(I + sC\big)\big(I + (s+u)C\big)^{-1}\big(\omega_0 - uy_0 - u^2Cy_0\big).$$

(16) has not been used yet, so we use it to eliminate ω_0:

$$\omega = \frac{s(t_0-u)}{s-u}\big(I + sC\big)\big(I + (s+u)C\big)^{-1}\big(I + (t_0+u)C\big)y_0.$$

Finally substitute this y and ω into (15) to obtain

$$\begin{pmatrix} \frac{(s-t)(t_0-u)}{s-u}\big(I + (s+t)C\big)\big(I + (s+u)C\big)^{-1}\big(I + (t_0+u)C\big)y_0 \\ t \end{pmatrix} \quad (19)$$

as the parametrisation of the locus we are interested in. Note that if $C = 0$ then this reduces to the plane $\big(\begin{smallmatrix} ry_0 \\ t \end{smallmatrix}\big)$ as expected. In general however, there are three parameters (u, s, t) and so the locus is not a plane nor even a surface but rather some fat object. What we need to know is whether a curve belonging to this locus has its direction and/or its base point fixed up to one parameter. The following proposition determines when this is so.

PROPOSITION 14.   *Suppose that the curve $\Gamma_y(\omega)$ is included in the locus (19). Then*

  (i) *The point ω must belong to a family described by only one parameter if and only if $C \parallel I$ or $C^2 = 0$, in which cases*

$$\omega = r(I + t_0C)y_0 \quad (20)$$

     for some parameter r.

  (ii) *The direction y must belong to a family described by only one parameter if and only if we have the straight-line case $C \parallel I$.*

Proof. From

$$\omega - t(I + tC)y = \frac{(s-t)(t_0 - u)}{s - u}\left(I + (s+t)C\right)\left(I + (s+u)C\right)^{-1}\left(I + (t_0 + u)C\right)y_0$$

we obtain

$$\omega = \frac{s(t_0 - u)}{s - u}(I + sC)\left(I + (s+u)C\right)^{-1}\left(I + (t_0 + u)C\right)y_0$$

$$y = \frac{t_0 - u}{s - u}\left(I + (s+u)C\right)^{-1}\left(I + (t_0 + u)C\right)y_0$$

where for a given curve, s and u will be fixed. However, we are considering the sets of all such y and ω, so we allow s and u to vary. We also require the property for all y_0 and t_0.

(i) To show that the locus of all ω is one-dimensional we require that the derivatives of ω with respect to s and u are always parallel. These are

$$\frac{t_0 - u}{s - u}\left[s\left(C - (I + sC)(I + (s+u)C)^{-1}C\right)\right.$$

$$\left. - \frac{u}{s - u}(I + sC)\right]\left(I + (s+u)C\right)^{-1}\left(I + (t_0 + u)C\right)y_0$$

and

$$\frac{s}{s - u}(I + sC)\left(I + (s+u)C\right)^{-1}\left[\frac{t_0 - s}{s - u}\left(I + (t_0 + u)C\right) + \right.$$

$$\left. (t_0 - u)\left(C - C(I + (s+u)C)^{-1}(I + (t_0 + u)C)\right)\right]y_0.$$

We need this for all y_0, so that in fact the matrices themselves must be "parallel", by which we mean that one is a scalar multiple of the other. Next we may rewrite the above, but ignore the initial (scalar) function of (s, u) and multiply on the left by $\left(I + (s+u)C\right)\left(I + sC\right)^{-1}$ and on the right by $\left(I + (t_0 + u)C\right)^{-1}\left(I + (s+u)C\right)$. We thus require the following two expressions to be parallel:

$$\frac{1}{s - u}\left(I + (s+u)C\right) - sC\left(I + sC\right)^{-1}C$$

$$\frac{1}{s - u}\left(I + (s+u)C\right) - (t_0 - u)C\left(I + (t_0 + u)C\right)^{-1}C$$

Setting $s = 0$ and $u = -t_0$ we find that $I - t_0C$, which is invertible, is parallel to $I - t_0C - 2t_0C^2$. This implies that either $C \parallel I$ or $C^2 = 0$.

(ii) For y, the two derivatives are

$$\frac{t_0 - u}{s - u}\left(I + (s+u)C\right)^{-1}\left[-C\left(I + (s+u)C\right)^{-1} - \frac{1}{s - u}I\right]\left(I + (t_0 + u)C\right)$$

$$\frac{1}{s-u}\bigl(I+(s+u)C\bigr)^{-1}\left[\frac{t_0-s}{s-u}\bigl(I+(t_0+u)C\bigr)\right.$$

$$\left.+(t_0-u)C\bigl(I-\bigl(I+(s+u)C\bigr)^{-1}\bigl(I+(t_0+u)C\bigr)\bigr)\right].$$

Ignoring scalar functions and multiplying by invertible matrices on the right and left, we thus require the following two expressions to be parallel:

$$I+2sC$$

$$I+(s+u)C-(t_0-u)(s-u)C\bigl(I+(t_0+u)C\bigr)^{-1}C$$

Setting $u=-t_0$ and subtracting gives

$$I+2sC \parallel (s+t_0)C\left[I+2t_0C\right]$$

and since $I+2sC$ and $I+2t_0C$ are invertible by (14), we can deduce that C is a (possibly zero) multiple of I.

In order to convince ourselves, we check that if $C^2=0$, then y is given by $\frac{2(t_0-u)}{s-u}\bigl(I+(t_0-s)C\bigr)y_0$, which does have two parameters unless $C\parallel I$. $\square$

This result clearly shows that for parabolas, the Kakeya and Nikodym versions of the problem are not the same at all. Indeed, the "worst case" example of Bourgain had $C^2=0$, and we already know that no non-trivial Kakeya estimate can hold for this.

We are now ready to prove the hairbrush lemma, and hence complete the proof of Theorem 12.

Proof of Lemma 13. We have a set $\mathbb{H}$ of indices which forms a hairbrush with central tube T. By linearity assume that the central tube is $T_0(0)$. Denote the other tubes by $T_j = T_{y_j}(\omega_j)$, where $\omega_j \in Q_j$ and so $\omega_j \approx j$. We partition the set $\mathbb{H}$ in several ways. First, let H_k be the set of all those indices whose tubes meet T at "angle" 2^{-k}; that is,

$$H_k := \left\{i \in \mathbb{H} : |y_i| \sim 2^{-k}\right\}.$$

For a fixed ω, there can be only one k such that $\mathcal{LN}_{H_k}(\omega) \neq 0$, so it is enough to prove $\|\mathcal{LN}_{H_k}\| \leq C(\log 1/\delta)^\alpha$ since there are only logarithmically many k. Then by the arguments used previously, this bound is true if and only if

$$\sum_{i \in H_k} M_{H_k}(i) \leq C(\log 1/\delta)^\alpha \#H_k.$$

For fixed $j \in H_k$ split up H_k into further sets $H_{j,k,l,m}$ as follows:

$$H_{j,k,l,m} := \left\{i \in H_k : |y_i - y_j| \sim 2^{-l} \text{ and } \operatorname{dist}(T_i \cap T_j, T_j \cap T) \sim \delta 2^{l+m}\right\}.$$

Note that this set is empty unless $l \geq k - 2$. Now it is enough to show that $M_{H_{j,k,l,m}}(j) \leq C(\log 1/\delta)^{\alpha}$, because

$$M_{H_k}(i) \leq \sum_{j \in H_k} \sum_l \sum_m M_{H_{j,k,l,m}}(i)$$

and there are only logarithmically many l and m and the sum over j introduces a factor $\#H_k$.

