

*Pacific
Journal of
Mathematics*

ON SOME VARIANTS OF THE KAKEYA PROBLEM

LAWRENCE KOLASA AND THOMAS WOLFF

ON SOME VARIANTS OF THE KAKEYA PROBLEM

LAWRENCE KOLASA AND THOMAS WOLFF

We study the question of lower bounds for the Hausdorff dimension of a set in \mathbb{R}^n containing spheres of every radius. If $n \geq 3$ then such a set must have dimension n . If $n = 2$ then it must have dimension at least $11/6$. We also study the analogous maximal function problem and related problem of Besicovitch sets with an axis of symmetry.

Besicovitch and Rado [1] and Kinney [5] proved the following result: There is a closed set $E \subset \mathbb{R}^2$ with measure zero which contains a circle of every radius.

The construction of Besicovitch and Rado works in \mathbb{R}^d : If $d \geq 3$ there is a closed set $E \subset \mathbb{R}^d$ with measure zero which contains a sphere of every radius. We will give an exposition of this construction in Section 1 below.

One can ask whether a set containing a sphere of every radius must have Hausdorff dimension d . As it turns out, this question is easily answered in higher dimensions.

Theorem 1. *A Borel set in \mathbb{R}^d , $d \geq 3$ which contains a sphere of every radius has Hausdorff dimension d .*

In \mathbb{R}^2 this may still be true but appears harder. One purpose of this paper is to prove the following partial result.

Theorem 2. *A Borel set in \mathbb{R}^2 which contains a circle of every radius has Hausdorff dimension at least $\frac{11}{6}$.*

Following a known pattern (see [2] for example) we will derive Theorems 1 and 2 from $L^p \rightarrow L^q$ estimates for a related maximal function. Fix $\delta > 0$. For $x \in \mathbb{R}^d$, $r \in [\frac{1}{2}, 2]$, define

- (1) $C(x, r) = \{z \in \mathbb{R}^d : |z - x| = r\}$
- (2) $C_\delta(x, r) = \{z \in \mathbb{R}^d : r - \delta < |z - x| < r + \delta\}$.

If $f : \mathbb{R}^d \rightarrow \mathbb{R}$ then define $M_\delta f : [1/2, 2] \rightarrow \mathbb{R}$ via

$$M_\delta f(r) = \sup_x \frac{1}{|C_\delta(x, r)|} \int_{C_\delta(x, r)} |f|,$$

where $|E|$ means Lebesgue measure of E . In higher dimensions we have the following estimate:

Theorem 1'. $\|M_\delta f\|_{L^2([1/2,2])} \leq C_d (\log \frac{1}{\delta})^{1/2} \|f\|_2$, $d \geq 3$. In fact, if $d \geq 3$ there are $L^p \rightarrow L^q$ estimates

$$\|M_\delta f\|_{L^q([1/2,2])} \leq C \left(\log \frac{1}{\delta} \right)^{1/p} \|f\|_p$$

if $p \geq 2$ and $q \leq p$ and

$$\|M_\delta f\|_{L^q([1/2,2])} \leq C \delta^{-\alpha} \|f\|_p$$

for other p, q , where $\alpha = \max\left(\frac{1}{p} - \frac{1}{p'}, \frac{1}{p} - \frac{1}{q}\right)$.

Theorem 1' follows easily from (for example) the method of A. Cordoba [3] and we expect that it may have been observed before, although we don't know a reference and will therefore give a proof, in Section 2.

Theorem 1' is easily seen to fail in \mathbb{R}^2 , as we will show in Section 3. The correct question appears to be whether there is an estimate

$$(3) \quad \forall \epsilon > 0 \exists C_\epsilon : \|M_\delta f\|_{L^3([1/2,2])} \leq C_\epsilon \delta^{-\epsilon} \|f\|_3, \quad d = 2.$$

By interpolation with the trivial estimate $\|M_\delta f\|_\infty \leq C \delta^{-1} \|f\|_1$, an equivalent question is whether there is an estimate

$$(4) \quad \forall \epsilon > 0 \exists C_\epsilon : \|M_\delta f\|_{L^q([1/2,2])} \leq C_\epsilon \delta^{-\frac{1}{2}(\frac{3}{p}-1)-\epsilon} \|f\|_p, \quad p \leq 3, q \leq 2p', d = 2.$$

When $p = 2$ (hence $q = 4$) this estimate is due to Pecher [9] (Theorem 1, the case $n = 2$ and $q = \infty$), in an equivalent form referring to solutions of the wave equation, but non- L^2 estimates appear to require different ideas.

We will prove (4) when $p \leq \frac{8}{3}$, using a geometric approach related to that of Marstrand [7]. Actually, we will prove (as does Pecher when $p = 2$) the following slightly stronger "endpoint" result.

Theorem 2'. If $d = 2$, $p < 8/3$, $q \leq 2p'$ then

$$\|M_\delta f\|_q \leq C \delta^{-\frac{1}{2}(\frac{3}{p}-1)} \|f\|_p.$$

Whether (4) holds when $8/3 < p \leq 3$, and whether a set in \mathbb{R}^2 containing a circle of every radius has Hausdorff dimension 2, at present are open problems. These problems can be regarded as special cases of the local smoothing problem for the wave equation [10], [8] - a slightly weaker form of (3) (namely: The same estimate with the L^3 norm replaced by the L^4 norm on both sides of the inequality) would follow readily from the "sharp local smoothing" conjecture made in [10], [8].

Another closely related problem, which was the original motivation for this paper, is the three dimensional Kakeya problem specialized to the cylindrically symmetric case. More precisely define a rotation invariant function

to be a function $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ such that

$$x = (x_1, x_2, x_3) \in \mathbb{R}^3, \theta \in \mathbb{R} \\ \Rightarrow f(x_1 \cos \theta + x_2 \sin \theta, -x_2 \cos \theta + x_1 \sin \theta, x_3) = f(x)$$

and a rotation invariant set to be a set whose characteristic function is rotation invariant. A Besicovitch set is a Borel set in \mathbb{R}^3 containing a line segment in every direction. We will prove:

Theorem 3. *Any rotation invariant Besicovitch set has Hausdorff dimension at least $\frac{17}{6}$.*

For general sets, the best bound presently known is $\frac{5}{2}$, proved in [13] which contains references to relevant previous work.

Bourgain [2] realized that the maximal function most closely related to the Kakeya problem is the following f_δ^* . If $f : \mathbb{R}^3 \rightarrow \mathbb{R}$, $\delta > 0$, \mathbb{P}^2 is two dimensional projective space then one defines $f_\delta^* : \mathbb{P}^2 \rightarrow \mathbb{R}$ by

$$f_\delta^*(e) = \sup_{T_e^\delta} \frac{1}{|T_e^\delta|} \int_{T_e^\delta} |f|$$

where T_e^δ runs over all cylinders with length 1, cross section radius δ and axis in the e direction. We will prove the following.

Theorem 3'. *If $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ is rotation invariant then*

$$\|f_\delta^*\|_{L^q(\mathbb{P}^2)} \leq C \delta^{-(\frac{3}{p}-1)} \|f\|_p, \quad p < \frac{17}{6}, \quad q \leq 2p'$$

To see why Theorems 2' and 3' are related, note first that although Theorem 3' is formulated in \mathbb{R}^3 , it is really a two dimensional result because of the rotation invariance. We will show in Section 5 that it basically reduces to a result like Theorem 2' with hyperbolas instead of circles. Namely, if $\delta > 0$ is fixed and if x, y, z are real numbers with $x > 0$ and $\frac{1}{2} \leq z \leq 2$ then we define

$$(5) \quad H(x, y, z) = \left\{ (s, t) \in \mathbb{R}^2 : s > 0, s^2 - \frac{(t - y)^2}{z} = x \right\}$$

$$(6) \quad H_\delta(x, y, z) = \left\{ (s, t) \in \mathbb{R}^2 : \frac{1}{2} \leq s \leq 2, \text{dist}((s, t), H(x, y, z)) < \delta \right\},$$

and for $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ we define

$$(7) \quad \tilde{M}_\delta f(z) = \sup_{x>0, y \in \mathbb{R}} \frac{1}{\delta} \int_{H_\delta(x, y, z)} |f|.$$

Lemma 5.1 below is the result analogous to Theorem 2' for \tilde{M}_δ . In fact, we will obtain both Theorem 2' and Lemma 5.1 as special cases of a more

general result (Theorem 4.1) for families of curves satisfying the “cinematic curvature” condition from [10].

Theorem 3' bears the same relation to Theorem 3 as Theorem 2' does to Theorem 2. In fact, the implication (Theorem 3' \Rightarrow Theorem 3) follows by specializing the argument in [2], Lemma 2.15 to the cylindrically symmetric case, while the following lemma shows that Theorems 1' and 2' imply Theorems 1 and 2.

Lemma 0.1. *An estimate in \mathbb{R}^d of the form*

$$\|M_\delta f\|_q \leq C\delta^{-\alpha}\|f\|_p$$

implies that a Borel set $E \subset \mathbb{R}^d$ containing a positive ($d - 1$ dimensional) measure subset of a sphere of radius r for a positive measure set of r has Hausdorff dimension at least $d - \rho\alpha$.

The proof of this is identical to the proof of Lemma 2.15 in [2], so we omit it.

We use the notation $x \lesssim y$ to mean that $x \leq Cy$ for a suitable fixed constant C , and similarly with $x \gtrsim y$ and $x \approx y$. The rest of the paper is organized as follows: In Section 1 we carry out the Besicovitch-Rado construction in general dimensions and in Section 2 we prove Theorem 1'. In Section 3 we extend lemmas from [7] to the setting of curves with cinematic curvature, and in Section 4 we prove Theorem 2' and the above mentioned generalization. The argument in Section 4 is the main point of the paper. In Section 5 we prove Theorem 3'.

Acknowledgments. We are indebted to P. Mattila who pointed out the similarity between our arguments and Marstrand's. The fact that the $p = 2$ case of Theorem 2' is in the wave equation literature was pointed out by C. Sogge and by C. Kenig and L. Vega.

Added 5/15/97. This paper was first written in the fall of 1994 and underwent several revisions as the authors became aware of relevant previous references, etc. Since that time several further related results have been proved by the authors and by W. Schlag. In particular, the conjecture (3) has been proved by the second author in “A Kakeya-type problem for circles” (to appear in Amer. J. Math), using extensions of the techniques of the present paper.

1. The Besicovitch-Rado construction.

In this section we generalize the Besicovitch-Rado-Kinney result to \mathbb{R}^d .

Proposition 1.1. *For any $d \geq 2$ there is a closed subset $E \subset \mathbb{R}^d$, with Lebesgue measure zero which contains a sphere of every radius.*

We will follow the idea of Besicovitch and Rado, i.e., will set up a recursive Perron tree type construction where the sets to be translated are annuli rather than triangles. The Besicovitch-Rado construction can in fact be generalized to higher dimensions with essentially no changes. However, we will give a slightly different construction instead. This leads to a better estimate (at least in low dimensions) in Proposition 2.2 below.

We need a certain geometric fact, Lemma 1.1 below. In \mathbb{R}^d consider two spheres S_1 and S_0 of radii P_1 and P_0 , where $P_1 > P_0$. Suppose these spheres have centers C_1 and C_0 respectively such that $|C_1 - C_0| < P_1 - P_0$, and let $\alpha = P_1 - P_0 - |C_1 - C_0|$. Consider a ray, R , which emanates from C_1 , and let θ be the angle between R and the ray through the two centers C_1 and C_0 which also originates at C_1 .

Lemma 1.1. *The length of the segment of R between S_0 and S_1 is*

$$\approx \alpha + \frac{1}{2}\theta^2|C_1 - C_0| \left(1 + \frac{|C_1 - C_0|}{P_0} \right),$$

when $\theta \frac{|C_1 - C_0|}{P_0} \ll 1$ and $\theta \ll 1$.

Proof. Let x denote the portion of R between C_1 and S_0 . What we wish to calculate is $P_1 - |x|$. By the law of cosines $P_0^2 = |C_1 - C_0|^2 + |x|^2 - 2|x| |C_1 - C_0| \cos \theta$ and therefore

$$|x| = |C_1 - C_0| \cos \theta + P_0 \sqrt{1 - \frac{|C_1 - C_0|^2}{P_0^2} \sin^2 \theta}.$$

Given that $\theta \frac{|C_1 - C_0|}{P_0} \ll 1$ and $\theta \ll 1$ are small,

$$\begin{aligned} P_1 - |x| &= P_1 - P_0 \sqrt{1 - \frac{|C_1 - C_0|^2}{P_0^2} \sin^2 \theta} - |C_1 - C_0| \\ &\quad + |C_1 - C_0|(1 - \cos \theta) \\ &\geq P_1 - P_0 - |C_1 - C_0| + \frac{1}{2}\theta^2|C_1 - C_0| \left(\left(\frac{|C_1 - C_0|}{P_0} \right) \frac{\sin^2 \theta}{\theta^2} \right. \\ &\quad \left. + 2\frac{1 - \cos \theta}{\theta^2} \right) \\ &\approx \alpha + \frac{1}{2}\theta^2|C_1 - C_0| \left(\left(\frac{|C_1 - C_0|}{P_0} \right) + 1 \right). \end{aligned}$$

The inequality on the second line holds since $\sqrt{1 - x} \leq 1 - \frac{1}{2}x$ for small positive x . The reverse inequality also holds provided the constant $\frac{1}{2}$ is replaced by a slightly larger one and $\theta \frac{|C_1 - C_0|}{P_0}$ is small. This finishes the proof of the lemma. \square

The typical step in the recursive process is as follows. Let A be an annulus in \mathbb{R}^d with center x_0 , $A = \{x \in \mathbb{R}^d : P_0 \leq |x - x_0| \leq P_1\}$, and call $P_1 - P_0$ the *width* of A . We assume that $1/2 \leq P_0 < P_1 \leq 2P_0 \leq 2$ throughout; this serves to make all constants independent of P_0 and P_1 . For fixed n and fixed $e \in S^{d-1}$ we describe $\phi_e^n(A)$. First divide A into n concentric annuli (each of width $(P_1 - P_0)/n$), and translate them in the e -direction relative to the center, x_0 , of A as so: Numbering the annuli from 1 to n in order of largest to smallest, the k -th annulus is translated a distance $(k - 1)(P_1 - P_0)/n$. In this way the n new annuli intersect tangentially¹ along a ray in the e -direction; we call $\phi_e^n(A)$ a *figure* and x_0 the *center* of this figure.