Next comes the geometric part of the argument, which is a quantitative version of the fact explained on page 1338. We need to show that $\#\mathbb{H}_{j,k,l,m}$ is not too big, which means that given the central tube T and another fixed tube $T_j : j \in H_k$ there are few other tubes T_i meeting these with all the correct "angles" and distances. In the straight-line case this follows from simple consideration of similar triangles as in Figure 5. In the curved
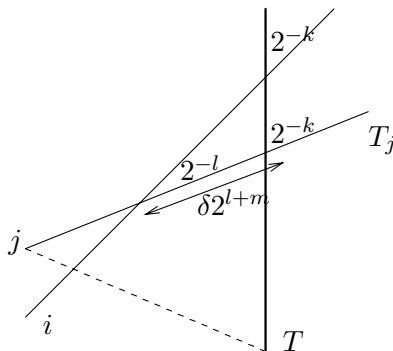

Figure 5: In the straight-line case, by similar triangles we have $|i - j| \lesssim 2^{-l}$ and $\mathrm{dist}(i, \mathrm{line}) \lesssim 2^{-(l+m)}$.

case, the dotted line in the picture is instead the line (20), which is the intersection of the base plane $x_n = 0$ with the surface (19) determined by T and T_j. Since we cannot appeal to similar triangles with curves, we state and prove our claim more formally:

CLAIM. *Let C satisfy $C^2 = 0$. Suppose that we are given three curved tubes $T = T_0(0)$, $T_j = T_{y_j}(\omega_j)$ and $T_i = T_{y_i}(\omega_i)$ with $|y_j|, |y_i| \in \left(\frac{1}{2^{k+1}}, \frac{1}{2^k}\right)$ and $|y_j - y_i| \in \left(\frac{1}{2^{l+1}}, \frac{1}{2^l}\right)$. Here $l \geq k - 2$ and all the powers of 2 that occur are greater than δ. Suppose that T_j meets the axis at height t_j, T_i meets it at t_i, and they meet each other at s, where $\delta 2^{l+m} \leq |s - t_j| \leq \delta 2^{l+m+1}$. Then $|\omega_j - \omega_i| \leq \frac{1}{2^l}$, and ω_i is at distance at most $\frac{1}{2^{l+m}}$ from the line (20).*

We have the following equations:

$$\omega_j = t_j(I + t_jC)y_j$$
$$\omega_i = t_i(I + t_iC)y_i + \varepsilon$$
$$\omega_i - s(I + sC)y_i = \omega_j - s(I + sC)y_j + \eta$$

where ε and η are errors due to the thickness of the tubes, and are of order at most δ. The first assertion is easy:

$$\begin{aligned}
|\omega_j - \omega_i| &= \big|2s(I + sC)(y_j - y_i) - \eta\big| \\
&\leq C|y_j - y_i| + |\eta| \\
&\lesssim \tfrac{1}{2^l}.
\end{aligned}$$

For the second, begin by eliminating the ωs:

$$(t_j - s)\big(I + (t_j + s)C\big)y_j + \eta = (t_i - s)\big(I + (t_i + s)C\big)y_i + \varepsilon. \qquad (21)$$

This can be rearranged to give y_i in terms of y_j. Substituting back into the 2nd of our original equations gives

$$\begin{aligned}
\omega_i &= \frac{t_i(t_j - s)}{t_i - s}(I + t_iC)\big(I + (t_i + s)C\big)^{-1}\big(I + (t_j + s)C\big)y_j \\
&\quad + \frac{t_i}{t_i - s}(I + t_iC)\big(I + (t_i + s)C\big)^{-1}(\eta - \varepsilon) + \varepsilon \\
&= \frac{t_i(t_j - s)}{t_i - s}(I + t_jC)y_j + \frac{t_i}{t_i - s}(I - sC)(\eta - \varepsilon) + \varepsilon
\end{aligned}$$

where we have used the fact that $C^2 = 0$. Looking back at (20) we discover that the first term belongs to the intersection of the surface determined by the first two curves with the horizontal plane. So the distance we are interested in is at most the absolute value of the other two terms, so at most $\frac{C}{|t_i - s|}\delta + \delta$. Finally we just ensure that $|t_i - s|$ is comparable to $|t_j - s|$. From (21) using the fact that on $\mathrm{supp}(a)$ the eigenvalues of $I + x_nC$ are bounded above and below, we get

$$\begin{aligned}
|t_i - s|2^{-k} &\gtrsim \delta 2^{l+m}2^{-k} - \delta \\
|t_i - s| &\gtrsim \delta 2^{l+m} - \delta 2^k \\
&\gtrsim \delta 2^{l+m}
\end{aligned}$$

provided that $k - l + m$ is not too large. Since $l \geq k - 2$ this could happen only with l close to k and m small, in which case the claim is trivial anyway. So the distance of ω_i from the curve of intersection is at most $2^{-(l+m)}$ and we have proved the claim.

We can now complete the proof of the hairbrush lemma, and hence the whole theorem. The claim tells us that

$$\#H_{j,k,l,m} \lesssim 2^{-l}\big(2^{-(l+m)}\big)^{n-2}\delta^{n-1}$$

which we use as follows:

$$M_{H_{j,k,l,m}}(i) := \frac{1}{|T_i|} \int_{T_i} \Big(\sum_{p \in H_{j,k,l,m}} \mathbb{1}_{T_p}(x) \Big)^{\frac{1}{n-1}} \mathrm{d}x$$

$$= \frac{1}{|T_i|} \int_{\{x \in T_i : \mathrm{dist}(x, T_i \cap T) \sim \delta 2^{l+m}\}} \Big(\sum_{p \in H_{j,k,l,m}} \mathbb{1}_{T_p}(x) \Big)^{\frac{1}{n-1}} \mathrm{d}x$$

$$\text{by defn. of } H_{j,k,l,m}$$

$$\leq \frac{1}{|T_i|} \big| \{x \in T_i : \mathrm{dist}(x, T_i \cap T) \sim \delta 2^{l+m}\} \big|^{1 - \frac{1}{n-1}}$$

$$\cdot \Big(\int_{T_i} \sum_{p \in H_{j,k,l,m}} \mathbb{1}_{T_p}(x) \Big)^{\frac{1}{n-1}} \mathrm{d}x \qquad \text{by Hölder}$$

$$\leq \frac{1}{\delta^{n-1}} (\delta^n 2^{l+m})^{1 - \frac{1}{n-1}} \Big(\sum_{p \in H_{j,k,l,m}} |T_i \cap T_p| \Big)^{\frac{1}{n-1}}$$

$$\lesssim \frac{1}{\delta^{n-1}} (\delta^n 2^{l+m})^{1 - \frac{1}{n-1}} \big(\#H_{j,k,l,m} \delta^n 2^l \big)^{\frac{1}{n-1}} \qquad \text{by Lemma 6}$$

$$\lesssim \frac{1}{\delta^{n-1}} (\delta^n 2^{l+m})^{1 - \frac{1}{n-1}} \big(2^{-l} (2^{-(l+m)})^{n-2} \delta^{n-1} \delta^n 2^l \big)^{\frac{1}{n-1}}$$

$$\text{by the claim}$$

$$= 1 \,.$$

Summing over all the index sets gives the result. $\qquad \square$

5 Arithmetic Methods

5.1 Introduction. Sections 2 and 4 showed how geometric methods could give lower bounds for the set dimension of the form $\frac{n}{2} + \mathrm{const}$. The best-known results in the straight-line case in low dimensions ($n = 3$ or 4) are still of this form [KLT], [LT], but in higher dimensions far better results are obtained by an arithmetic approach, since these improve the coefficient of n to something greater than $1/2$.

The arithmetic arises in the form of *sumset inequalities*. For these we require some notation.

NOTATION. Let $\mathcal{A}, \mathcal{B} \subseteq \mathbb{Z}^{n-1}$ be finite sets and let $\mathcal{G} \subseteq \mathcal{A} \times \mathcal{B}$. For any $(n-1) \times (n-1)$ real matrix X define the X-sumset of $\mathcal{A}$ and $\mathcal{B}$ by

$$\mathcal{A} + X\mathcal{B} := \big\{ a + Xb : (a,b) \in \mathcal{G} \big\} \,.$$

In the case $X = -I$ write $\mathcal{A} - \mathcal{B}$ and call it the difference set.

The structure of sumsets, and inequalities regarding the relative sizes of sum and difference sets, have been extensively studied by combinatorialists when the matrix X is an integer multiple of the identity, but they have generally considered only $\mathcal{G} = \mathcal{A} \times \mathcal{B}$. See [N], [R]. The link with the Kakeya problem was noticed in 1999 by Bourgain [B3], and since then many inequalities with $\mathcal{G} \subseteq \mathcal{A} \times \mathcal{B}$ have been proved. However, the case where X is not a multiple of I arises only with curves, and seems to be a new problem.

The most general matrix sumset problem with $N + 2$ "slices" is as follows:

QUESTION 4. *Let $X_1, X_2, \ldots, X_N$ be $(n-1) \times (n-1)$ real matrices. To avoid trivialities assume that they are non-zero, distinct, and not equal to $-I$. Does there exist an $\varepsilon > 0$ depending only on the X_j's such that for all $\mathcal{A}, \mathcal{B} \in \mathbb{Z}^{(n-1)}$, $\mathcal{G} \subseteq \mathcal{A} \times \mathcal{B}$ we have*

$$\#(\mathcal{A} - \mathcal{B}) \leq \max\left\{ \#\mathcal{A}, \#\mathcal{B}, \max_j \#(\mathcal{A} + X_j \mathcal{B}) \right\}^{2-\varepsilon} ?$$

If so, what is the largest possible ε?

The idea is to let $\mathcal{A}$ and $\mathcal{B}$ correspond to two horizontal slices through our δ-discretised Kakeya set, and $\mathcal{G} \subseteq \mathcal{A} \times \mathcal{B}$ the set of all pairs which are joined by a line. Then the difference set $\mathcal{A} - \mathcal{B}$ corresponds to the set of directions, so must be large. However, if our set had small dimension then $\mathcal{A}$ and $\mathcal{B}$ must be small, and moreover the set of midpoints of the lines, which (for straight lines) has cardinality $\#(\mathcal{A} + \mathcal{B})$, must also be small. Inequalities regarding the relative sizes of sumsets and difference sets thus lead to lower bounds for the dimension of Kakeya sets.

In the curved case, we must discover how to determine from two endpoints the location of any other point on the curve, and its direction.

LEMMA 15. *Let E be a set of curves of the form (15) and let $\mathcal{A}, \mathcal{B} \subset \mathbb{Z}^{n-1}$ be the (δ-discretised) intersections of $\mathrm{nbd}_\delta(E)$ with the planes $x_n = t_0$ and $x_n = t_1$ respectively, where $t_0 \neq t_1$. Let*

$$\mathcal{G} := \left\{ (a, b) : a \text{ and } b \text{ lie on the same tube in } \mathrm{nbd}_\delta(E) \right\} \subseteq \mathcal{A} \times \mathcal{B} .$$

Then

(i) *The set of directions y has the same cardinality as the difference set $\mathcal{A} - \mathcal{B}$.*

(ii) *Assume that $t_0, t_1 \neq 0$. The set of centres ω has the same cardinality as $\mathcal{A} - T\mathcal{B}$ where T is the $(n-1) \times (n-1)$ matrix*
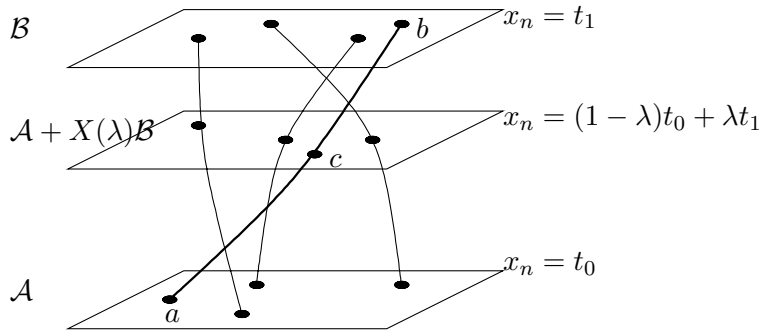
$$T = \tfrac{t_0}{t_1}(I + t_0 C)(I + t_1 C)^{-1}.$$

Figure 6: Slices through a curved Kakeya set

(iii) *The intersection of the set with the plane $x_n = (1 - \lambda)t_0 + \lambda t_1$ has the same cardinality as the sumset $\mathcal{A} + X(\lambda)\mathcal{B}$, where $X(\lambda)$ is the $(n-1) \times (n-1)$ matrix*

$$X(\lambda) = \frac{\lambda}{1 - \lambda}\left[I + \lambda(t_1 - t_0)C(I + (t_0 + t_1)C)^{-1}\right]^{-1}$$
$$\cdot \left[I - (1 - \lambda)(t_1 - t_0)C(I + (t_0 + t_1)C)^{-1}\right].$$

Proof. Consider a curve through the points (a, t_0) and (b, t_1). The equation (15) of the curves gives

$$a = \omega - t_0 y - t_0^2 C y \tag{22}$$
$$b = \omega - t_1 y - t_1^2 C y. \tag{23}$$

Subtracting these we find that $y = \frac{1}{t_0 - t_1}(I + (t_0 + t_1)C)^{-1}(b - a)$, and so the first assertion follows, since multiplication by an invertible matrix does not change the cardinality. This part is analogous to Lemma 6.

If we write $M_j = t_j(I + t_j C)$ for $j = 0, 1$ so that $a = \omega - M_0 y$ and $b = \omega - M_1 y$, then solving gives

$$\omega = (M_0^{-1} - M_1^{-1})^{-1}(M_0^{-1}a - M_1^{-1}b).$$

We can always multiply through by an invertible matrix to get the a on its own. Therefore an appropriate "difference set" is $\mathcal{A} - T\mathcal{B}$ where $T = M_0 M_1^{-1} = \frac{t_0}{t_1}(I + t_0 C)(I + t_1 C)^{-1}$ as in the second assertion. Note that for the Nikodym problem, we cannot take slices through $x_n = 0$ anyway, because these sets arise only when $x_n = 0$ is not in the support of the cutoff function in (1).

For the third, denote the point of intersection of this curve with the intermediate plane by c. It helps to take $(1 - \lambda)(22) + \lambda(23)$, which gives

$$\omega = (1 - \lambda)a + \lambda b + (1 - \lambda)(t_0 y + t_0^2 C y) + \lambda(t_1 y + t_1^2 C y).$$

This allows lots of cancellation, so that

$$
\begin{aligned}
c &= \omega - \big((1-\lambda)t_0 + \lambda t_1\big)y - \big((1-\lambda)t_0 + \lambda t_1\big)^2 Cy \\
&= (1-\lambda)a + \lambda b + \lambda(1-\lambda)(t_0 - t_1)^2 Cy \\
&= (1-\lambda)a + \lambda b + \lambda(1-\lambda)(t_0 - t_1)^2 C \frac{1}{t_0 - t_1}\big(I + (t_0 + t_1)C\big)^{-1}(b - a) \\
&= \big[(1-\lambda)I + \lambda(1-\lambda)(t_1 - t_0)C(I + (t_0 + t_1)C)^{-1}\big]a \\
&\qquad\qquad + \big[\lambda I - \lambda(1-\lambda)(t_1 - t_0)C(I + (t_0 + t_1)C)^{-1}\big]b\,.
\end{aligned}
$$

Multiplying through by an invertible matrix gives the result.   □

It is easy to check that all the matrices occurring above are indeed invertible, because of the non-degeneracy criterion (14). Recall also that in the straight-line case we have $C = 0$ and hence $X(\lambda)$ is really just a scalar. The sumset in the second assertion does not appear in the literature on the straight line problem since it is only appropriate when dealing with Nikodym rather than Kakeya sets. Although the elements of the matrices C and the parameters t_0, t_1, λ are real, for our application we should consider only matrices over $\mathbb{Q}$, since each real number may be approximated to within $O(\delta)$ by a rational, which corresponds to the same point in the δ-discretisation.

We now show how sumset inequalities imply results about Kakeya sets. For straight-line Kakeya sets this is due to Bourgain [B3], with the "plane-varying" improvement noticed by Katz and Tao [KT1].

LEMMA 16.   (i) *Suppose that for some matrix C we can choose $t_0, t_1 \in [-1, 1]$ and $\lambda_j \in (0, 1)$, $j = 1, \ldots, N$ such that Question 4 with $X_j = X(\lambda_j)$ has a positive answer. Then Kakeya sets of curves of the form (15) for this C have Minkowski dimension at least $\frac{n-1}{2-\varepsilon}$.*

(ii) *If the same holds but with $X_j = X(\lambda_j)T^{-1}$ where T is as in Lemma 15 and none of the heights $t_0, t_1, (1-\lambda_j)t_0 + \lambda_j t_1$ is 0, then the corresponding curved Nikodym sets have Minkowski dimension at least $\frac{n-1}{2-\varepsilon}$.*

In both cases, if in fact we have a range of solutions, meaning that Question 4 remains true as t_0 is allowed to vary over some small interval and the other heights to vary correspondingly, then we can obtain the better lower bound of $\frac{n-1}{2-\varepsilon} + 1$.

Proof. Let E, $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{G}$ be as in Lemma 15.

(i) If E is a curved Kakeya set, then we may assume that $\mathrm{nbd}_\delta(E)$ consists of $\delta^{-(n-1)}$ tubes in distinct δ-separated directions. So by

Lemma 15 we have $\delta^{-(n-1)} \sim \#(\mathcal{A} - \mathcal{B})$. By the assumption, this means that
$$\delta^{-(n-1)} \lesssim \max\left\{\#\mathcal{A}, \#\mathcal{B}, \max_j \#(\mathcal{A} + X(\lambda_j)\mathcal{B})\right\}^{2-\varepsilon}$$
and so one of the sets on the right-hand side has cardinality at least $\delta^{-\frac{n-1}{2-\varepsilon}}$. So $\mathrm{nbd}_\delta(E)$ includes a δ-ball at each of these points, so that $|\mathrm{nbd}_\delta(E)| \gtrsim \delta^n \delta^{-\frac{n-1}{2-\varepsilon}}$, which says that E has Minkowski dimension at least $\frac{n-1}{2-\varepsilon}$.