Clearly the measure of a figure is smaller than that of the original annulus. Let $N_\delta(E) = \{x : \text{dist}(x, E) < \delta\}$ — the δ -neighborhood of E — and let $\Gamma_\theta^e(x_0)$ denote the cone based at x_0 of width θ in the direction of e — i.e.,

$$\Gamma_\theta^e(x_0) = \{x_0 + th : t \geq 0, h \in S^{d-1}, \text{dist}(h, e) \leq \theta\},$$

where distance is measured along S^{d-1} . Note that if x_0 is the center of $\phi_e^n(A)$ and if c_0 is the center of its smallest annulus, then

$$\begin{aligned} \phi_e^n(A) \cap \Gamma_\theta^e(x_0) &\subset \Gamma_\theta^e(x_0) \cap \\ &\{x \in \mathbb{R}^d : |x - x_0| \leq P_1\} \cap \{x \in \mathbb{R}^d : |x - c_0| \geq P_0\}. \end{aligned}$$

Lemma 1.2. *Let A, n, e be as above. Then for (absolutely) large n ,*

$$\left| N_{2(P_1 - P_0)/n} \left(\phi_e^n(A) \cap \Gamma_{2/\sqrt{n}}^e(x_0) \right) \right| \lesssim (P_1 - P_0)n^{-\frac{d+1}{2}}.$$

Proof. We calculate volume using polar coordinates. Consider any ray in $\Gamma_{2/\sqrt{n}}^e(x_0)$; we claim its intersection with $\phi_e^n(A)$ is contained in a segment of length $\approx (P_1 - P_0)/n$. Indeed by Lemma 1.1 with $\alpha = (P_1 - P_0)/n$ and $|C_1 - C_0| = (P_1 - P_0)\frac{n-1}{n}$, taking $\theta = 2/\sqrt{n}$ we see that such a segment has length $\approx \alpha = (P_1 - P_0)/n$. Then $\phi_e^n(A) \cap \Gamma_{2/\sqrt{n}}^e(x_0)$ is contained in a rectangle which has dimensions $\approx (P_1 - P_0) \cdot n^{-1} \underbrace{\times n^{-1/2} \times \dots \times n^{-1/2}}_{d-1 \text{ times}}$, and

hence so is $N_{2(P_1 - P_0)/n}(\phi_e^n(A) \cap \Gamma_{2/\sqrt{n}}^e(x_0))$.

Our construction involves a repeated application of ϕ_n^e where we vary the direction, e , of translation. Let us therefore cover S^{d-1} with a maximal $\frac{1}{C\sqrt{n}}$ -separated subset $D = \{e_1, e_2, \dots, e_M\}$, where $M \approx n^{\frac{d-1}{2}}$. Here C is a sufficiently large constant. If $A = \bigcup A_j$ is the union of non-concentric annuli A_j , let $\phi_e^n(A) = \bigcup \phi_n^e(A_j)$. □

Lemma 1.3. *Let $A = \{x \in \mathbb{R}^d : P_0 \leq |x - x_0| \leq P_1\}$, and consider points $e_k, e_{k+1}, \dots, e_{k+j} \in D$. Then for (absolutely) large n ,*

¹Besicovitch and Rado use a slightly different translation rule and obtain transverse intersections.

$$\phi_{e_{k+j}}^n \circ \dots \circ \phi_{e_{k+1}}^n \circ \phi_{e_k}^n(A) \cap \Gamma_{2/\sqrt{n}}^{e_k}(x_0) \subset N_{2(P_1-P_0)/n} \left(\phi_{e_k}^n(A) \cap \Gamma_{2/\sqrt{n}}^{e_k}(x_0) \right).$$

Proof. By construction, a point of $\phi_{e_{k+j}}^n \circ \dots \circ \phi_{e_{k+1}}^n \circ \phi_{e_k}^n(A)$ has been translated a distance no more than $(P_1 - P_0)(\frac{1}{n} + \frac{1}{n^2} + \dots + \frac{1}{n^M}) \leq 2(P_1 - P_0)/n$ from one of its preimages in $\phi_{e_k}^n(A)$. \square

The following lemma is the main step in the proof of Proposition 1.1.

Lemma 1.4. *Let R_0, R_1 be numbers such that $1/2 \leq R_0 < R_1 \leq 2R_0 \leq 2$, n a positive integer, and let $M \approx n^{\frac{d-1}{2}}$ be the cardinality of D , the maximal $\frac{1}{C\sqrt{n}}$ -separated subset of S^{d-1} . Then any annulus $\{x \in \mathbb{R}^d : R_0 \leq |x - x_0| \leq R_1\}$ may be divided into n^M closed subannuli, each of width $(R_1 - R_0)n^{-M}$, which when properly translated about yield a compact set E_n such that $|E_n| \lesssim (R_1 - R_0)/n$. Moreover $E_n \subset \{x \in \mathbb{R}^d : R_0 - (R_1 - R_0) \leq |x - x_0| \leq R_1\}$.*

Proof. Without loss of generality we may assume that $x_0 = 0$. Define $A_0 = \{x \in \mathbb{R}^d : R_0 \leq |x| \leq R_1\}$, and for $k = 1, 2, \dots, M$ let $A_k = \phi_{e_k}^n(A_{k-1})$; take $E_n = A_M$. Clearly E_n is a compact subset which is the union of n^M annuli of width $(R_1 - R_0)n^{-M}$. Also, since in general for an annulus $A = \{x \in \mathbb{R}^d : P_0 \leq |x - x_0| \leq P_1\}$, $\phi_e^n(A) \subset \{x \in \mathbb{R}^d : P_0 - \frac{n-1}{n}(P_1 - P_0) \leq |x - x_0| \leq P_1\}$, we see that $E_n \subset \{x \in \mathbb{R}^d : R_0 - (R_1 - R_0) \leq |x - x_0| \leq R_1\}$.

To obtain the desired upper bound for $|E_n|$ we devise a cover of E_n by sets whose measure we may readily estimate.

To facilitate the construction of this cover we must delineate the n^{k-1} figures of A_k . To each annulus of E_n we associate uniquely a multi-index I , the *address* of the annulus, as follows. For a given annulus of E_n its center can be written

$$(P_1 - P_0) \sum_{j=1}^M i_j \frac{e_j}{n^j} \quad e_j \in D,$$

where the coefficients i_j are integers in $[0, n - 1]$ and hence uniquely given; the address for this annulus is then $I = (i_1, i_2, \dots, i_M)$. Denote the center of this annulus as $x(I)$. Let S_k denote the set of indices of the form $(i_1, \dots, i_{k-1}, 0, \dots, 0)$.² It is easy to see that $x(I), I \in S_k$, is the center of a figure of A_k ; denote this figure as $A_k(x(I))$. With this notation we remark that if $z \in E_n$ is contained in an annulus with address (i_1, i_2, \dots, i_M) , then $z \in A_M(x(i_1, \dots, i_{M-1}, 0))$, and for any $k = 1, \dots, M$ and any $I \in S_k$, $z \in \phi_n^{e_M} \circ \dots \circ \phi_n^{e_{k+1}}(A_k(x(I)))$.

²Here $S_1 = \{(0, \dots, 0)\}$.

For each $k = 1, 2, \dots, M$ and for each $I \in S_k$ we take as an element of our proposed cover of E_n

$$(8) \quad N_{2(R_1-R_0)/n^k} \left(A_k(x(I)) \cap \Gamma_{2/\sqrt{n}}^{e_k}(x(I)) \right).$$

By Lemma 1.2 with $P_1 - P_0 = (R_1 - R_0)/n^{k-1}$,

$$\left| N_{2(R_1-R_0)/n^k} \left(A_k(x(I)) \cap \Gamma_{2/\sqrt{n}}^{e_k}(x(I)) \right) \right| \lesssim \frac{1}{n^{k-1}} \cdot \frac{(R_1 - R_0)}{n^{\frac{d+1}{2}}},$$

and so

$$(9) \quad \left| \bigcup_{I \in S_k} N_{2(R_1-R_0)/n^k} \left(A_k(x(I)) \cap \Gamma_{2/\sqrt{n}}^{e_k}(x(I)) \right) \right| \lesssim \frac{(R_1 - R_0)}{n^{\frac{d+1}{2}}}.$$

The claim is that the collection of all sets defined in (8) covers E_n . Let $z \in E_n$ be given and choose an index $I = (i_1, \dots, i_{M-1}, 0) \in S_M$ such that $z \in A_M(x(I))$ —i.e., find one of the n^{M-1} figures of E_n that z lies in. If C is large enough then

$$\begin{aligned} \mathbb{R}^d &= \Gamma_{1/\sqrt{n}}^{e_1}(0) \cup \Gamma_{1/\sqrt{n}}^{e_2}(x(i_1, 0, \dots, 0)) \cup \Gamma_{1/\sqrt{n}}^{e_3}(x(i_1, i_2, 0, \dots, 0)) \cup \dots \\ &\quad \cup \Gamma_{1/\sqrt{n}}^{e_M}(x(i_1, \dots, i_{M-1}, 0)). \end{aligned}$$

This is easy to see if one remembers that $x(i_1, \dots, i_k, 0, \dots, 0)$ and $x(i_1, \dots, i_{k+1}, 0, \dots, 0)$ are very close together when $k \geq 1$.

So choose k such that $z \in \Gamma_{2/\sqrt{n}}^{e_k}(x(i_1, \dots, i_{k-1}, 0, \dots, 0))$. The claim is that

$$z \in N_{\frac{2(R_1-R_0)}{n^k}} \left(A_k(x(i_1, \dots, i_{k-1}, 0, \dots, 0)) \cap \Gamma_{2/\sqrt{n}}^{e_k}(x(i_1, \dots, i_{k-1}, 0, \dots, 0)) \right).$$

This is established by Lemma 1.3 with $P_1 - P_0 = (R_1 - R_0)/n^{k-1}$ once we recall, given the above remarks, that for this choice of k , $z \in \phi_n^{e_M} \circ \dots \circ \phi_n^{e_{k+1}}(A_k(x ik))$.

Finally by (9)

$$\begin{aligned} |E_n| &\leq \sum_{k=1}^M \sum_{I \in S_k} \left| N_{2(R_1-R_0)/n^k} \left(A_k(x(I)) \cap \Gamma_{2/\sqrt{n}}^{e_k}(x(I)) \right) \right| \\ &\lesssim (R_1 - R_0) \underbrace{\left(n^{-\frac{d+1}{2}} + n^{-\frac{d+1}{2}} + \dots + n^{-\frac{d+1}{2}} \right)}_{M \text{ times}} \approx (R_1 - R_0)n^{-1}, \end{aligned}$$

which establishes the lemma. □

The rest of the proof of Proposition 1.1 follows a standard pattern, and furthermore is identical to the corresponding argument in [1], so we will omit it.

2. The higher dimensional case.

The purpose of this section is to prove Theorem 1'. Before doing so we show that the result is sharp except for the logarithmic factor (for this factor, see Proposition 2.2).

Proposition 2.1. $\|M_\delta\|_{L^p \rightarrow L^q} \gtrsim \max\{\delta^{-(\frac{1}{p}-\frac{1}{q})}, \delta^{-(\frac{1}{p}-\frac{1}{q})}\}.$

Proof. Let E_δ be a $C^{-1}\delta$ -neighborhood of the set $\{x \in \mathbb{R}^d : x_d = 0, |x| = 1/2\}$, and $f = \chi_{E_\delta}$. Then $\|f\|_p = |E_\delta|^{1/p} \approx \delta^{2/p}$. If C is large then E_δ is contained in $C_\delta(y_r, r)$ for any $r \in [1/2, 2]$, where $y_r = (0, \dots, 0, \sqrt{r^2 - 1/4}) \in \mathbb{R}^d$, and therefore $M_\delta f(r) \geq \frac{|E_\delta|}{|C_\delta(y_r, r)|} \approx \delta = \delta^{-(\frac{1}{p}-\frac{1}{q})} \|f\|_p$ for every r . On the other hand let $f = \chi_{C_\delta(0,1)}$. Then $\|f\|_p = \delta^{1/p}$, while for $r \in [1 - \delta, 1 + \delta]$ $M_\delta f(r) \approx 1$. Thus $\|M_\delta f\|_q \gtrsim \delta^{1/q} = \delta^{1/q-1/p} \|f\|_p$. \square

The proof of Theorem 1' necessitates a lemma, whose proof we omit.

Lemma 2.1. *If $d \geq 3$, $1/2 \leq r, s \leq 2$, then for any $a, b \in \mathbb{R}^d$ we have*

$$|C_\delta(a, r) \cap C_\delta(b, s)| \leq C \frac{\delta^2}{|r - s| + \delta}.$$

Proof of Theorem 1'. The $L^p \rightarrow L^\infty$ case is trivial. By interpolation and by Hölder's inequality it suffices to prove the theorem in the $L^p \rightarrow L^{p'}$ case, $1 < p \leq 2$.

First we "discretize": If $|s-r| < \delta$ then $M_\delta f(s) \lesssim M_{3\delta} f(r)$ since $C_\delta(x, s) \subset C_{3\delta}(x, r)$. Consequently it suffices to show that

$$\begin{aligned} \left(\sum_j \delta |M_\delta f(r_j)|^{p'} \right)^{\frac{1}{p'}} &\lesssim \delta^{1-\frac{2}{p}} \|f\|_p, \quad 1 < p < 2 \\ \left(\sum_j \delta |M_\delta f(r_j)|^2 \right)^{\frac{1}{2}} &\lesssim \left(\log \left(\frac{1}{\delta} \right) \right)^{\frac{1}{2}} \|f\|_2, \end{aligned}$$

where say $r_j = 1/2 + j\delta$, $0 \leq j \leq 3/2\delta$.

Next we use duality: For suitable nonnegative numbers a_j with $\sum \delta a_j^p = 1$ and suitable points $x_j \in \mathbb{R}^d$ we have (with $1 < p \leq 2$)

$$\begin{aligned} \left(\sum_j \delta |M_\delta f(r_j)|^{p'} \right)^{1/p'} &= \sum_j \delta a_j M_\delta f(r_j) \\ &= \int f \cdot \sum_j b_j \chi_j, \end{aligned}$$

where $b_j = \frac{\delta a_j}{|C_\delta(x_j, r_j)|}$ satisfies $\sum \delta b_j^p \sim \sum \delta a_j^p = 1$, and $\chi_j = \chi_{C_\delta(x_j, r_j)}$. It therefore suffices by Hölder's inequality to show that

$$\left\| \sum_j b_j \chi_j \right\|_{p'} \lesssim \delta^{1-2/p} \left(\sum_j \delta b_j^p \right)^{1/p}, \quad 1 < p < 2$$

$$\left\| \sum_j b_j \chi_j \right\|_2 \lesssim (\log(1/\delta))^{1/2} \left(\sum_j \delta b_j^2 \right)^{1/2}.$$

Let $\chi_{jk} = \chi_{C_\delta(x_j, r_j) \cap C_\delta(x_k, r_k)}$. For $1 < p \leq 2$

$$\begin{aligned} \left\| \sum_j b_j \chi_j \right\|_{p'}^2 &= \left(\int \left| \sum_{jk} b_k b_j \chi_{jk}(y) \right|^{p'/2} dy \right)^{2/p'} \\ &\leq \sum_{jk} b_k b_j \left(\int |\chi_{jk}(y)|^{p'/2} dy \right)^{2/p'} \\ &\lesssim \sum_{jk} b_k b_j \left(\frac{\delta}{1 + |j - k|} \right)^{2/p'} \quad \text{by Lemma 2.1} \\ &\leq \delta^{2(1-2/p)} \left(\sum_k \delta b_k^p \right)^{1/p} \left\| \sum_j \delta^{1/p} b_j (1 + |j - k|)^{-2/p'} \right\|_{l^{p'}}. \end{aligned}$$

On the one hand, by the Hardy-Littlewood-Sobolev inequality for the integrals³, when $1 < p < 2$,

$$\left\| \sum_j \delta^{1/p} b_j (1 + |j - k|)^{-2/p'} \right\|_{l^{p'}} \lesssim \left(\sum_j \delta b_j^p \right)^{1/p}.$$

On the other hand, by Schur's test

$$\left\| \sum_j \delta^{1/2} b_j (1 + |j - k|)^{-1} \right\|_{l^2} \lesssim (\log(1/\delta))^{1/2} \left(\sum_j \delta |b_j|^2 \right)^{1/2},$$

and the theorem is proved. □

Proposition 2.2. *A necessary condition for an estimate of the form*

$$\|M_\delta f\|_{L^q([1/2, 2])} \leq C(\log(1/\delta))^\alpha \|f\|_{L^p(\mathbb{R}^d)}$$

to hold is that $\alpha \geq \frac{2}{(d-1)p}$.

³The referee points out that this was the original version proved by Hardy and Littlewood.

Proof. Let E_n be the set described in Lemma 1.4 (with $R_0=1/2$ and $R_1=2$). Take $\delta = n^{-M}$ ($M \approx n^{\frac{d-1}{2}}$) and let $f = \chi_{E_n}$. Then $|E_n| \lesssim \frac{1}{n}$, but $M_\delta f(r) \gtrsim 1$ for $r \in [1/2, 2]$. Thus

$$1 \lesssim (\log(n^M))^\alpha n^{-\frac{1}{p}} = n^{\frac{\alpha(d-1)}{2} - \frac{1}{p}} (\log(n))^\alpha,$$

which is only possible when $\alpha \geq \frac{2}{(d-1)p}$. □

Remark. When $d = 3$, this shows that the powers of $\log \frac{1}{\delta}$ in Theorem 1' are the optimal ones. Taking $d = 2, p = 3$ it also gives a lower bound $\frac{2}{3}$ for the number of logarithms which would be needed in (3). Note that this shows that Theorem 1' cannot hold as stated when $d = 2$.

Further remarks.

(1) Theorem 1' (and 2') have equivalent formulations in terms of the maximal function $u^*(t) = \sup_x |u(x, t)|$, where u is a solution of the wave equation. The argument relating the two maximal functions is a standard one and we will only sketch it. Namely, let $C_f(x, t)$ be the average of f over the sphere centered at x with radius t . If (in d space dimensions) $u_{tt} - \Delta_x u = 0$ and $u(\cdot, 0) = f, u_t(\cdot, 0) = g$, and if $\hat{f}(\xi)$ and $\hat{g}(\xi)$ vanish when $|\xi| \leq 1$ then one can obtain $u(\cdot, t)$ in the form

$$(10) \quad \mathcal{A}_t \left(C_{(-\Delta)^{\frac{d-1}{4}}} f(\cdot, t) \right) + \mathcal{B}_t \left(C_{(-\Delta)^{\frac{d-3}{4}}} g(\cdot, t) \right) \\ + \mathcal{C}_t \left(\frac{d}{dt} C_{(-\Delta)^{\frac{d-3}{4}}} f(\cdot, t) \right) + \mathcal{D}_t \left(\frac{d}{dt} C_{(-\Delta)^{\frac{d-5}{4}}} g(\cdot, t) \right)$$

where $\mathcal{A}_t, \mathcal{B}_t, \mathcal{C}_t, \mathcal{D}_t$ are operators which are bounded on L^∞ , uniformly in $t \in [\frac{1}{2}, 2]$. This may be seen by taking Fourier transforms and using the asymptotics for the Fourier transform of surface measure and the fact that a multiplier m supported outside the unit disc with $D^\alpha m(\xi) = \mathcal{O}(|\xi|^{-(|\alpha|+1)})$ at ∞ will be L^∞ bounded. Furthermore, by differentiation under the integral sign

$$\left| \frac{d}{dt} C_{(-\Delta)^{\frac{d-3}{4}}} f(x, t) \right| \lesssim \sum_{i=1}^d C_{|R_i(-\Delta)^{\frac{d-1}{4}}} f(x, t)$$

where the R_i are the Riesz transforms, and similarly with the fourth term in (10). It follows (with $R_0 \stackrel{\text{def}}{=} \text{identity operator}$) that

$$u^*(t) \lesssim \sup_x \sum_{i=0}^d C_{|R_i(-\Delta)^{\frac{d-1}{4}}} f(x, t) + C_{|R_i(-\Delta)^{\frac{d-3}{4}}} g(x, t),$$

so that any estimate for M_δ gives a corresponding estimate for u^* . For example the first statement in Theorem 1' corresponds in this way to the

estimate

$$(11) \quad \|u^*\|_{L^2(\left(\frac{1}{2}, 2\right])} \lesssim \|f\|_\alpha + \|g\|_{\alpha-1}, \quad \alpha > \frac{d-1}{2}, \quad d \geq 3.$$

Here $\|\cdot\|_\alpha$ is the inhomogeneous L^2 Sobolev space norm with α derivatives. One would expect (11) to be a well-known estimate, although we only know a reference in the three dimensional case (namely [9]).

(2) One can also extend the domain of $M_\delta f$ from $[1/2, 2]$ to $(0, \infty)$. This of course requires modifying the definition of M_δ . We let

$$C'_\delta(x, r) = \{z \in \mathbb{R}^d : r(1 - \delta) \leq |z - x| \leq r(1 + \delta)\}.$$

Then if $f : \mathbb{R}^d \rightarrow \mathbb{R}$ we define $M'_\delta f : (0, \infty) \rightarrow \mathbb{R}$ via

$$M'_\delta f(r) = \sup_x \frac{1}{|C'_\delta(x, r)|} \int_{C'_\delta(x, r)} |f|.$$

Theorem 1''. *Assume $d \geq 3$. Then*

$$\left(\int_0^\infty |M'_\delta f(r)|^2 r^d \frac{dr}{r} \right)^{1/2} \leq C \log \left(\frac{1}{\delta} \right)^{\frac{3}{2}} \|f\|_{L^2(\mathbb{R}^d)}.$$

If $1 < p < 2$ then

$$\left(\int_0^\infty |M'_\delta f(r)|^{p'} r^{\frac{d}{p-1}} \frac{dr}{r} \right)^{1/p'} \leq C_p \delta^{-\left(\frac{1}{p} - \frac{1}{p'}\right)} \|f\|_{L^p(\mathbb{R}^d)}.$$

This is proved in a standard way using Theorem 1' and a small amount of Littlewood-Paley theory.

3. Proof of Theorem 2', part 1.

Before giving the proof of Theorem 2' we want to make some remarks concerning the difference between the two- and higher-dimensional cases. Note to begin with that Lemma 2.1 fails in the two dimensional case: The corresponding estimate is

$$(12) \quad |C_\delta(x, r) \cap C_\delta(y, s)| \lesssim \frac{\delta^{\frac{3}{2}}}{(|r - s| + \delta)^{\frac{1}{2}}}.$$

This is sharp when the circles $\{z : |z - x| = r\}$ and $\{z : |z - y| = s\}$ are tangent. The bound (12) leads via the argument of Section 2 to the estimate

$$(13) \quad \|M_\delta f\|_2 \lesssim \delta^{-\frac{1}{4}} \|f\|_2.$$

This is the optimal $L^2 \rightarrow L^2$ estimate (by the example in the next paragraph) but implies only the bound $\dim E \geq \frac{3}{2}$ for the Hausdorff dimension of a set containing a circle of every radius.

To see why (3) is the natural conjecture to make, let E_δ be a $\delta \times \delta^{\frac{1}{2}}$ rectangle and let f be the characteristic function of E_δ . For any $r \in [\frac{1}{2}, 2]$ there is an annulus $C_\delta(x, r)$ which contains a fixed fraction of E_δ (namely: Choose x so that the circle $\{z : |z - x| = r\}$ is tangent to a long side of E_δ) and therefore $M_\delta f(r) \gtrsim \delta^{\frac{1}{2}}$. It follows that a bound

$$\forall \epsilon \exists C_\epsilon : \|M_\delta f\|_p \leq C_\epsilon \delta^{-\epsilon} \|f\|_p$$

can only hold if $\delta^{\frac{1}{2}} \leq \delta^{\frac{3}{2p}}$, i.e., $p \geq 3$. Similar considerations show that the bound (4) would be best possible for each p and q , and the example $f =$ characteristic function of $C_\delta(0, 1)$ shows that the range $q \leq 2p'$ in (4) would also be best possible.

As we pointed out in the introduction, the $p = 2, q = 4$ case of Theorem 2' is due to Pecher [9]. When $p > 2$ it is natural to use a geometric argument in order to deal with the problems arising from tangent circles. Our way of doing this will be based on a geometric fact used also by Marstrand [7], together with a well-known argument from elementary combinatorics. In order to avoid certain irrelevant technicalities, let us say that two circles are internally tangent if they are tangent and one is contained in the bounded component of the complement of the other. Then we have the following geometrical fact, which we call (M), and which we believe was first used in a similar context in [7]: Given three circles C_1, C_2, C_3 , there are at most two circles which are internally tangent to C_1, C_2 and C_3 at three different points. It is easy to see that this leads to control over the number of possible tangencies among N circles, N large:

Proposition 3.1. *Let C_1, \dots, C_N be circles. Let k be a number and assume: For each of at least $\frac{N}{2}$ values of j , we are given k values of i such that C_i is internally tangent to C_j with (for fixed j) no more than $\frac{k}{3}$ of these tangencies occurring at any one point. Then $k \lesssim N^{\frac{2}{3}}$.*

Proof. This basically follows by combining fact (M) with [4], Theorem 3 on p. 111: An $N \times N$ $(0, 1)$ matrix with no 3×3 submatrix of 1's contains $\lesssim N^{\frac{5}{3}}$ 1's altogether. However, we have stated the proposition in a way which makes the result from [4] not directly applicable, so we give a proof (assuming (M)). Let Q be the set of all quadruples (j, i_1, i_2, i_3) such that C_j is internally tangent to C_{i_1}, C_{i_2} and C_{i_3} at three different points. Then $\text{card}(Q) \leq 2N^3$, since for any given i_1, i_2, i_3 , (M) implies there are at most two choices for j . On the other hand, $\text{card}(Q) \geq \frac{N}{2} \cdot k \cdot \frac{2k}{3} \cdot \frac{k}{3}$, since if we take j to be one of the $\frac{N}{2}$ values in the statement, then there are at least k possibilities for i_1 , and then $k - \frac{k}{3}$ possibilities for i_2 and $k - \frac{2k}{3}$ possibilities for i_3 . Hence $2N^3 \geq \frac{Nk^3}{9}$ and the proposition follows. \square

The proof of Theorem 2' will be basically a quantitative version of the proof of Proposition 3.1. We note that the example of circles with integer radius ($\leq n^{\frac{1}{3}}$) and center at an integer lattice point (lying in $D(0, n^{\frac{1}{3}})$) shows that the optimal exponent to replace $\frac{2}{3}$ in the proposition must be at least $\frac{1}{3}$, since tangencies may be identified with pythagorean triples.