(ii) If E is a curved Nikodym set, then we may assume that $\mathrm{nbd}_\delta(E)$ consists of $\delta^{-(n-1)}$ tubes whose centres ω are distinct and δ-separated. So by Lemma 15, we have $\delta^{-(n-1)} \sim \#(\mathcal{A} - T\mathcal{B}) := \#(\mathcal{A} - \mathcal{B}')$ where $\mathcal{B}' := \{Tb : b \in \mathcal{B}\}$. By the assumption, this means that
$$\delta^{-(n-1)} \lesssim \max\left\{\#\mathcal{A}, \#\mathcal{B}', \max_j \#(\mathcal{A} + X(\lambda_j)T^{-1}\mathcal{B}')\right\}^{2-\varepsilon}$$
and so one of the sets on the right-hand side has cardinality at least $\delta^{-\frac{n-1}{2-\varepsilon}}$, and since $T^{-1}\mathcal{B}' = \mathcal{B}$, this implies that $\mathrm{nbd}_\delta(E)$ includes δ-balls as before, giving the same bound for the dimension.

If we have a range of solutions, as we vary the heights t_0, t_1 of $\mathcal{A}$, $\mathcal{B}$ and hence those of the other N slices, we always have at least one of the intersections having large cardinality. So one of them has this property for a range of heights of positive measure. This means that instead of δ-balls, $\mathrm{nbd}_\delta(E)$ actually includes small cylinders of width δ and height some constant c. Hence
$$\left|\mathrm{nbd}_\delta(E)\right| \gtrsim c\delta^{n-1-\frac{n-1}{2-\varepsilon}}$$
which shows that $\dim(E) \geq \frac{n-1}{2-\varepsilon} + 1$ as required. $\square$

So we must try to answer Question 4. Clearly it always holds with $\varepsilon = 0$; this recovers the "trivial bound" for which we proved the stronger maximal function version in section 2. If we could obtain $\varepsilon = 1$ then the sets would have dimension n; unfortunately this is not known for any matrices, and in the scalar case with $N = 1$ has been shown to be false by Ruzsa [R].

5.2   The scalar case.   We will begin by reviewing the known results in the case where all of the X_j are multiples of the identity, and seeing what results for curves can be deduced from them.

With three slices ($N = 1$) and $X_1 = I$ we have the problem Bourgain originally used in [B3]. He proved the estimate with $\varepsilon = 1/13$, which was quickly improved to $1/6$ by Katz and Tao [KT1], i.e.
$$\#(\mathcal{A} - \mathcal{B}) \leq \max\left\{\#\mathcal{A}, \#\mathcal{B}, \#(\mathcal{A} + \mathcal{B})\right\}^{2-1/6}. \tag{24}$$

For all other rational multiples of I the existence of positive improvements ε has been proved by Christ [Ch], although it is tedious to compute their values.

The first four-slice estimate was again due to Katz and Tao, namely

$$\#(\mathcal{A} - \mathcal{B}) \leq \max \left\{ \#\mathcal{A}, \#\mathcal{B}, \#(\mathcal{A} + \mathcal{B}), \#(\mathcal{A} + 2\mathcal{B}) \right\}^{2 - 1/4} \qquad (25)$$

as shown in [KT1]. In [KT2, Theorem 3.3] they showed that $\varepsilon = 1/4$ still holds if, instead of using 1 and 2 as here, the two non-zero scalars simply differ by 1. We shall generalise this for matrices shortly.

In the same theorem, they showed that six slices with scalars $x, y, \bar{x}, \bar{y}$ satisfying

$$\left(1 + \tfrac{1}{\bar{x}} \right) x = \left(1 + \tfrac{1}{\bar{y}} \right) y \,, \qquad (26)$$

also gave $\varepsilon = 1/4$. This relation allows us to obtain results for five slices also, by taking two scalars to be equal (or by taking one to be ∞, in which case we interpret $\mathcal{A} + \infty\mathcal{B} := \mathcal{B}$).

They also proved an iteration result:

Theorem 17 (Katz & Tao [KT2]). *If we can obtain $\varepsilon = \varepsilon_0$ in Question 4 for some finite set of scalars, then for some larger set of scalars we can obtain $\varepsilon = \frac{2 - \varepsilon_0^2}{8 - 7\varepsilon_0 + \varepsilon_0^2}$. Hence by choosing larger and larger sets, the improvement ε may be made as close to the fixed point $0.32486\ldots$ as we wish.*

This result gives the lower bound of approximately $0.5969n + 0.403$ for the Minkowski dimension of straight-line Kakeya sets, which is currently the best known for large n.

In order to apply these results in the curved case we must discover whether $X(\lambda)$, or $X(\lambda)T^{-1}$ can be multiples of the identity.

LEMMA 18. (i) $X(\lambda)$ *cannot be a multiple of the identity except in the straight-line case.*

(ii) $X(\lambda)T^{-1}$ *is a multiple of the identity for all λ if $C^2 = 0$, but this condition is not necessary to obtain the equality for some λ.*

Proof. It will be helpful to write $M = M_{t_0, t_1} := (t_1 - t_0)C\big(I + (t_0 + t_1)C\big)^{-1}$ so that $X(\lambda) := \frac{\lambda}{1 - \lambda}[I + \lambda M]^{-1}[I - (1 - \lambda)M]$. Suppose that $X(\lambda) = \frac{\lambda}{1 - \lambda}kI$ where $k = k(t_0, t_1, \lambda)$ is some scalar function. Then

$$I - (1 - \lambda)M = k(I + \lambda M)$$
$$(1 - k)I = (k\lambda + 1 - \lambda)M$$

which implies that M is some (possibly zero) multiple of I. By the definition of M this implies that C is a multiple of I. As observed before (page 1330) this reduces to the straight-line case.

On the other hand, if $C^2 = 0$, then $M = (t_1 - t_0)C$ and hence $X(\lambda) = \frac{\lambda}{1-\lambda}(I - (t_1 - t_0)C)$, while $T := \frac{t_0}{t_1}(I + t_0 C)(I + t_1 C)^{-1} = \frac{t_0}{t_1}(I - (t_1 - t_0)C)$. So $X(\lambda)$ and T are parallel. Theorem 20 gives examples where $X(\lambda) = T$ but $C^2 \neq 0$.   □

Theorem 19 (Nikodym result for $C^2 = 0$).   *Nikodym sets of curves of the form* (15) *with $C^2 = 0$ have Minkowski dimension at least $\frac{n-1}{2-\varepsilon} + 1 \approx 0.5969n + 0.403$, where ε is the smallest root of $\varepsilon^3 - 6\varepsilon^2 + 8\varepsilon - 2$.*

Proof. In this case, for all t_0, t_1 the sumsets are just the scalar ones $\mathcal{A} + \frac{\lambda}{1-\lambda}\mathcal{B}$. Clearly by choosing suitable heights these can be any scalars we like, so this follows immediately from Katz and Tao's sumset result (Theorem 17).   □

Many other families of curves admit some good bound for the Nikodym sets, however. To use Katz and Tao's simple three-slice estimate (24), all we require is that there exist $t_0, t_1 \in [-1, 1] \setminus \{0\}$ and $\lambda \in (0, 1) \setminus \{\frac{t_0}{t_0 - t_1}\}$ such that $X(\lambda) = T$. These equations are difficult to solve, but where the matrix C is invertible, or where one of A, B is positive definite so that we may assume that C is diagonal and hence commuting, we can simplify the problem.

Theorem 20.   *Suppose that C is either diagonal or invertible. Then $X(\lambda) = T$ if and only if C satisfies the quadratic equation*

$$(t_0^2 t_2^2 + t_1^2 t_2^2 - 2t_0^2 t_1^2)C^2 + (t_0 + t_1 + t_2)(t_0 t_2 + t_1 t_2 - 2t_0 t_1)C$$
$$+ (t_0 t_2 + t_1 t_2 - 2t_0 t_1)I = 0 \quad (27)$$

where t_0, t_1 and $t_2 := (1 - \lambda)t_0 + \lambda t_1$ are the heights of the planes. If this is so for some choice of $t_0, t_1, t_2 \in (-1, 1)$, then the corresponding curved Nikodym sets have Minkowski dimension at least $\frac{6n-6}{11}$, while if there is a whole range of such heights then the dimension is at least $\frac{6n+5}{11}$.