Added 5/15/97. More sophisticated methods for dealing with combinatorial problems like the one in Proposition 3.1 have been developed in the discrete geometry literature and can be applied in connection with (3). See the remark at the end of the introduction.

In fact, as mentioned in the introduction we will work in the general situation of a family of curves satisfying the “cinematic curvature hypothesis” from [10], instead of with circles.

Let $\Phi : U \rightarrow \mathbb{R}$, U a neighborhood of (a, b) in $\mathbb{R}^2 \times \mathbb{R}^2$ be a function satisfying the following conditions:

$$(14) \quad \nabla_x \Phi|_{(x,y)=(a,b)} \neq 0.$$

$$(15) \quad \text{If } e \text{ is a unit vector with } e \perp \nabla_x \Phi(a, b),$$

$$\text{then } d_y \left(\frac{\nabla_x \Phi \cdot e}{\frac{\nabla_x^2 \Phi(e,e)}{|\nabla_x \Phi|}} \right) \Big|_{(x,y)=(a,b)} \text{ is invertible.}$$

We use the notation $d_y f$ for the differential of f with respect to y and also use $\nabla_y \Phi$ interchangeably with $d_y \Phi$ if Φ maps into \mathbb{R} , and $\nabla_y^2 \Phi$ for the Hessian of Φ in the y variable, etc.

We define

$$C(y, r) = \{x \in \mathbb{R}^2 : \Phi(x, y) = r\}.$$

Condition (15) has a number of equivalent formulations, cf. [10], [11]. We recall: If y and z are close to b and $C(y, r)$ and $C(z, s)$ intersect near a then (since Φ is smooth) $|r - s| \lesssim |y - z|$. If we assume (14) then (locally) the $C(y, r)$ are smooth curves, and (15) is equivalent to the following condition: If $|y - b|, |r - \Phi(a, b)|, |z - b|, |s - \Phi(a, b)|$ and $|x - a|$ are small and $C(y, r)$ and $C(z, s)$ intersect at x then either the unit tangent vectors to $C(y, r)$ and $C(z, s)$ at x , or else their curvatures at x , must differ by $\gtrsim |y - z| + |r - s|$.

In what follows, $\mathcal{Q}(\beta)$ will always mean the set $\{(z, s, y, r) \in \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}^2 \times \mathbb{R} : \max(|z - b|, |s - \Phi(a, b)|, |y - b|, |r - \Phi(a, b)|) < \beta\}$. We fix small positive numbers $\alpha_2 \ll \alpha_1 \ll \alpha_0$ and β_0 , and define a function $\Delta : \mathcal{Q}(\beta_0) \rightarrow \mathbb{R}$ via

$$\Delta(z, s, y, r) = \inf_{\substack{x \in C(y,r) \cap D(a,\alpha_1) \\ x' \in C(z,s) \cap D(a,\alpha_1)}} |x - x'| + \left| \frac{\nabla_x \Phi(x, y)}{|\nabla_x \Phi(x, y)|} - \frac{\nabla_x \Phi(x', z)}{|\nabla_x \Phi(x', z)|} \right|.$$

Thus $\Delta(z, s, y, r)$ measures the minimum distance between the curves obtained by lifting $C(z, s)$ and $C(y, r)$ to the unit tangent bundle. We further define (for small δ)

$$C_\delta(y, r) = \{x \in D(0, \alpha_2) : |\Phi(x, y) - r| < \delta\}.$$

The constants in Lemma 3.1 below are of course independent of δ, ϵ, t and λ .

Lemma 3.1 below is the main result of this section. In (19), $N_\delta(S^*)$ is the δ -entropy of S^* , i.e., the maximum possible cardinality for a δ -separated subset of S . Part (i) and estimate (19) are what are needed in the subsequent sections; the remaining statements follow from the same proof and are included for the sake of possible future applications.

Lemma 3.1. *There are $\beta_1 > 0$ and $A_1 < \infty$ such that if $\delta > 0$ is small enough then:*

(i) *If $(z, s, y, r) \in \mathcal{Q}(\beta_1)$ then*

$$(16) \quad |C_\delta(z, s) \cap C_\delta(y, r)| \lesssim \frac{\delta^2}{\sqrt{(|y - z| + |r - s| + \delta)(\Delta(z, s, y, r) + \delta)}}$$

$$(17) \quad \text{diam}(C_\delta(z, s) \cap C_\delta(y, r)) \lesssim \sqrt{\frac{\Delta(z, s, y, r) + \delta}{|y - z| + |r - s| + \delta}}.$$

(ii) *Let $\epsilon, t, \lambda \in (0, 1)$ be such that $\epsilon \geq \delta$ and $\lambda\sqrt{\frac{t}{\epsilon}} \geq A_1$. For fixed $(z_i, r_i) \in \mathbb{R}^2 \times \mathbb{R}, i = 1, 2, 3$ with $\max(|z_i - b|, |r_i - \Phi(a, b)|) < \beta_1$, define*

$$(18) \quad S = \{(y, r) \in \mathbb{R}^2 \times \mathbb{R} : (z_i, r_i, y, r) \in \mathcal{Q}(\beta_1), \Delta(z_i, r_i, y, r) < \epsilon \forall i, \\ |y - z_i| + |r - r_i| > t \forall i, C_\delta(z_i, r_i) \cap C_\delta(y, r) \neq \emptyset \forall i, \\ \text{dist}(C_\delta(z_i, r_i) \cap C_\delta(y, r), C_\delta(z_j, r_j) \cap C_\delta(y, r)) > \lambda \forall (i, j) : i \neq j\}.$$

Then S is the union of two sets, each of volume $\lesssim \frac{\epsilon^3}{\lambda^3}$ and diameter $\lesssim \frac{\epsilon}{\lambda^2}$. In particular, if $S^ = \{r \in \mathbb{R} : (y, r) \in S \text{ for some } y \in \mathbb{R}^2\}$, then*

$$(19) \quad N_\delta(S^*) \lesssim \frac{\epsilon}{\delta\lambda^2}.$$

Remarks.

(1) The main point is part (ii); part (i) is implicit in [10] (as are several of the sublemmas below) and in any case is quite simple.

(2) The case of circles corresponds of course to $\Phi(x, y) = |x - y|, a \neq b$. The function $\Delta(z, s, y, r)$ then satisfies $\Delta(z, s, y, r) \gtrsim ||z - y| - |r - s||$. In this case Lemma 3.1 is essentially in [7] (cf. Lemma 5.2 there), but some of the conclusions of Lemma 3.1 are not stated there. It may therefore be worth recording the following “explicit” form of Lemma 3.1 (ii) for circles even though we will not use it below. It is not quite a formal consequence

of Lemma 3.1 but can be proved by a much abbreviated version of the proof of Lemma 3.1 - as has already mentioned, it is also implicit in [7].

Fix three circles $C(x_i, r_i) \stackrel{\text{def}}{=} \{z \in \mathbb{R}^2 : |z - x_i| = r_i\}$, where $r_i \in [\frac{1}{2}, 2]$, and let $C_\delta(x, r) = \{z \in \mathbb{R}^2 : r - \delta < |z - x| < r + \delta\}$. Fix $\epsilon, t, \lambda \in (0, 1)$ such that $\epsilon \geq \delta$ and $\lambda \sqrt{\frac{t}{\epsilon}} \geq A_1$. Define two annuli $C_\delta(x, r)$ and $C_\delta(y, s)$ to be interior (ϵ, t) -tangent if their intersection is nonempty and also $||x - y| - |r - s|| < \epsilon$ and $|x - y| + |r - s| > t$. Let $\Omega_{\epsilon t \lambda}$ be the set of all $(x, r) \in \mathbb{R}^2 \times [\frac{1}{2}, 2]$ such that $C_\delta(x, r)$ is interior (ϵ, t) -tangent to all three annuli $C_\delta(x_i, r_i)$ and furthermore the distance between any two of the sets $C_\delta(x_i, r_i) \cap C_\delta(x, r)$ and $C_\delta(x_j, r_j) \cap C_\delta(x, r)$ ($1 \leq i, j \leq 3, i \neq j$) is at least λ . Then $\Omega_{\epsilon t \lambda}$ is contained in the union of two ellipsoids in \mathbb{R}^3 each of volume $\lesssim \frac{\epsilon^3}{\lambda^3}$ and diameter $\lesssim \frac{\epsilon}{\lambda^2}$.

(3) In the proof of Lemma 3.1 we often assume that $a = b = 0$ and $\Phi(a, b) = 0$.

We start the proof by making some further observations about the conditions (14), (15). First, it is clear that

$$(20) \quad C_\delta(z, s) \cap C_\delta(y, r) = \emptyset \text{ if } |r - s| > C|y - z| + 2\delta$$

for a suitable constant C . (14), (15) are invariant under changes of coordinates in the x variable, so for many purposes we may assume that $\Phi(x, 0) = x_{(2)}$. Here we use the notation $x_{(2)}$ to mean the second coordinate of the point $x \in \mathbb{R}^2$. If $\Phi(x, 0) = x_{(2)}$, then (15) is equivalent with

$$(21) \quad C^{-1}|y| \leq \left| \frac{\partial \Phi}{\partial x_{(1)}} \right| + \left| \frac{\partial^2 \Phi}{\partial x_{(1)}^2} \right| \leq C|y|$$

for suitable C , locally near $(0, 0) \in \mathbb{R}^2 \times \mathbb{R}^2$. Next we have the following (perhaps well-known) fact, which will be used in the last part of the proof of Lemma 3.1.

Lemma 3.2. *Assume that Φ satisfies (14), (15). Suppose that $x_1, x_2, x_3, y \in \mathbb{R}^2$ are sufficiently close to the origin and satisfy $\Phi(x_1, y) = \Phi(x_2, y) = \Phi(x_3, y)$. Then*

$$\left| \det \begin{pmatrix} -\nabla_y \Phi(x_1, y) & 1 \\ -\nabla_y \Phi(x_2, y) & 1 \\ -\nabla_y \Phi(x_3, y) & 1 \end{pmatrix} \right| \approx |x_1 - x_2| |x_2 - x_3| |x_3 - x_1|.$$

Proof. We may assume that $y = 0$ and $\Phi(x, 0) = x_{(2)}$, and furthermore that $x_1 = (0, 0)$, $x_2 = (\sigma, 0)$, $x_3 = (\tau, 0)$, with $\sigma < 0 < \tau$. Define $v(t) = \nabla_y \Phi((t, 0), 0)$. Thus $v(t)$ is an \mathbb{R}^2 -valued function on an interval. Condition (15) implies that

$$\dot{v}(0) \wedge \ddot{v}(0) \neq 0,$$

where we have denoted $\frac{dv}{dt}$ by \dot{v} etc. Consequently, by Taylor's theorem,

$$v(\tau) - v(0) = \tau\dot{v}(0) + \frac{\tau^2}{2}\ddot{v}(0) + \mathcal{O}(\tau^3)$$

$$v(\sigma) - v(0) = \sigma\dot{v}(0) + \frac{\sigma^2}{2}\ddot{v}(0) + \mathcal{O}(|\sigma|^3),$$

and therefore

$$\left| \det \begin{pmatrix} -\nabla_y \Phi(x_1, 0) & 1 \\ -\nabla_y \Phi(x_2, 0) & 1 \\ -\nabla_y \Phi(x_3, 0) & 1 \end{pmatrix} \right| = |(v(\tau) - v(0)) \wedge (v(\sigma) - v(0))|$$

$$= \frac{1}{2}\tau|\sigma|(\tau - \sigma)|\dot{v}(0) \wedge \ddot{v}(0)| + \mathcal{O}(|\tau\sigma|(\tau^2 + \sigma^2)).$$

The lemma follows, since $\sigma < 0 < \tau$ so that $\tau^2 + \sigma^2$ is small compared with $\tau - \sigma$. □

Next we prove an elementary lemma about functions of one variable. The implicit constants depend on C_0 only.

Lemma 3.3. *Let $I \subset \mathbb{R}$ be an interval, \tilde{I} the double of I , I_0 the double of \tilde{I} , and let ϕ be a smooth real-valued function on I_0 . Assume that*

$$(22) \quad C_0^{-1}\rho \leq |\dot{\phi}(t)| + |\ddot{\phi}(t)| \leq C_0\rho$$

for all $t \in I_0$. Let $\Delta = \min_{t \in \tilde{I}} (|\phi(t)| + |\dot{\phi}(t)|)$. Then provided $|I|$ and η_0 are small enough we have:

- (i) If $\Delta \leq \eta_0\rho$, then $\dot{\phi}(\tau) = 0$ for a unique $\tau \in I_0$. Furthermore $|\phi(\tau)| \lesssim \Delta$.
- (ii) If $\tau \in I_0$ and $\dot{\phi}(\tau) = 0$, then

$$(23) \quad |\dot{\phi}(t)| \approx \rho|t - \tau|$$

and

$$(24) \quad |\phi(t) - \phi(\tau)| \approx \rho|t - \tau|^2$$

for all $t \in I_0$.

- (iii) $n(\phi, c) \leq 2$ for all $c \in \mathbb{R}$.
- (iv) Let $E = \{(t, y) \in \mathbb{R}^2 : t \in I : |\phi(t)| < \delta, \text{ and } |y| < \delta\}$. Then $|E| \lesssim \frac{\delta^2}{\sqrt{(\rho+\delta)(\Delta+\delta)}}$ and $\text{diam}(E) \lesssim \sqrt{\frac{\Delta+\delta}{\rho+\delta}}$. If $\Delta \lesssim \eta_0\rho$, then E is contained in a disc of radius $\lesssim \sqrt{\frac{\Delta+\delta}{\rho+\delta}}$ centered at $(\tau, \phi(\tau))$.

In (iii) $n(\phi, c)$ is the multiplicity of c as a value of ϕ , defined in the usual way, i.e.,

$$n(\phi, c) = \sum_{t \in I_0: \phi(t)=c} \min(\{k \geq 1 : \phi^{(k)}(t) \neq 0\}),$$

where the minimum of the empty set is $+\infty$. We will also denote $n(\phi, 0)$ by $n(\phi)$.

Proof. (i) If $\eta_0 + C_0|I_0| \leq \frac{1}{2}C_0^{-1}$ then the assumption (22) implies that $|\dot{\phi}(s)| \leq \frac{1}{2}C_0^{-1}\rho$ for all $s \in I_0$, and therefore that $|\ddot{\phi}(s)| \geq \frac{1}{2}C_0^{-1}\rho$ for all $s \in I_0$. In particular $\ddot{\phi}$ does not change sign. If $t \in I$ with $|\dot{\phi}(t)| \leq \Delta$ then (assuming as we may that $\ddot{\phi} > 0$) we have $\frac{\dot{\phi}(s) - \dot{\phi}(t)}{s-t} \geq \frac{1}{2}C_0^{-1}\rho$ for all $s \in I_0$. If $\frac{|I_0|}{8C_0} \geq \eta_0$ then it follows that $\dot{\phi}$ changes sign, i.e., we have proved the existence part of (i). We will prove the rest of (i) after proving (ii).

(ii) If τ is as in (ii) then by the same argument as above $|\dot{\phi}(t) - \dot{\phi}(\tau)| \approx |t - \tau|\rho$, and therefore also $|\phi(t) - \phi(\tau)| \approx |t - \tau|^2\rho$.

The uniqueness part of (i) clearly follows from (ii). Furthermore, if $t \in I$ and $|\phi(t)| + |\dot{\phi}(t)| = \Delta$, then by (ii), $|\tau - t| \lesssim \frac{\Delta}{\rho}$ and therefore also $|\phi(\tau) - \phi(t)| \lesssim \frac{\Delta^2}{\rho}$, so $\phi(\tau) \leq \Delta + C\frac{\Delta^2}{\rho} \lesssim \Delta$.

(iii) Either $\dot{\phi}$ vanishes somewhere or it doesn't. If $\dot{\phi}$ never vanishes then $n(\phi, c) \leq 1$ for all c . If $\dot{\phi}$ vanishes at τ then by the above considerations ϕ is either convex or concave and furthermore $\ddot{\phi}(\tau) \neq 0$. This implies (iii).

(iv) Let $E^* \stackrel{\text{def}}{=} \{t \in I : |\phi(t)| < \delta\}$, i.e., E^* is the projection of E on the t axis.

We can assume that $\Delta \leq C_0\rho + \delta$, since otherwise E is clearly empty. Given this, we may also assume that $\rho \geq 2\eta_0^{-1}\delta$, since if $\Delta \lesssim \rho + \delta \lesssim \delta$ then all the statements in (iv) just follow since $E^* \subset I$.

Suppose first that $\Delta \leq \eta_0\rho$. Let τ be as in (i). If $t \in E^*$ then by (24),

$$(25) \quad |t - \tau| \approx \sqrt{\frac{|\phi(t) - \phi(\tau)|}{\rho}} \lesssim \sqrt{\frac{\Delta + \delta}{\rho}} \approx \sqrt{\frac{\Delta + \delta}{\rho + \delta}}.$$

This implies the last statement of (iv) and therefore also the diameter bound. If $\Delta \leq 2\delta$ then the measure bound follows from the diameter bound and we're done. It remains to prove the measure bound when $\Delta \geq 2\delta$. We may assume that $\tau \in \tilde{I}$, since otherwise E is empty by (25) if η_0 has been chosen small enough. But if $\tau \in \tilde{I}$ then $|\phi(\tau)| \geq \Delta$, by definition of Δ . Since $\Delta \geq 2\delta$, it then follows by (24) that $|t - \tau| \approx \sqrt{\frac{\Delta}{\rho}}$, $t \in E^*$. Hence (by (23)) $|\dot{\phi}(t)| \approx \sqrt{\rho\Delta}$, and then the measure bound follows using (iii).

Now suppose that $\Delta \geq \eta_0 \rho$. Then the diameter bound in (iv) follows since $E^* \subset I$, and the measure bound follows (using (iii)) since $|\dot{\phi}(t)| \geq \Delta - \delta \geq \eta_0 \rho - \delta \gtrsim \rho$.

Proof of part (i) of Lemma 3.1. This follows from (iv) of Lemma 3.3 by choosing appropriate local coordinates. Namely, make a local change of variable on $D(0, \alpha_0)$ so that $\Phi(x, y) = x_{(2)} + r$. Then $D(0, \alpha_1) \cap C(y, r)$ goes over to an interval I_1 on the $x_{(1)}$ axis, $D(0, \alpha_1) \cap C(z, s)$ goes over to the graph of a function $x_{(2)} = \phi(x_{(1)})$ over a (possibly slightly different) interval I_2 , and the part of the graph corresponding to $D(0, \alpha_2)$ will lie over an interval I whose double will be contained in $I_1 \cap I_2$ if α_2 has been chosen small enough. By (20), ϕ will satisfy (22) with $\rho = |y - z|$. Furthermore $\Delta(z, s, y, r)$ is clearly comparable to $\min_{s \in I_1} (|\phi(t)| + |\dot{\phi}(t)| + |t - s|)$, and this in turn is easily seen to be comparable to $\min_{t \in I_1 \cap I_2} (|\phi(t)| + |\dot{\phi}(t)|)$. So (i) of Lemma 3.1 follows by applying (iv) of Lemma 3.3 with $\tilde{I} = I_1 \cap I_2$.

For future reference, we note that the above argument also shows the following: Assume that $(z, s, y, r) \in \mathcal{Q}(\beta_0)$ and that $\Delta(z, s, y, r) < A_0^{-1}|y - z|$, for suitable fixed constants β_0 and A_0 . Then there is a unique point $\xi(z, s, y) \in C(z, s) \cap D(0, \alpha_0)$ such that $\nabla_x \Phi(\xi(z, s, y), y) \wedge \nabla_x \Phi(\xi(z, s, y), z) = 0$. Furthermore we have the estimates

$$(26) \quad |\Phi(\xi(z, s, y), y) - r| \lesssim \Delta$$

$$(27) \quad C_\delta(y, r) \cap C_\delta(z, s) \subset D \left(\xi(z, s, y), C \sqrt{\frac{\Delta(z, s, y, r) + \delta}{|y - z| + \delta}} \right).$$

Namely, the tangency point $\xi(z, s, y)$ corresponds to the point $(\tau, \phi(\tau))$ of Lemma 3.3 under the change of variable, and the estimates follow from (i) and (iv) of Lemma 3.3.

It remains to prove (ii) of Lemma 3.1. This will be done in two steps:

(1) There are at most two curves $C(z, s)$ which are tangent to the $C(z_i, r_i)$ at three different points.

(2) If (y, r) belongs to the set S , then (y, r) must be close to one of the points (z, s) in (1).

Step (1) is accomplished in the next two lemmas.

Lemma 3.4. *Assume that $a_1, a_2, a_3, b_1, b_2, b_3$ are smooth functions on an interval $I_0 \subset \mathbb{R}$, with $n(a_i - b_j) \leq 2$ for all i and j and $n(a_i - a_j) \leq 2$ and $n(b_i - b_j) \leq 2$ for all i and j with $i \neq j$. Assume furthermore that $a_i - b_j$ has a double zero for all i and j . Then there are $i \neq j, k \neq l$, and $\tau \in I_0$ such that $a_i = a_j = b_k = b_l$ to second order at τ .*

Remark added 5/15/97. This type of statement is well known in discrete geometry. Step (V) below is essentially the fact that a word on three letters with no abab subword has length ≤ 5 , which is a trivial case of the Davenport-Schinzel theorem.

Proof. We will argue by contradiction, so will always assume the conclusion fails.

(I) $a_i - b_j$ cannot change sign, and vanishes at just one point.

This is because $n(a_i - b_j) \leq 2$, so that the double zero of $a_i - b_j$ is its only zero.

(II) We can assume that $a_i - a_j$ and $b_i - b_j$ ($i \neq j$) have no double zero.

For suppose that $a_i = a_j$ to second order at $\tau \in I_0$. Then $a_i - a_j$ cannot change sign, and we may suppose that $a_i > a_j$ on $I_0 \setminus \{\tau\}$. Since we are assuming the conclusion fails there is at most one k such that $b_k(\tau) = a_i(\tau) = a_j(\tau)$. Fix $l \neq k$. We can assume by symmetry that $b_l(\tau) > a_i(\tau)$. Then $b_l \geq a_i$ everywhere, hence $b_l > a_j$ everywhere, a contradiction.

(III) We can assume that there do not exist τ , $i \neq j$ and k with $a_i(\tau) = a_j(\tau) = b_k(\tau)$ (and similarly with the roles of the a 's and b 's reversed).

This is because τ would have to be a double zero for $a_i - b_k$ and $a_j - b_k$ but (by (II)) a simple zero for $a_i - a_j$.

Because of (I), there is a definite ordering independent of $t \in I_0$ of the form

$$\epsilon_1 \geq \epsilon_2 \geq \epsilon_3 \geq \epsilon_4 \geq \epsilon_5 \geq \epsilon_6$$

where each symbol ϵ_i is either an a or a b .

For example, one possible ordering would be $a \geq a \geq b \geq b \geq b \geq a$, signifying that after renumbering, $a_1(t)$ and $a_2(t)$ are always $\geq b_1(t), b_2(t), b_3(t)$, which in turn are $\geq a_3(t)$.

(IV) Only the following orderings (if any) can occur:

$$\begin{aligned} \text{(i)} \quad & a \geq a \geq a \geq b \geq b \geq b & \text{(ii)} \quad & b \geq b \geq b \geq a \geq a \geq a \\ \text{(iii)} \quad & a \geq a \geq b \geq b \geq b \geq a & \text{(iv)} \quad & b \geq b \geq a \geq a \geq a \geq b \\ \text{(v)} \quad & a \geq b \geq b \geq b \geq a \geq a & \text{(vi)} \quad & b \geq a \geq a \geq a \geq b \geq b. \end{aligned}$$

If the ordering is not one of these six, then there would be $i \neq j$ and $k \neq l$ such that $a_i \geq b_k \geq a_j \geq b_l$, or else $b_l \geq a_j \geq b_k \geq a_i$. But $a_i(\tau) = b_l(\tau)$ for some τ and then also $a_j(\tau) = a_i(\tau)$ contradicting (III).

It suffices by symmetry to consider orderings (i) and (iii). In either case, we define a b_j interval to be an interval where $b_j = \max(b_1, b_2, b_3)$ and which is maximal with respect to this property. It is clear that I_0 is the union of the b_j intervals ($j \in \{1, 2, 3\}$) and (using (II)) that any two have at most an endpoint in common. A common endpoint of a b_j and b_k interval is clearly

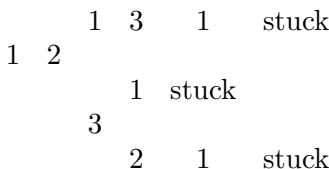
a point where $b_j - b_k$ changes sign. Accordingly there is a diagram of b_j intervals of the form

$$(28) \quad (j_1 \ j_2 \ \dots \ j_n)$$

where each j_i is either a 1, 2 or 3 and $j_i \neq j_{i+1}$, signifying that the leftmost interval is a b_{j_1} interval, the next a b_{j_2} interval, etc., with the rightmost being a b_{j_n} interval.

(V) The number n in (28) is ≤ 5 . If $n \geq 4$ then the pattern (28) is one of the following: (12321), (12131), (1213), (1232) (1231) or may be reduced to one of these by renumbering.

The proof may be given diagrammatically as follows.



Namely, we can always renumber so that the first two intervals are a b_1 interval followed by a b_2 interval. The third is then either a b_1 or a b_3 . If the third is a b_1 then there can never be another b_2 , since $b_1 - b_2$ has already changed sign twice. Thus the fourth must be a b_3 , the fifth a b_1 , and now $b_1 - b_3$ has also changed sign twice so there is no possibility for a sixth interval. Similarly, if the third interval is a b_3 then the fourth is a b_1 or b_2 . In the first case, $b_1 - b_2$ and $b_1 - b_3$ have both changed sign twice so there is no possible fifth interval. In the second case, $b_2 - b_3$ has changed sign twice, so the fifth interval can only be a b_1 , and then $b_1 - b_2$ and $b_1 - b_3$ have also changed sign twice.