REMARK.   For any such heights to exist, C must have at most two eigenvalues h and k with $\left|\frac{1}{h} + \frac{1}{k}\right| < 3$. Some cases in which h and k satisfy this quadratic for a range of heights are as follows:

  (i) If the eigenvalues h and k are real, then the necessary condition $\left|\frac{1}{h} + \frac{1}{k}\right| < 3$ is also sufficient.
  (ii) If they are complex, so $h = \bar{k} = \alpha + i\beta$, $\alpha, \beta \in \mathbb{R}$, then it is enough for either of the following to hold:
      (a) $\max\left\{2|\alpha| - \alpha^2, \frac{1}{2}\left(-1 - 2\alpha^2 + \sqrt{1 + 16\alpha^2}\right)\right\} \leq \beta^2 \leq 3\alpha^2$.
      (b) $\beta^2 \geq \max\left\{2|\alpha| - \alpha^2, \frac{1}{2}\left(-1 - 2\alpha^2 + \frac{1}{2}\sqrt{16\alpha^2 + 24|\alpha| + 1}\right)\right\}$.
      In particular, $|\alpha| \geq 2.36\ldots$ suffices.

Proof of the theorem and remark.   When C is diagonal or invertible, the equation $X(\lambda) = T$ can be easily rearranged to give (27). Thus for these cases, we require that C should have at most two eigenvalues. Moreover, if the eigenvalues are distinct, then we require C to be diagonalisable (over $\mathbb{C}$), while in the case of one repeated eigenvalue we need the Jordan normal form of C to contain only 1×1 and 2×2 blocks. By considering the ratio of the last two coefficients we find that the sum of the reciprocals of the eigenvalues must be equal to $-(t_0+t_1+t_2) \in (-3, 3)$, which imposes further restriction on C.

(i) If C has real eigenvalues, then they must lie in $(-1/2, 1/2)$ by (14), whence the necessary condition $\left|\frac{1}{h}+\frac{1}{k}\right| < 3$ implies that they are of opposite sign. Without loss of generality $|h| \leq |k|$. We exploit homogeneity by setting $t_0 = t$, $t_1 = bt$, $t_2 = ct$. Then (27) becomes

$$Q(X) := (2b^2 - b^2c^2 - c^2)X^2 + (b+c+1)(2b-bc-c)X + (2b-bc-c) = 0 \quad (28)$$

which should have roots ht and kt. These lie in $(-1/2, 1/2)$ and have opposite sign so that the sum of their reciprocals is in $(-3, 3)$. So we choose b, c so that $Q(0) > 0$, $Q(1/2) < 0$ and $Q(-1/2) < 0$. It is easy to check that this is so if we choose $b \in (0, 1)$ and

$$\frac{-6 - 4b - 2b^2 + 2\sqrt{7b^4 + 28b^3 + 52b^2 + 48b + 9}}{2(b^2 + 2b + 3)} < c < \frac{2b}{1+b} \ .$$

Consider the region of those b, c satisfying this for which $b+c+1 > \left|\frac{1}{h}+\frac{1}{k}\right|$. This is shown in Figure 7, and is not empty provided that $\left|\frac{1}{h} + \frac{1}{k}\right| < 3$.

Let $X^+(b,c), X^-(b,c)$ denote the roots of Q obtained by the quadratic formula, taking the positive or negative square root respectively, and define the function $f(b,c) := -(b + c + 1)X^+(b,c)$, which is continuous on the interior of the region.

On the upper curve we have $X^+ = X^- = 0$ and so $f = 0$, while on the lower curve $X^- = 1/2$. The ratio of the last two coefficients of the quadratic function Q tells us that $\frac{1}{X^+} + \frac{1}{X^-} = -(b+c+1)$, so that on the lower curve we have $f(b,c) = \frac{b+c+1}{b+c+3}$ which tends to $3/5$ as $(b,c) \to (1,1)$. So provided that $1 + h/k \in (0, 3/5)$ we can find a curve of points (b_0, c_0) on which $f(b_0, c_0) = 1 + h/k$. We have $|X^+(b_0, c_0)| = \left|\frac{1+h/k}{b_0+c_0+1}\right| < \frac{|1+h/k|}{\left|\frac{1}{h}+\frac{1}{k}\right|} = |h|$, so that the heights $t := X^+(b_0, c_0)/h$, $t_1 := b_0 t$ and $t_2 := c_0 t$ are all in $(-1, 1)$. Finally observe that

$$1 + h/k = -X^+(b_0, c_0)(b_0 + c_0 + 1) = X^+ \left(\frac{1}{X^+} + \frac{1}{X^-}\right) = 1 + th/X^-$$

so that $X^-(b_0, c_0) = tk$ as required. Observing that the inequalities $\left|\frac{1}{h}+\frac{1}{k}\right| < 3$ and $|h| \leq |k|$ imply $0 < 1 + h/k < 3/5$ for $h, k \in (-1/2, 1/2)$ gives the result stated.
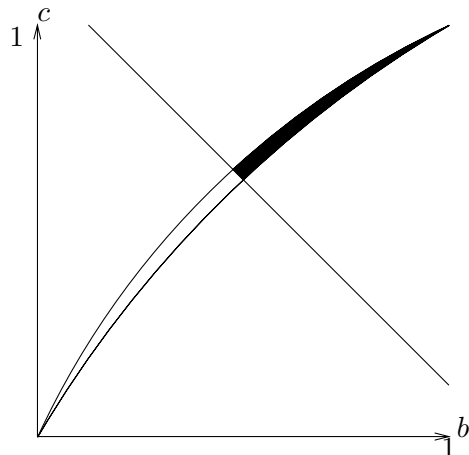
Figure 7: The region $\frac{-6-4b-2b^2+2\sqrt{7b^4+28b^3+52b^2+48b+9}}{2(b^2+2b+3)} < c < \frac{2b}{1+b}$ and $b + c + 0 > \left|\frac{1}{h} + \frac{1}{k}\right|$.

(ii) If the eigenvalues of C are complex conjugates $\alpha \pm i\beta$, consider the map $g : \mathbb{R}^3 \to \mathbb{R}^2$ given by

$$g(t_0, t_1, t_2) = \left(-(t_0 + t_1 + t_2), \frac{t_0 t_2 + t_1 t_2 - 2t_0 t_1}{t_0^2 t_2^2 + t_1^2 t_2^2 - 2t_0^2 t_1^2}\right).$$

The two expressions on the right are the sum of the reciprocals of the roots and the product of the roots respectively. So we have to solve $g(t_0, t_1, t_2) = \left(\frac{2\alpha}{\alpha^2+\beta^2}, \alpha^2 + \beta^2\right)$.

(a) Take $t_1 = -t_0$. Then solving gives

$$t_0 = -t_1 = \frac{\sqrt{3\alpha^2 - \beta^2}}{\alpha^2 + \beta^2} \qquad\qquad t_2 = \frac{-2\alpha}{\alpha^2 + \beta^2}$$

which are in $(-1, 1)$ if and only if α and β are as claimed.

(b) Take $t_2 = -t_0$. This is not so easily solved, but $t_1 = \frac{-2\alpha}{\alpha^2+\beta^2}$ while t_0 satisfies the following cubic:

$$(\alpha^4 + 2\alpha^2\beta^2 + \beta^4)t_1^3 + (\beta^2 - 3\alpha^2)t_1 - 6\alpha = 0.$$

By considering the sign of this cubic at $t_1 = -1, 0, 1$, we can force a sign change in the interval $(-1, 1)$ by letting α and β satisfy the inequality stated.

In both cases it is easy to check that the Jacobian of g has full rank, so that we can use the implicit function theorem to find not just one solution for (t_0, t_1, t_2) but a whole range.

Once we have found ranges of heights so that the quadratic is satisfied, the result follows from (24) and Lemma 16. □

It is likely that there are many other complex pairs h, k which work, however it seems difficult to describe the set of all such pairs concisely.

5.3 Non-scalar matrices.

In the case of curved Kakeya rather than Nikodym sets, the matrix $X(\lambda)$ is never a multiple of I, nor is there a matrix T which it might cancel with. So we have no option but to try to answer Question 4 with the X_j not multiples of I. This seems hard. However, we have some negative results, and have been able to generalise one of the positive results from the scalar case. We begin with a rather trivial observation.

LEMMA 21. *If all of the X_j are block diagonal with blocks of the same size, then a sumset inequality for these X_j implies one for each of the sets of blocks, with the same ε.*

Proof. Obvious by letting $\mathcal{A}, \mathcal{B}$ consist of vectors with zeros everywhere except in the block of interest. □

The converse seems likely to be false – we would need not only that "collisions" often occur in each block of coordinates, but that they often occur in all coordinates at the same time.

We now reveal the easy but disappointing fact that three slices are simply not enough in the matrix case.

PROPOSITION 22. *If X is not a multiple of the identity, then the power of 2 in*

$$\#(\mathcal{A} - \mathcal{B}) \leq \max\left\{\#\mathcal{A}, \#\mathcal{B}, \#(\mathcal{A} + X\mathcal{B})\right\}^2$$

is best possible.

Proof. Choose a vector v that is not an eigenvector of X, and let $\mathcal{B}$ consist of M equally spaced points along this direction. Set $\mathcal{A} = \{Xb : b \in \mathcal{B}\}$, that is, M equally spaced points along the direction Xv. Then with $\mathcal{G} = \mathcal{A} \times \mathcal{B}$, clearly $\#(\mathcal{A} + X\mathcal{B})$ is about $2M$, while since v and Xv are linearly independent, $\#(\mathcal{A} - \mathcal{B})$ is about M^2. □

Similar observations with more slices give another negative result.

PROPOSITION 23. *Suppose that there exists $v \in \mathbb{R}^n$ such that all the vectors $X_j v$ are rational multiples of some fixed vector w which is not parallel to v itself. (That is, v is a secular vector of each pair of matrices, but is not an eigenvector.) Then there can be no positive answer to Question 4.*

This theorem is rather weak, but it does at least rule out the case where the matrices X_j are all multiples of each other but not of the identity, and combining this with Lemma 21 gives further examples. This makes sense because taking more than three slices is not really giving much more information.

Proof. We have $X_j v = \frac{p_j}{q_j} w$ where p_j, q_j are non-zero coprime integers. Let $M > \prod_{i=1}^{N} p_i q_i$ be a large integer, and set

$$\mathcal{A} = \left\{ n \Big(\prod_{i=1}^{N} p_i q_i \Big) w : n = 1, \dots, M \right\}$$

$$\mathcal{B} = \left\{ n \Big(\prod_{i=1}^{N} q_i \Big) v : n = 1, \dots, M \right\}.$$

Then if $\mathcal{G} = \mathcal{A} \times \mathcal{B}$, we find that

$$\mathcal{A} + X_j \mathcal{B} = \left\{ p_j \Big(\prod_{k \neq j} q_k \Big) \Big[q_j \Big(\prod_{i \neq j} p_i \Big) m + n \Big] w : m, n = 1, \dots, M \right\}$$

$$\mathcal{A} - \mathcal{B} = \left\{ m \Big(\prod_{j=1}^{N} p_i q_i \Big) w - n \Big(\prod_{i=1}^{N} q_i \Big) v, n = 1, \dots, M \right\}$$

and we have the combinatorial task of finding their cardinality. Since w and v are linearly independent it is obvious that $\#(\mathcal{A} - \mathcal{B}) = M^2$. To find the cardinality of $\mathcal{A} + X_j \mathcal{B}$, first write $Q_j := q_j \prod_{i \neq j} p_i$. We need to know how many distinct numbers $(mQ_j + n)$ there are, so first let $m < M$ be fixed. Since $M > Q_j$, there are Q_j such numbers between $mQ_j + 1$ and $(m + 1)Q_j$ inclusive, after which the remaining values overlap with those for the next m. When $m = M$ we just get all the numbers $MQ_j + 1$ up to $MQ_j + M$. Hence we find that $\#(\mathcal{A} + X_j \mathcal{B}) = Q_j(M - 1) + M$. So $\mathcal{A}$ and $\mathcal{B}$ and all the sumsets have cardinality about M while the difference set is about M^2. So there can be no positive answer to Question 4. $\square$

This result shows us exactly what goes wrong in Bourgain's "worst case" example, and more generally for the parabolas for which we have proved good results in the Nikodym case.

COROLLARY 24. *For a Kakeya set of curves of the form* (15) *with* $C^2 = 0$ *we cannot prove any non-trivial bound by sumset methods.*

Proof. If $C^2 = 0$, then we can calculate

$$X(\lambda) = \frac{\lambda}{1 - \lambda} \big(I + (t_0 - t_1)C \big).$$

So we would need a sumset result where the X_j were all multiples of each other but not of I. But we have already seen in Proposition 23 that in such a case no non-trivial estimate can hold.   □

It is interesting that curves that behave well for Nikodym should not do so for Kakeya. We shall discuss this in the final section.

So far this picture looks bleak. However, we can prove one or two results in the positive direction, analogous to those in the scalar case. Here we generalise the four-slice result (25) to the matrix setting. For legibility, write $X_1 = X$, $X_2 = Y$.

PROPOSITION 25. *If $Y - X = I$ and X is invertible, then*

$$\#(\mathcal{A} - \mathcal{B}) \leq \max\left\{\#\mathcal{A}, \#\mathcal{B}, \#(\mathcal{A} + X\mathcal{B}), \#(\mathcal{A} + Y\mathcal{B})\right\}^{7/4}.$$

Proof. This is just as in [KT1], so we only give an outline. Start by discarding elements of $\mathcal{G}$ until $\#(\mathcal{A} - \mathcal{B}) = \#\mathcal{G}$, and denote the maximum on the right-hand side by M. We need to show that $\#\mathcal{G} \leq M^{7/4}$.

The idea is to count *trapezia*: sets of four elements of $\mathcal{G}$ consisting of two "sides" whose endpoints have the same value of a while the endpoints of the remaining two sides share values of $a + Yb$ and b respectively. More precisely, a trapezium is a set

$$\left\{(a_0, b_0), (a_0, b_0'), (a_1, b_1), (a_1, b_1')\right\} \subseteq \mathcal{G}$$

such that $a_0 + Yb_0 = a_1 + Yb_1$ and $b_0' = b_1'$.

First count the number of pairs in $\mathcal{G}$ that share their value of a. This is

$$\#\left\{(a, b), (a, b') \in \mathcal{G}\right\} = \sum_{a \in \mathcal{A}} \#\left\{b : (a, b) \in \mathcal{G}\right\}^2$$
$$\geq \frac{\#\mathcal{G}^2}{\#\mathcal{A}}$$

by Cauchy–Schwarz. A trapezium consists of two such pairs that share their value of $(a + Yb, b')$, so by Cauchy–Schwarz again we find that the number of trapezia is at least

$$\frac{(\#\mathcal{G}^2/\#\mathcal{A})^2}{\#(\mathcal{A} + Y\mathcal{B})\#\mathcal{B}} \geq \frac{\#\mathcal{G}^4}{M^4}\,.$$

But we also have the following algebraic fact:

$$a_1 - b_1' = (I + X^{-1})(a_0 + Xb_0) - X^{-1}(a_0 + Xb_0') - Yb_1\,.$$

So, since $\#(\mathcal{A} - \mathcal{B}) = \#\mathcal{G}$, knowing $(a_0 + Xb_0), (a_0 + Xb_0')$, and b_1 is enough to determine (a_1, b_1'), and hence the whole trapezium by substituting back. So the number of trapezia is at most M^3, which together with the lower bound of $\#\mathcal{G}^4/M^4$ gives the result.   □

To apply this to the Kakeya problem with curves, we need conditions
on the curves (in terms of C) that guarantee the existence of $t_0, t_1 \in [-1, 1]$
and $\lambda, \mu \in (0, 1)$ such that $X(\lambda) - X(\mu) = I$. Unfortunately, in many cases
this cannot be done. This is hardly surprising, since for fixed C, we are
trying to satisfy $(n - 1)^2$ equations with only four unknowns. As with the
Nikodym sets, we end up trying to make C satisfy a polynomial, whose
coefficients now depend on the four heights. However, in the Kakeya case
we are able to deal with the λ and μ first, independently of t_0 and t_1, which
makes the problem easier.

PROPOSITION 26. *If the matrix $M := (t_1 - t_0)C(I + (t_0 + t_1)C)^{-1}$ is
nilpotent, then $X(\lambda) - X(\mu)$ is never equal to the identity.*

Proof. Let k be the highest power of M that is non-zero. Then $[I + \lambda M]^{-1} = \sum_0^k (-1)^n \lambda^n M^n$, and hence

$$X(\lambda) - X(\mu) = \left(\frac{\lambda}{1 - \lambda} - \frac{\mu}{1 - \mu} \right) I + \sum_1^k (-1)^n M^n \left(\frac{\lambda^n}{1 - \lambda} - \frac{\mu^n}{1 - \mu} \right).$$

If this equals the identity, then some linear combination of $I, M, M^2, \ldots, M^k$
is zero. But this cannot happen because the minimum polynomial of a
nilpotent matrix is x^{k+1}. $\square$

This cuts down the list of possible candidates for C. Note in particular
that all matrices with $C^2 = 0$ make M nilpotent for every choice of t_0, t_1.
Obviously invertible matrices C are not ruled out, and nor are diagonal
matrices, and whenever either A or B is positive definite we can fix C to
be diagonal by a change of coordinates.