(VI) Ordering (i) cannot occur.

Define a_i intervals similarly to b_j intervals: An a_i interval is an interval where $a_i = \min(a_1, a_2, a_3)$ and which is maximal with respect to this property. Of course (V) applies to the a_i intervals as well. Observe that in (V) the patterns with length 5 start and end with the same digit. Accordingly, there are at most nine intervals of the form

$$a_i \text{ interval} \cap b_j \text{ interval}$$

and containing more than one point, and if there are nine such intervals, then the first and last correspond to the same pair (i, j) . So at most eight pairs (i, j) actually occur. On the other hand, a point where $a_i = b_j$ must (by (III)) belong to the interior of an a_i interval and a b_j interval. Since there are nine choices for (i, j) we have a contradiction.

(VII) Ordering (iii) cannot occur. (This will of course complete the proof of the lemma.)

We number so that $a_1 \& a_2 \geq b_1 \& b_2 \& b_3 \geq a_3$. Then we define an a_i interval ($i = 1$ or 2) to be a maximal interval where $a_i = \min(a_1, a_2)$. It is easy to see that the pattern of a_i intervals is (121), (12) or (1) or can be reduced to one of these by interchanging 1 and 2.

There are six pairs (i, j) with $i \in \{1, 2\}$ and $j \in \{1, 2, 3\}$. If there are less than four b_j intervals, or if there are four and the first and last have the same value of j , then we see as in (VI) that for one of these six pairs there is no interval of the form (a_i interval \cap b_j interval) and containing more than one point, which is a contradiction. Thus the pattern of b_j intervals may be taken to be either (21312), (12131), (1213) or (2131)⁴. In any of these cases 1 appears in two positions not the first and last. (Namely, the 2nd and 4th positions in the first case, the 1st and 3rd in the second case, etc.) Now there must be a point τ where $b_1(\tau) = a_3(\tau)$, hence $b_1(\tau) < \min(b_2(\tau), b_3(\tau))$. In moving from τ to a b_1 interval, b_1 must cross both b_2 and b_3 .

Now consider two cases: Either τ separates two b_1 intervals or not. If so, then b_1 crosses both b_2 and b_3 twice between the two b_1 intervals. Since one of the b_1 intervals is neither the first nor the last interval, b_1 must also cross either b_2 or b_3 at its opposite endpoint and we have a contradiction. If τ does not separate two b_1 intervals then b_1 nevertheless crosses either b_2 or b_3 twice in moving between the two b_1 intervals, and crosses both b_2 and b_3 between τ and its closest b_1 interval. This is again a contradiction.

Lemma 3.5. *Suppose Φ satisfies (14) and (15), and α_0, β_0 are small enough. Assume that $(z_i, r_i) \in \mathbb{R}^2 \times \mathbb{R}$ ($i = 1, 2, 3$) are distinct and satisfy $|z_i| + |r_i| < \beta_0$. Then there are at most two values of (z, s) with $|z| + |s| < \beta_0$ and such that $C(z, s)$ is tangent to the three curves $C(z_i, r_i)$ at three different points of $D(0, \alpha_0)$.*

Proof. This follows from Lemma 3.4 and (iii) of Lemma 3.3 using appropriate changes of coordinates. Namely, suppose that $C(y_j, s_j)$, $j = 1, 2, 3$ are each tangent to each of the curves $C(z_i, r_i)$ at three different points as indicated. Since we are working locally we can assume that all the curves in question are graphs. Letting $C(z_i, r_i)$ be the graph of a_i and $C(y_j, s_j)$ the graph of b_j , we have $n(a_i - b_j) \leq 2$, $n(a_i - a_j) \leq 2$, $n(b_i - b_j) \leq 2$, by Lemma 3.3 (iii), and furthermore the tangency assumption implies that $a_i - b_j$ has a double zero. We conclude by Lemma 3.4 that for some i_1, i_2 and j the tangency point between $C(z_{i_1}, r_{i_1})$ and $C(y_j, s_j)$ must coincide with the tangency point between $C(z_{i_2}, r_{i_2})$ and $C(y_j, s_j)$, clearly a contradiction. \square

We now carry out step (2) of the proof of Lemma 3.1(ii). We note to begin with that the points $\xi(z_i, r_i, y, r)$ as well as the definition of $\Delta(z_i, r_i, y, r)$ depend only on the curve family $C(y, r) = \{x : \Phi(x, y) = r\}$ and not on the

⁴It was convenient to renumber some of the cases in (V): (21312) and (2131) are of course equivalent to (12321) and (1232) respectively.

particular choice of defining function Φ . We recall (26): Provided we have appropriate derivative bounds for Φ , there is an estimate

$$(29) \quad |r - \Phi(\xi_i(y), y)| \lesssim \epsilon$$

when $(y, r) \in S$. Here $\xi_i(y) \stackrel{\text{def}}{=} \xi(z_i, r_i, y)$. We will also repeatedly use (27) which guarantees that the set $C_\delta(y_0, r_0)$ lies within distance $\ll \lambda$ from the point $\xi_j(y_0)$.

We let $C_j = C(z_j, r_j)$, and given a curve $C(y_0, r_0)$ we can find numbers μ and ν with $\lambda \leq \nu \leq \mu$, and may order the C_j 's so that $|\xi_2(y_0) - \xi_3(y_0)| = \nu$ and $|\xi_1(y_0) - \xi_2(y_0)| \approx \mu$, $|\xi_1(y_0) - \xi_3(y_0)| \approx \mu$. Let C_* be a suitable large constant which should be chosen before the constant A_1 in Lemma 3.1; we note that (30) below is therefore satisfied if $|a_j - \xi_j(y_0)|$ is bounded by a fixed multiple of $\frac{\epsilon}{t\lambda}$. We will say that Φ is (y_0, r_0) -adapted if there are points $a_j \in C_j$ such that

$$(30) \quad |a_j - \xi_j(y_0)| \leq C_*^{-1}\lambda$$

and

$$(31) \quad \Phi(a_1, y) = 0$$

$$(32) \quad \nabla_y \Phi(a_2, y) = (e \cdot (a_2 - a_1))\beta$$

for all y ; here e is the tangent vector to C_1 at a_1 , and β is a vector independent of y with length ≈ 1 .

Lemma 3.6. *If $(y_0, r_0) \in S$ then a (y_0, r_0) -adapted defining function will exist (satisfying uniform derivative bounds). Furthermore, it can be taken to be (y_1, r_1) -adapted for all (y_1, r_1) such that $|\xi_j(y_1) - \xi_j(y_0)| \leq (2C_*)^{-1}\lambda$ for all j .*

Proof. It is easy to see that $|e \cdot (a_2 - a_1)| \approx |a_2 - a_1|$, and we will prove this below (see (36)). We take it as known for now.

Starting from any defining function Φ , we can achieve (31) by replacing Φ by $\Phi(\cdot, y) - \Phi(a_1, y)$. Assuming this done, we can then achieve (32) by choosing appropriate curvilinear coordinates in the y -plane (the “straightening out theorem” in the terminology of [6], applied to the function $\Phi(a_2, \cdot)$), since the hypothesis (15) guarantees that $|\nabla_y \Phi(a_2, y)| \approx |a_2 - a_1| \approx |e \cdot (a_2 - a_1)|$. The last statement follows from the definition of adapted provided we use points with $|a_j - \xi_j(y_0)| \leq (2C_*)^{-1}\lambda$ in the preceding construction. \square

We fix $(y_0, r_0) \in S$ and choose a (y_0, r_0) -adapted defining function Φ , and we let

$$T(y) = \begin{pmatrix} \nabla_y \Phi(\xi_1(y), y) & -1 \\ \nabla_y \Phi(\xi_2(y), y) & -1 \\ \nabla_y \Phi(\xi_3(y), y) & -1 \end{pmatrix}.$$

We record the formula for $T(y)^{-1}$ for use below. If $v = (v_1, v_2) \in \mathbb{R}^2$, then let $v^* = (-v_2, v_1)$. Then, provided $\det T(y) \neq 0$, $T(y)^{-1}$ is equal to

$$(33) \quad (\det T(y))^{-1} \begin{pmatrix} (\nabla_y \Phi(\xi_2(y), y) - \nabla_y \Phi(\xi_3(y), y))^* & \nabla_y \Phi(\xi_2(y), y) \wedge \nabla_y \Phi(\xi_3(y), y) \\ (\nabla_y \Phi(\xi_3(y), y) - \nabla_y \Phi(\xi_1(y), y))^* & \nabla_y \Phi(\xi_3(y), y) \wedge \nabla_y \Phi(\xi_1(y), y) \\ (\nabla_y \Phi(\xi_1(y), y) - \nabla_y \Phi(\xi_2(y), y))^* & \nabla_y \Phi(\xi_1(y), y) \wedge \nabla_y \Phi(\xi_2(y), y) \end{pmatrix}^t$$

where T^t means transpose of T . We also define $s_j = \Phi(\xi_j(y_0), y_0)$, and note that $|s_j - r_0| \lesssim \epsilon$ by (29).

Lemma 3.7. $|\det T(y_0)| \approx \mu^2 \nu$ and $\|T(y_0)^{-1}\| \lesssim (\mu \nu)^{-1}$.

Proof. If $y = y_0$ then the entries of the cofactor matrix in (33) are clearly $\lesssim \mu$ and it therefore suffices to show that $|\det T(y_0)| \approx \mu^2 \nu$. However, we have

$$|\Phi(\xi_i(y_0), y_0) - \Phi(\xi_j(y_0), y_0)| \lesssim \epsilon$$

so by (14) there are points $x_1, x_2, x_3 \in \mathbb{R}^2$ with $|x_i - \xi_i(y_0)| \lesssim \epsilon$ and $\Phi(x_1, y_0) = \Phi(x_2, y_0) = \Phi(x_3, y_0)$. Note that then also

$$(34) \quad |x_i - x_j| \approx |\xi_i(y_0) - \xi_j(y_0)|$$

since ϵ is small compared with λ . Define

$$M = \begin{pmatrix} \nabla_y \Phi(x_1, y_0) & -1 \\ \nabla_y \Phi(x_2, y_0) & -1 \\ \nabla_y \Phi(x_3, y_0) & -1 \end{pmatrix}.$$

Then Lemma 3.2 implies that

$$\begin{aligned} |\det M| &\approx |x_1 - x_2| |x_2 - x_3| |x_3 - x_1| \\ &\approx \mu^2 \nu \end{aligned}$$

where the last line follows from (34). On the other hand, $|\det(T(y_0)) - \det(M)| \lesssim \epsilon(\mu + \epsilon)$ since $|\nabla_y \Phi(\xi_i(y_0), y_0) - \nabla_y \Phi(x_i, y_0)| \lesssim \epsilon$ and the cofactors of $T(y_0)$ are $\lesssim \mu$. We know that $\epsilon \ll \lambda^2$, so $\epsilon(\mu + \epsilon)$ is small compared with $\mu^2 \nu$. This completes the proof. \square

Lemma 3.8. *If y is such that $|T(y_0)(y - y_0, r - r_0)| < \epsilon$ for some r then*

$$T(y)T(y_0)^{-1} = I + E(y)$$

where I is the identity matrix and $E(y)$ is a matrix with norm $\leq \frac{1}{100}$.

Proof. We can assume that a_1 is the origin (thus by (30), all the points $\xi_j(y_0)$ are at distance $\lesssim \mu$ from the origin) and furthermore that the unit tangent vector to C_1 at the origin is parallel to the $x^{(1)}$ axis.

Let $e_j(x)$ be the unit tangent vector to C_j at the point $x \in C_j$. We claim first that if $x \in C_j$ for some j and $|x| \leq \mu$, then

$$(35) \quad |x^{(2)}| \lesssim \mu^2$$

$$(36) \quad |e_j(x)^{(2)}| \lesssim \mu.$$

We will repeatedly use the following simple fact: If C is a smooth curve and there is one point $x \in C$ with $|x| \lesssim \mu$ such that $|x^{(2)}| \lesssim \mu^2$ and the $x^{(2)}$ component of the unit tangent vector to C at x is $\lesssim \mu$, then the same properties hold at every point of C which is at distance $\lesssim \mu$ from the origin.