PROPOSITION 27. *Suppose that C is invertible or diagonal.   Then
$X(\lambda) - X(\mu) = I$ if and only if $M := (t_1 - t_0)C(I + (t_0 + t_1)C)^{-1}$ sat-
isfies the following quadratic:*

$$M^2 + \left(\frac{2}{\mu} + \frac{1}{1 - \mu} - \frac{1}{1 - \lambda} \right) M + \left(\frac{1}{\lambda\mu} - \frac{1}{\mu(1 - \lambda)} + \frac{1}{\lambda(1 - \mu)} \right) I = 0.$$
(29)

*This quadratic cannot have both its roots real and in $(-1, 1)$, but suitable
$\lambda, \mu \in (0, 1)$ can be chosen to give any desired roots l and m such that
$l + m < -2(1 + \sqrt{2})$.*

Proof. By Hörmander's criterion (14), C is invertible or diagonal if and
only if M is. It is then easy to rearrange the equation $X(\lambda) - X(\mu) = I$
to give (29). So M must either be diagonalisable with at most two distinct
eigenvalues, or have one repeated eigenvalue and have Jordan normal form

consisting only of 1×1 and 2×2 blocks. Suppose that M has two eigenvalues l and m. These are either real or form a complex conjugate pair, and so both their sum and their product are real. We obtain two simultaneous equations by considering the sum and product of the roots of (29):

$$\frac{1}{\lambda\mu} - \frac{1}{\mu(1-\lambda)} + \frac{1}{\lambda(1-\mu)} = lm \,, \tag{30}$$

$$\frac{2}{\mu} + \frac{1}{1-\mu} - \frac{1}{1-\lambda} = -(l+m) \,. \tag{31}$$

Now (31) is linear in λ so we solve it to obtain

$$\lambda = 1 - \frac{\mu(1-\mu)}{2 - \mu + \mu(1-\mu)(l+m)} \,.$$

Of course this needs to lie in $(0,1)$. Tedious calculation shows that in the case $l+m > -2(1+\sqrt{2})$ it does so for all $\mu \in (0,1)$. For $l+m < -2(1+\sqrt{2})$ it does so provided we take

$$\mu \in \left(0, \frac{2-(l+m)-\sqrt{(l+m+2)^2-8}}{2(1-l-m)} \right) \text{ or } \mu \in \left(\frac{2-(l+m)+\sqrt{(l+m+2)^2-8}}{2(1-l-m)}, 1 \right). \tag{32}$$

Next we substitute this expression for λ back into (30). After rearranging we obtain an equation which is quartic in μ and quadratic in l and m. With the help of MAPLE we express it as

$$\begin{aligned}
0 &= [lm(l+m-1)]\mu^4 + \left[2lm + (l+m)\left((l+m)^2 - lm - 1 \right) \right]\mu^3 \\
&\quad + \left[(l+m)(4-(l+m)) - 1 - 2lm \right]\mu^2 + 4(1-l-m)\mu - 4
\end{aligned} \tag{33}$$

$$\begin{aligned}
&= -\left[\mu^2(1-\mu)(\mu m+1) \right]l^2 + \left[\mu(\mu^2 m - \mu m + \mu - 2)(\mu m - \mu + 2) \right]l \\
&\quad + (\mu^2 m - \mu m + \mu - 2)(\mu m - \mu + 2)
\end{aligned} \tag{34}$$

$$=: q(\mu, l, m) \,.$$

Note that this is a real-valued function of μ. For the second part we use a naïve approach via the intermediate value theorem. Setting $\mu = 0$ gives -4, while $\mu = 1$ gives $-(l+1)(m+1)$. This is positive only when l and m are real with exactly one being less than -1. Similarly we obtain a positive result by setting $\mu = \frac{-2}{m-1}$ or $\frac{-2}{l-1}$, but these too are only permitted if l, m are real and one is less than -1. However, if we instead substitute in either endpoint from equation (32) (which is allowable if and only if $l+m < -2(1+\sqrt{2})$) we obtain

$$\frac{(l+m)^2 \pm (l+m-2)\sqrt{(l+m+2)^2-8}}{2(l+m-1)^2}$$

which by more tedious rearranging is seen to be positive for all l, m.

For the first assertion we show that the maximum of the function q over the region $(\mu, l, m) \in [0,1] \times [-1,1] \times [-1,1]$ is zero, and moreover that this

is attained only for $\mu = 1$. For interior maxima we use the version (34) of the equation as a quadratic in l. Its stationary point (a maximum) occurs at

$$l = \frac{(\mu^2 m - \mu m + \mu - 2)(\mu m - \mu + 2)}{2\mu(1 - \mu)(\mu m + 1)}\,.$$

Now if $|m| < 1$ then the denominator is positive, so that the whole fraction will be less than -1 if

$$(\mu^2 m - \mu m + \mu - 2)(\mu m - \mu + 2) < -2\mu(1 - \mu)(\mu m + 1)$$

which rearranges to

$$\mu^2(1 - \mu)m^2\big(3(1 - \mu)^2 + 1\big)(\mu m + 1) > 0\,.$$

So there is no zero of $\partial q/\partial l$ in the region, except perhaps when $\mu = 1$ and $m = -1$. We find that $q(1, l, -1) \equiv 0$. Now to check the other boundaries:

$$q(\mu, 1, m) = -m^2\mu^2(1 - \mu^2) - 2(\mu m + 1)(2 - \mu^2)$$

$$q(\mu, -1, m) = -m^2\mu^2(1 - \mu)^2 - 2(\mu m + 1)(1 - \mu)\big((1 - \mu)^2 + 1\big)\,.$$

Both of these are clearly non-positive, and give zero only at $(1, l, -1)$ as we have already seen, and at $(1, 1, m)$. $\qquad\square$

Unfortunately, M cannot have real eigenvalues outwith $(-1, 1)$. This follows from the non-degeneracy criterion (14) and that fact that l is an eigenvalue of M if and only if

$$\det\big[I + \big(\big(1 - \tfrac{1}{l}\big)t_0 + \big(1 + \tfrac{1}{l}\big)t_1\big)C\big] = 0\,.$$

However, if the eigenvalues are complex conjugate we obtain a Kakeya result as follows.

Theorem 28.   *Suppose that C is invertible and diagonalisable over $\mathbb{C}$ and has only two eigenvalues $\alpha \pm \beta i$. Then if either α is sufficiently large ($|\alpha| > \frac{1+\sqrt{2}}{2}$ will do) or β is large compared to α, there is a lower bound of $\frac{4n+3}{7}$ for the curved Kakeya sets associated to C.*

Proof. We have seen that this holds if we can make $l + m < -2(1 + \sqrt{2})$. But $l + m$ is simply twice the real part of the eigenvalues of $M := (t_1 - t_0)C(I + (t_0 + t_1)C)^{-1}$, so if we let the eigenvalues of C be $\alpha \pm i\beta$ as before, we require

$$(t_1 - t_0)\frac{\alpha + (t_0 + t_1)(\alpha^2 + \beta^2)}{1 + 2(t_0 + t_1)\alpha + (t_0 + t_1)^2(\alpha^2 + \beta^2)} < -\big(1 + \sqrt{2}\big)\,.$$

It helps to write $t_1 = 1 - 2\varepsilon$, $t_0 = -1 + \varepsilon$, where $\varepsilon < 2/3$ may be taken as small as we wish. The inequality becomes

$$\frac{\alpha(1 - \varepsilon\alpha) - \varepsilon\beta^2}{(1 - \varepsilon\alpha)^2 + \varepsilon^2\beta^2} < -\frac{1 + \sqrt{2}}{2 - 3\varepsilon}\,.$$

Clearly this is satisfied for small ε and large β: Choosing $\varepsilon \lesssim 1/|\alpha|$ shows that $\beta \gtrsim |\alpha| + 1$ will work. Alternatively if $\alpha < -\frac{1+\sqrt{2}}{2}$ then we simply need to take ε very small. If $\alpha > \frac{1+\sqrt{2}}{2}$ we simply swap t_0 and t_1. In all of these cases, we have in fact found a whole family of solutions for varying ε so there is no problem with using the argument about varying the heights of the planes which gave the extra $+1$ for the dimension bound in Lemma 16. $\qquad\square$

So we get a non-trivial result in some cases, although it is not easy to give the criteria any geometric interpretation.