In particular, (35) and (36) are obvious if $j = 1$, since C_1 passes through the origin and has a horizontal tangent vector there. If $j = 2$ or 3 , then set $x = \xi_j(y_0) \in C_j \cap C(y_0, s_j)$. We have $|s_j - s_1| \lesssim \epsilon$, so by (14) we can find a point $\tilde{x} \in C(y_0, s_1)$ which is at distance $\lesssim \epsilon$ from $\xi_j(r_0)$. Let \tilde{e} be the unit tangent vector to $C(y_0, s_1)$ at \tilde{x} . Clearly then $|\tilde{x}| \lesssim \mu$. $C(y_0, s_1)$ is tangent to C_1 at $\xi_1(y_0)$, which is within μ of the origin, and it follows that $|\tilde{x}^{(2)}| \lesssim \mu^2$ and $|\tilde{e}^{(2)}| \lesssim \mu$. Hence $|x^{(2)}| \lesssim \mu^2 + \epsilon \lesssim \mu^2$. On the other hand $e_j(x)$ coincides with the unit tangent vector to $C(y_0, s_j)$ at x and therefore $|e_j(x) - \tilde{e}| \lesssim \epsilon$, so $|e_j(x)^{(2)}| \lesssim \mu$. This proves (35) and (36). We note that if $x, \bar{x} \in C_j$ and $|x| + |\bar{x}| \lesssim \mu$, then (36) implies

$$(37) \quad |x^{(2)} - \bar{x}^{(2)}| \lesssim \mu|x - \bar{x}|.$$

Next, we claim that if $|x| \lesssim \mu$ and $|x^{(2)}| \lesssim \mu^2$ then

$$(38) \quad \frac{d}{dx^{(1)}} \nabla_y \Phi(x, y) = \beta + \mathcal{O}(\mu).$$

We start by proving (38) when $x = 0$. We expand $\nabla_y \Phi(\cdot, y)$ in a Taylor series with remainder:

$$(39) \quad \begin{aligned} \nabla_y \Phi(x, y) &= \nabla_y \Phi(0, y) + x^{(1)} \frac{\partial \nabla_y \Phi}{\partial x^{(1)}}(0, y) + \mathcal{O}(x^{(2)} + (x^{(1)})^2) \\ &= x^{(1)} \frac{\partial \nabla_y \Phi}{\partial x^{(1)}}(0, y) + \mathcal{O}(x^{(2)} + (x^{(1)})^2) \end{aligned}$$

using (31). Now we set $x = a_2$ and use (32):

$$\begin{aligned} a_2^{(1)} \beta &= \nabla_y \Phi(a_2, y) = a_2^{(1)} \frac{\partial \nabla_y \Phi}{\partial x^{(1)}}(0, y) + a_2^{(2)} \frac{\partial \nabla_y \Phi}{\partial x^{(2)}}(0, y) + \mathcal{O}(|a_2|^2) \\ &= a_2^{(1)} \frac{\partial \Phi}{\partial x^{(1)}}(0, y) + \mathcal{O}(\mu^2) \end{aligned}$$

by (35). This proves (38) when $x = 0$. It follows that (38) holds for any x with $|x| \leq \mu$.

An immediate consequence of (38) is that if $|x| + |\bar{x}| \lesssim \mu$ and $|x^{(2)}| + |\bar{x}^{(2)}| \lesssim \mu^2$ then

$$(40) \quad \nabla_y \Phi(x, y) - \nabla_y \Phi(\bar{x}, y) = (x - \bar{x})^1 \beta + \mathcal{O}(\mu^2).$$

We now prove two more bounds of the same type. First we claim that if $x, \bar{x} \in C_j$ and $|x| + |\bar{x}| \lesssim \mu$, then

$$(41) \quad \nabla_y \Phi(x, y) - \nabla_y \Phi(\bar{x}, y) = (x - \bar{x})^{(1)} \beta + \mathcal{O}(\mu|x - \bar{x}|).$$

This follows from the preceding estimates and the mean value theorem: Let γ be the line segment connecting x and \bar{x} , then

$$\begin{aligned} & \left| (\nabla_y \Phi(x, y) - x^{(1)}\beta) - (\nabla_y \Phi(\bar{x}, y) - \bar{x}^{(1)}\beta) \right| \\ & \lesssim |x^{(1)} - \bar{x}^{(1)}| \max_{u \in \gamma} \left| \frac{d}{dx^{(1)}} \nabla_y \Phi(u, y) - \beta \right| + |x^{(2)} - \bar{x}^{(2)}| \max_{u \in \gamma} \left| \frac{d}{dx^{(2)}} \nabla_y \Phi(u, y) \right| \\ & \lesssim |x - \bar{x}| \cdot \mu + \mu |x - \bar{x}| \cdot 1 \end{aligned}$$

by (38) and (37). This proves (41).

Next, we claim that if $|x| \lesssim \mu$ and $|x^{(2)}| \lesssim \mu$, then

$$(42) \quad |\nabla_y^2 \Phi(x, y)| \lesssim \mu^2.$$

Namely, let e be the unit vector in the direction from 0 to a_2 . Fix i and j and let $f(x) = \nabla_y^2 \Phi(x, y)$. By (31) and (32), f is a smooth function vanishing at 0 and a_2 and therefore $|e \cdot \nabla f(0)| \lesssim \mu$. The $x^{(2)}$ component of e is $\lesssim \mu$ by (35) so we may conclude that $\left| \frac{\partial f}{\partial x^{(1)}}(0) \right| \lesssim \mu$, and therefore $\left| \frac{\partial f}{\partial x^{(1)}} \right| \lesssim \mu$ at all points within μ of the origin. Estimate (42) now follows from the mean value theorem and: $|f(x)| = |f(x) - f(0)| \lesssim \mu |x^{(1)} - a_i^{(1)}| + |x^{(2)}| \lesssim \mu^2$.

Now suppose that $(y, r) \in \mathbb{R}^2 \times \mathbb{R}$ is such that $|T(y_0)(y - y_0, r - r_0)| < \epsilon$. Then we claim that

$$(43) \quad |y - y_0| \lesssim \frac{\epsilon}{\mu\nu}$$

$$(44) \quad |\beta \cdot (y - y_0)| \lesssim \frac{\epsilon}{\nu}$$

$$(45) \quad |\xi_j(y) - \xi_j(y_0)| \lesssim \frac{\epsilon}{t\nu}.$$

Namely, (43) follows immediately from Lemma 3.7. Furthermore, by (40) we have $\nabla_y \Phi(\xi_2(y_0), y_0) - \nabla_y \Phi(\xi_1(y_0), y_0) = (\xi_2(y_0) - \xi_1(y_0))^{(1)}\beta + \mathcal{O}(\mu^2)$, with $|(\xi_2(y_0) - \xi_1(y_0))^{(1)}| \approx \mu$. The quantity $(\nabla_y \Phi(\xi_2(y_0), y_0) - \nabla_y \Phi(\xi_1(y_0), y_0)) \cdot (y - y_0)$ is the difference between the second and first components of $T(y_0)(y - y_0, r - r_0)$ and is therefore $\leq 2\epsilon$, so $|\beta \cdot (y - y_0)| \lesssim \frac{\epsilon}{\mu} + \mu|y - y_0| \lesssim \frac{\epsilon}{\nu}$ (by (43)), which is (44).

In order to prove (45) we consider the variation $y(s) = y_0 + s(y - y_0)$, $0 \leq s \leq 1$. We will denote the s -derivative of the function f by \dot{f} . Recall that $e_j(x), x \in C_j$, is the unit tangent vector field to C_j . Assume temporarily that $|\xi_j(y(s)) - \xi_j(y(0))| \leq \mu$ for all s . Consider the equation

$$e_j(\xi_j(s)) \cdot \nabla_x \Phi(\xi_j(y(s)), y(s))$$

which defines $\xi_j(s)$ implicitly. Differentiate it with respect to s and take the component of the resulting equation in the direction $e_j(\xi_j(s))$. This gives

$$\dot{\xi}_j = \frac{(e_j \cdot \nabla_x) \nabla_y \Phi(\xi_j(y(s)), y(s)) \cdot (y - y_0)}{(e_j \cdot \nabla_x)^2 \Phi(\xi_j(y(s)), y(s))}.$$

The expression in the denominator is $\gtrsim |y(s) - z_j|$ by the curvature hypothesis (15). Since $|y(s) - y_0| \leq \frac{\epsilon}{\mu\nu} \ll t$, we conclude that $|(e \cdot \nabla_x)^2 \Phi(\xi, y)| \gtrsim t$. On the other hand, the $x^{(2)}$ -component of e_j is $\lesssim \mu$ and therefore

$$\begin{aligned} & (e_j \cdot \nabla_x) \nabla_y \Phi(\xi_j(y(s)), y(s)) \cdot (y - y_0) \\ &= e^{(1)} \frac{d}{dx^{(1)}} \nabla_y \Phi(\xi_j(y(s)), y(s)) \cdot (y - y_0) + \mathcal{O}(\mu|y - y_0|) \\ &= e^{(1)} \beta \cdot (y - y_0) + \mathcal{O}(\mu|y - y_0|) \quad \text{by (38)} \\ &\lesssim \frac{\epsilon}{\nu} \quad \text{by (43) and (44)}. \end{aligned}$$

We conclude that $|\dot{\xi}_j| \lesssim \frac{\epsilon}{t\nu}$. Thus we have shown that under the a priori assumption on v that $|\xi_j(s) - \xi_j(0)| \leq \mu$ for all s we actually have $|\xi_j(s) - \xi_j(0)| \leq C \frac{\epsilon}{t\nu}$ for all s . Since $\frac{\epsilon}{t\nu}$ is small compared with μ it follows that the a priori assumption can be dropped, i.e., (45) holds.

Next we claim that if $|T(y_0)(y - y_0, r - r_0)| < \epsilon$ then

$$(46) \quad |\nabla_y \Phi(\xi_j(y), y) - \nabla_y \Phi(\xi_j(y_0), y_0)| \lesssim \frac{\epsilon}{t\nu}$$

$$(47) \quad |(\nabla_y \Phi(\xi_j(y), y) - \nabla_y \Phi(\xi_j(y_0), y_0)) \wedge \beta| \lesssim \mu \frac{\epsilon}{t\nu}.$$

For the proof we write

$$\begin{aligned} & \nabla_y \Phi(\xi_j(y), y) - \nabla_y \Phi(\xi_j(y_0), y_0) \\ &= (\nabla_y \Phi(\xi_j(y), y) - \nabla_y \Phi(\xi_j(y_0), y)) + (\nabla_y \Phi(\xi_j(y_0), y) - \nabla_y \Phi(\xi_j(y_0), y_0)) \\ &\stackrel{\text{def}}{=} \text{I} + \text{II} \end{aligned}$$

and make the following estimates:

$$|\text{II}| \lesssim \mu^2 \cdot \frac{\epsilon}{\mu\nu} = \mu \frac{\epsilon}{\nu} \leq \mu \frac{\epsilon}{t\nu}$$

by (42) and (43). Also

$$|\text{I}| \lesssim \frac{\epsilon}{t\nu}$$

by (45), and finally

$$|\text{I} \wedge \beta| \lesssim \mu \frac{\epsilon}{t\nu}$$

by (45) and (41). This proves (46) and (47).

By (40), which is applicable by (45) and (35), we have

$$\nabla_y \Phi(\xi_j(y_0), y_0) - \nabla_y \Phi(\xi_k(y_0), y_0) = (\xi_j(y_0) - \xi_k(y_0))^{(1)} \beta + \mathcal{O}(\mu^2),$$

with $|\xi_j(y_0) - \xi_k(y_0)| \lesssim \mu$. It follows using (46) and (47) that

(48)

$$|(\nabla_y \Phi(\xi_i(y), y) - \nabla_y \Phi(\xi_i(y_0), y_0)) \wedge (\nabla_y \Phi(\xi_j(y_0), y_0) - \nabla_y \Phi(\xi_k(y_0), y_0))| \lesssim \mu^2 \frac{\epsilon}{t\nu}$$

for all i, j and k .

We now prove Lemma 3.8. Namely, if $v = (v_1, v_2) \in \mathbb{R}^2$ then, using (33), each component of $(T(y) - T(y_0))T(y_0)^{-1}$ is of the form

$$(\det T(y_0))^{-1}(\nabla_y \Phi(\xi_i(y), y) - \nabla_y \Phi(\xi_i(y_0), y_0)) \wedge (\nabla_y \Phi(\xi_j(y_0), y_0) - \nabla_y \Phi(\xi_k(y_0), y_0))$$

for some i, j, k . We have $|\det T(y_0)| \approx \mu^2\nu$ by Lemma 3.7 and we therefore conclude by (48) that $\|(T(y) - T(0))T(0)^{-1}\| \lesssim \frac{\epsilon}{t\nu^2}$, which is small. The lemma follows. \square

We now define

$$G(y, r) = \begin{pmatrix} \Phi(\xi_1(y), y) - r \\ \Phi(\xi_2(y), y) - r \\ \Phi(\xi_3(y), y) - r \end{pmatrix}.$$

We let C_0 be a suitable constant and $E(y, r) = (y, r) + DG(y, r)^{-1}D(0, C_0\epsilon)$, i.e., $E(y, r)$ is the ellipsoid centered at (y, r) which the derivative of G at (y, r) maps onto the disc $D(C_0, \epsilon)$.

- Lemma 3.9.** (i) $E(y_0, r_0)$ has volume $\approx \frac{\epsilon^3}{\mu^2\nu}$ and diameter $\lesssim \frac{\epsilon}{\mu\nu}$
 (ii) G defines a diffeomorphism from a subset of $E(y_0, r_0)$ onto $D(0, \epsilon)$.
 (iii) If $(y, r) \in E(y_0, r_0)$ then $E(y_0, r_0)$ and $E(y, r)$ are comparable ellipsoids, i.e., each is contained in the dilation of the other around its center by a factor C_1 .

Proof. Observe that $DG(y, r) = T(y)$, since $\nabla_x \Phi(\xi_i(y), y)$ is parallel to $\nabla_x \Phi(\xi_i(y), z_i)$ and therefore perpendicular to any column of $d\xi_i(y)$.

Part (i) is immediate from Lemma 3.7. For (iii), we use that matrices A and B such that $\|A^{-1}B - I\| \leq \frac{1}{2}$ map any fixed disc centered at the origin to comparable ellipsoids. If we set $A = T(y)^{-1}$, $B = T(y_0)^{-1}$ then $\|A^{-1}B - I\| \leq \frac{1}{2}$ by Lemma 3.8 and (iii) follows.