However, about the case where C is not invertible, or where C has more than two eigenvalues or two real ones, we cannot say anything other than that the above proof will not work.

We have not yet considered using four slices in the Nikodym case. This is more complicated, because we require $X(\lambda) - X(\mu) = T$ instead of I, which means that we cannot write this in terms of M and so we must look at all four variables t_0, t_1, λ, μ together, rather than in two stages as we did above. By the methods already used, we can show that if C is diagonal or invertible then it must satisfy a cubic equation, and that the reciprocals of the roots (the eigenvalues of C) must have the same sum as minus the heights of the slices, as we found in section 5.2. As one would expect, it is difficult to say anything more than that explicitly.

So what hope is there for the use of arithmetic methods? If we still want to use only four slices for cases not covered above, then we shall have to prove a new sumset result, that is, find a more flexible condition than $X(\lambda) - X(\mu) = I$ which guarantees that the difference set is not too much larger than the two original sets and their $X(\lambda)$ and $X(\mu)$ matrix sumsets. Or we could instead look at using more slices – the techniques in [KT2] suggest that relations like

$$0 = X(\lambda) - X(\mu) + X(\nu)^{-1}X(\lambda)$$
$$\text{or } 0 = X(\lambda) - X(\mu) + X(\nu)^{-1}X(\lambda) - X(\kappa)^{-1}X(\mu)\,,$$

which are analogues of (26), would suffice. But of course these lead to higher degree polynomials in more variables which make it harder to compute sufficient conditions for suitable solutions to exist.

6   Final Remarks

We have now seen two non-trivial positive results for curves of the form (15) where $C^2 = 0$, namely the $\frac{n+2}{2}$ bound for the Nikodym maximal function (Theorem 12) and the bound of approximately $0.5969n + 0.403$ for the Minkowski dimension of the Nikodym sets (Theorem 19). This condition on the matrix C has a surprising link with Bourgain's "worst case" example for the curved Kakeya problem. That example had $C = \left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)$ which clearly satisfies the condition. This is no accident; the following converse is also true:

PROPOSITION 29. *If $C^2 = 0$, then the corresponding Kakeya sets can have dimension as low as $n - \mathrm{rank}(C)$. In particular, in odd dimensions, the "trivial" lower bound of $\frac{n+1}{2}$ can be attained, while in even dimensions there can be sets with dimension at least as low as $\frac{n+2}{2}$.*

Proof. The curves are as in (15) where we may assume that C is in rational canonical form. Then $C^2 = 0$ if and only if C consists only of 1×1 blocks $(\,0\,)$ and 2×2 blocks $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$. The rank of B is the number of 2×2 blocks, which can be at most $\frac{n-1}{2}$ if n is odd, and $\frac{n-2}{2}$ if n is even. We can now imitate the proof on page 1327 in each block, and the result follows.   □

So it seems that the same curves that allow no good bound in the Kakeya case are particularly amenable to proving good bounds in the Nikodym case. This is rather curious, and may reveal a kind of duality between the Kakeya and Nikodym problems. Up until now Restriction/Kakeya and Bochner–Riesz/Nikodym have been thought of as essentially the same [C2], [T], but the above suggests they might be better described as dual in some way, or even opposite. This idea is not so strange when one remembers that curvature of the surface in question is *good* when considering Restriction (since it causes decay of the Fourier transform) but *bad* for Bochner–Riesz (Bochner–Riesz for squares is trivial).

This also shows the importance of Carbery's transformation

$$(x', x_n) \mapsto (x'/x_n, 1/x_n)$$

which relates the two classes of problems. This does not preserve parabolas; rather it maps them to hyperbolas. Another way of phrasing the above is that if for a given matrix, the parabolas can be tightly packed, then the hyperbolas cannot. This would leave straight lines as an overlapping middle case, the only family that this transformation leaves unchanged. These ideas will be explored further in [CW].

References

[B1]   J. BOURGAIN, Besicovitch type maximal operators and applications to Fourier analysis, GAFA, Geom. funct. anal. 1 (1991), 147–187.

[B2]   J. BOURGAIN, L^p estimates for oscillatory integrals in several variables, GAFA, Geom. funct. anal. 1 (1991), 321–374.

[B3]   J. BOURGAIN, On the dimension of Kakeya sets and related maximal inequalities, GAFA, Geom. funct. anal. 9(1999), 256–282.

[C1]   A. CARBERY, Covering lemmas revisited, Proc. Edinburgh Math. Soc. (2) 31 (1988), 145–150.

[C2]   A. CARBERY, Restriction implies Bochner–Riesz for paraboloids, Math. Proc. Cambridge Philos. Soc. 111 (1992), 525–529.

[CHS]   A. CARBERY, E. HERNÁNDEZ, F. SORIA, Estimates for the Kakeya maximal operator on radial functions in $\mathbb{R}^n$, in "Harmonic Analysis (Sendai, 1990)" (S. Igari, ed.), Springer, Tokyo (1991), 41–50.

[CW]   A. CARBERY, L. WISEWELL, Three notes on Restriction and Bochner–Riesz, in preparation.

[Ch]   M. CHRIST, On certain elementary trilinear operators, Math. Res. Lett. 8 (2001), 43–56.

[ChDR] M. CHRIST, J. DUOANDIKOETXEA, J.L. RUBIO DE FRANCIA, Maximal operators related to the Radon transform and the Calderón–Zygmund method of rotations, Duke Math. J. 53:1 (1986), 189–209.

[F]   K.J. FALCONER, The Geometry of Fractal Sets, Cambridge Tracts in Mathematics 85, Cambridge University Press, 1985.

[Fe]   C. FEFFERMAN, The multiplier problem for the ball, Ann. of Math. 94 (1971), 330–336.

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

[K]   N.H. KATZ, Social density and the Kakeya maximal operator, to appear.

[KLT]   N.H. KATZ, I. ŁABA, T. TAO, An improved bound on the Minkowski dimension of Besicovitch sets in $\mathbb{R}^3$, Ann. of Math. 152 (2000), 383–446.

[KT1]   N.H. KATZ, T. TAO, Bounds on arithmetic projections, and applications to the Kakeya conjecture, Math. Res. Lett. 6 (1999), 625–630.

[KT2]   N.H. KATZ, T. TAO, New bounds for Kakeya problems, J. Anal. Math. 87 (2002), 231–263.

[LT]   I. ŁABA, T. TAO, An improved bound for the Minkowski dimension of Besicovitch sets in medium dimension, GAFA, Geom. funct. anal. 11 (2001), 773–806.

[MS]   W.P. MINICOZZI I, C.D. SOGGE, Negative results for Nikodym maximal functions and related oscillatory integrals in curved space, Math. Res. Lett. 4 (1997), 221–237.

[N]   M.B. Nathanson, Additive Number Theory: Inverse Problems and the Geometry of Sumsets, Graduate Texts in Mathematics 165, Springer Verlag, New York, 1996.

[R]   I.Z. Ruzsa, Sums of finite sets, in "Number theory New York Seminar 1991–1995" (D.V. Chudnovsky, G.V. Chudnovsky, M.B. Nathanson, eds.), Springer, New York (1996), 281–293.

[S]   W. Schlag, A geometric inequality with applications to the Kakeya problem in three dimensions, GAFA, Geom. funct. anal. 8:3 (1998), 606–625.

[St]   E.M. Stein, Oscillatory integrals in Fourier analysis, in "Beijing Lectures in Harmonic Analysis (Beijing, 1984)" (E.M. Stein, ed.), Annals of Mathematics Studies 112, Princeton University Press, New Jersey (1986). 307–355.

[T]   T. Tao, The Bochner–Riesz conjecture implies the restriction conjecture, Duke Math. J. 96 (1999), 363–376.

[W1]   L. Wisewell, Families of surfaces lying in a null set, to appear in Mathematika.

[W2]   L. Wisewell, Oscillatory Integrals and Curved Kakeya Sets. PhD thesis, University of Edinburgh, 2003.

[Wo1]   T. Wolff, An improved bound for Kakeya type maximal functions, Rev. Mat. Iberoamericana 11:3 (1995), 651–674.

[Wo2]   T. Wolff, Recent work connected with the Kakeya problem, in "Prospects in Mathematics (Princeton, New Jersey, 1996)" (H. Rossi, ed.), American Mathematical Society, Providence, RI (1999), 129–162.

Laura Wisewell, Department of Mathematics, University College London, Gower Street, London WC1E 6BT United Kingdom   lwiswll@math.ucl.ac.uk