To prove (ii), we use the following version of the inverse function theorem: If $\Omega \subset \mathbb{R}^3$ is open, $f : \Omega \rightarrow \mathbb{R}^3$, $f(a) = b$ and $|Df(x) - I| \leq \frac{1}{100}$ when $x \in D(0, C_2\epsilon)$, then f maps some subset of $D(a, C_2\epsilon)$ diffeomorphically onto $D(b, \epsilon)$. This follows from the usual proof of the IFT: We can assume $a = b = 0$. Then with the stated hypotheses, if $y \in D(0, \epsilon)$ then the map $x \rightarrow x - f(x) + y$ will be a contraction mapping from $D(0, C_2\epsilon)$ into itself, etc.

We apply this version of the IFT to the map $G \circ DG(y_0, r_0)^{-1}$, which satisfies the hypothesis in view of Lemma 3.8 (ϵ in Lemma 3.8 can clearly be replaced by $C_2\epsilon$). The result follows. \square

We may now finish the proof of Lemma 3.1 as follows. Suppose that $(y_0, r_0) \in S$ is given and choose an adapted defining function. Then by (ii) of Lemma 3.9 one of the two points (η, ρ) must belong to $E(y_0, r_0)$. Also $|\xi_j(y_0) - \xi_j(\eta)|$ is small compared with λ by (45). It follows therefore that there is a defining function Φ which is (y_0, r_0) - adapted for all (y_0, r_0) with a given (η, ρ) . Using this Φ , it follows by (iii) of Lemma 3.9 that each point (y_0, r_0) must belong to a fixed dilation of $E(\eta, \rho)$, and the proof is complete. The final statement (19) follows since the projection of each ellipsoid $E(\eta, \rho)$ on any given axis will be an interval of length $\lesssim \frac{\epsilon}{\lambda^2}$.

4. Proof of Theorem 2', part 2.

We will prove the following general version of Theorem 2'.

Theorem 4.1. *Let $\Phi : U \rightarrow \mathbb{R}$ be a function satisfying hypotheses (14), (15), with $\Phi(a, b) = r_0$, and let ϵ_0 be a small constant. If $\delta > 0$ is small and $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, $\text{supp } f \subset D(a, \epsilon_0)$ then define $M_\delta f : (r_0 - \epsilon_0, r_0 + \epsilon_0) \rightarrow \mathbb{R}$ via*

$$M_\delta f(r) = \sup_{y \in D(b, \epsilon_0)} \delta^{-1} \int_{C_\delta(y, r)} |f|.$$

Then

$$(49) \quad \|M_\delta f\|_q \lesssim \delta^{-\frac{1}{2}(\frac{3}{p}-1)} \|f\|_p, \quad p < \frac{8}{3}, q = 2p'.$$

The $p = 1, q = \infty$ case of Theorem 4.1 is trivial, and it therefore suffices to prove the following restricted weak type bound:

$$(50) \quad \left| \left\{ r \in \left[\frac{1}{2}, 2 \right] : M_\delta \chi_E(r) > \lambda \right\} \right| \leq C \left(\frac{|E|}{\delta^{\frac{1}{6}} \lambda^{\frac{8}{3}}} \right)^{\frac{6}{5}}.$$

Here the set E is contained in $D(0, \epsilon_0) \subset \mathbb{R}^2$. As usual it is convenient to discretize; since we can estimate the measure of a set in terms of its δ -entropy it suffices to prove the following.

Assume there are M 3δ -separated values $r_j \in [r_0 - \epsilon_0, r_0 + \epsilon_0]$ and points $x_j \in D(0, \epsilon_0)$ such that $|E \cap C_\delta(x_j, r_j)| \geq \lambda |C_\delta(x_j, r_j)|$. Then

$$(51) \quad M_\delta \leq C \left(\frac{|E|}{\delta^{\frac{1}{6}} \lambda^{\frac{8}{3}}} \right)^{\frac{6}{5}}.$$

We can assume that M is large; for M smaller than any fixed constant (51) holds because $M \neq 0$ implies $|E| \gtrsim \lambda \delta$. We may also assume that $\lambda \leq 1$.

To prove (51) we let μ (“multiplicity”) be the smallest number with the following property: There are at least $\frac{M}{2}$ values of j such that

$$(52) \quad |E \cap C_\delta(x_j, r_j) \cap \{x : \text{card}(\{i : x \in C_\delta(x_i, r_i)\}) \leq \mu\}| \geq \frac{\lambda}{2} |C_\delta(x_j, r_j)|.$$

The main estimate is

$$(53) \quad \mu \lesssim M^{\frac{1}{6}} \lambda^{-\frac{5}{3}}.$$

Before proving (53) we introduce some more notation, as follows. For any $t \in [\delta, 1]$ and $\epsilon \in [\delta, 1]$, let

$$a(t, \epsilon) = C_1^{-1} \left(\frac{\delta}{\epsilon} \right)^\alpha \left(\frac{M\delta}{t} + \frac{t}{M\delta} \right)^{-\alpha}.$$

Here α is a sufficiently small positive constant, and C_1 is a positive constant (easily shown to exist) which is large enough that

$$(54) \quad \sum_{\substack{k \geq 0 \\ l \geq 0}} a(2^k \delta, 2^l \delta) < 1$$

for all M and δ . Also let

$$\begin{aligned} \bar{\lambda}(t, \epsilon) &= a(t, \epsilon) \frac{\lambda}{2} \\ \bar{\mu}(t, \epsilon) &= a(t, \epsilon) \mu \\ \bar{M}(t, \epsilon) &= a(t, \epsilon) \frac{M}{2}. \end{aligned}$$

Let $\Delta(z, s, y, r)$ be the function of Lemma 3.1, and for $i, j \in \{1, \dots, M\}$ define a number Δ_{ij} as follows:

$$(55) \quad \Delta_{ij} = \max(\delta, \Delta(x_i, r_i, x_j, r_j)).$$

For each $j \in \{1, \dots, M\}$, $t \in [\delta, 1]$, $\epsilon \in [\delta, 1]$, let

$$S_{t, \epsilon}(x_j, r_j) \stackrel{\text{def}}{=} \{i : C_\delta(x_j, r_j) \cap C_\delta(x_i, r_i) \neq \emptyset, t \leq |r_i - r_j| + |x_i - x_j| \leq 2t \\ \text{and } \epsilon \leq \Delta_{ij} \leq 2\epsilon\}$$

$$A_{t, \epsilon}(x_j, r_j) \stackrel{\text{def}}{=} \{x \in C_\delta(x_j, r_j) : \text{card}(\{i \in S_{t, \epsilon}(x_j, r_j) : x \in C_\delta(x_i, r_i)\}) \\ \geq \bar{\mu}(t, \epsilon)\}.$$

Lemma 4.1. *There are numbers $t \in [\delta, 1]$ and $\epsilon \in [\delta, 1]$ with the following property:*

There are $\geq \bar{M}(t, \epsilon)$ values of j such that $|A_{t\epsilon}(x_j, r_j)| \geq \bar{\lambda}(t, \epsilon) |C_\delta(x_j, r_j)|$.

Proof. This is a routine pigeonhole argument. By the minimality of μ there are at least $\frac{M}{2}$ values of j such that $|\tilde{E}_j| \geq \frac{\lambda}{2}|C_\delta(x_j, r_j)|$ where

$$\tilde{E}_j = E \cap C_\delta(x_j, r_j) \cap \{x : \text{card}(\{i : x \in C_\delta(x_i, r_i)\}) \geq \mu\}.$$

For any such j and any $x \in \tilde{E}_j$, (54) implies there are $t = 2^k\delta$ and $\epsilon = 2^l\delta$ such that $x \in A_{t\epsilon}(x_j, r_j)$. Consequently, using (54) again, for any such j there are $t = 2^k\delta$ and $\epsilon = 2^l\delta$ such that

$$(56) \quad |A_{t\epsilon}(x_j, r_j)| \geq \bar{\lambda}(t, \epsilon)|C_\delta(x_j, r_j)|.$$

By (54) once more, there must be a choice of t and ϵ such that (56) holds for at least $\bar{M}(t, \epsilon)$ values of j . This finishes the proof. \square

We also want to note the following simple fact which will be used in a crucial way below.

Lemma 4.2. *Let ϕ be a smooth function on a disc $D(0, r) \subset \mathbb{R}^2$, with $\nabla\phi \neq 0$ and $\phi(0) = 0$. Let $\Gamma^\delta = \{x \in D(0, \frac{r}{2}) : |\phi(x)| < \delta\}$. Suppose $F \subset \Gamma^\delta$. Then there are three subsets $F_1, F_2, F_3 \subset F$ such that (i) $|F_i| \geq C^{-1}|F|$, $i = 1, 2, 3$ and (ii) $\text{dist}(F_i, F_j) \geq (C\delta)^{-1}|F|$ if $i \neq j$.*

The constant C of course depends only on bounds for derivatives of ϕ , a lower bound for $|\nabla\phi|$, and r .

Proof of (53). We may assume by a partition of unity and change of variable that $\phi(x) = x_{(2)}$ and $r = 1$.

We partition $(-\frac{1}{2}, \frac{1}{2})$ into five subintervals I_1, \dots, I_5 (numbered from left to right) in such a way that $|\{x \in F : x_{(1)} \in I_k\}|$ is independent of k , and is therefore equal to $\frac{|F|}{5}$. Set $F_j = \{x \in F : x_{(1)} \in I_{2j-1}\}$. It is clear that property (i) then holds. To prove (ii), note that for $k = 2$ or 4 , $|I_k| \geq \frac{|\{x \in F : x_{(1)} \in I_j\}|}{2\delta} = \frac{|F|}{10\delta}$. Consequently $\text{dist}(F_i, F_j) \geq \frac{|F|}{10\delta}$, as claimed.

We split the proof of (53) into two cases.

(i) $\bar{\lambda} \geq C_2\sqrt{\frac{\epsilon}{t}}$

(ii) $\bar{\lambda} \leq C_2\sqrt{\frac{\epsilon}{t}}$

where C_2 is a sufficiently large constant.

In case (i), which is the main case, we let S be the set of M “tubes” in (51), and let \bar{S} be the set of at least \bar{M} tubes in Lemma 4.1. Let Q be the set of all quadruples (j, j_1, j_2, j_3) with $C(x_j, r_j) \in \bar{S}$, $C(x_{j_i}, r_{j_i}) \in S$ for $i = 1, 2, 3$ and such that $j_i \in S_{t,\epsilon}(x_j, r_j)$ for each $i \in \{1, 2, 3\}$ and furthermore the distance between any two of the three sets $C_\delta(x_j, r_j) \cap C_\delta(x_{j_i}, r_{j_i})$ is at least $C_3^{-1}\bar{\lambda}$. Here C_3 is a suitable constant which should be chosen before C_2 .

We will make two different estimates on the cardinality of Q . On the one hand, if C_2 is large enough then by (19)⁵, for any fixed j_1, j_2, j_3 there are $\lesssim \frac{\epsilon}{\delta} \bar{\lambda}^{-2}$ values of j such that $(j, j_1, j_2, j_3) \in Q$. Also it follows from the definition of Q that there are $\lesssim M \min(M, \frac{t}{\delta})^2$ possible choices for (j_1, j_2, j_3) : There are at most M choices for j_1 , and once j_1 is fixed there are $\lesssim \min(M, \frac{t}{\delta})$ possibilities for each of j_2 and j_3 , since $|r_{j_1} - r_{j_i}| \leq |r_{j_1} - r_j| + |r_j - r_{j_i}| \leq 4t$ for $i = 2$ or 3 . We conclude that

$$(57) \quad \text{card } Q \lesssim \frac{\epsilon}{\delta} \bar{\lambda}^{-2} M \min \left(M, \frac{t}{\delta} \right)^2.$$

On the other hand, if we fix j with $C(x_j, r_j) \in \bar{S}$ then (provided C_3 has been chosen large enough) Lemma 4.2 implies there are three subsets F_1, F_2, F_3 of the set $A_{t, \epsilon}(x_j, r_j)$ such that $\text{dist}(F_l, F_m) \geq 2C_3^{-1} \bar{\lambda}$, $l \neq m$, and $|F_l| \gtrsim \delta \bar{\lambda}$ for each l . For fixed l , we let S_l be those indices $i \in S_{t, \epsilon}(x_j, r_j)$ such that $F_l \cap C_\delta(x_i, r_i) \neq \emptyset$. The sets $C_\delta(x_i, r_i)$, $i \in S_l$ must cover F_l at least $\bar{\mu}$ times. So

$$\sum_{i \in S_l} |F_l \cap C_\delta(x_i, r_i)| \gtrsim \bar{\mu} \bar{\lambda} \delta.$$

Also, for each fixed i we have $|F_l \cap C_\delta(x_i, r_i)| \lesssim \frac{\delta^2}{\sqrt{t\epsilon}}$ by (16). Consequently

$$(58) \quad \text{card } S_l \gtrsim \delta^{-1} \bar{\mu} \bar{\lambda} \sqrt{t\epsilon}.$$

Estimate (17) implies that if $i \in S_{t, \epsilon}(x_j, r_j)$ then the diameter of $C_\delta(x_i, r_i) \cap C_\delta(x_j, r_j)$ is $\lesssim \sqrt{\frac{\epsilon}{t}}$, which is small compared with $\bar{\lambda}$ if C_2 has been chosen large enough. So if $l \neq m$, $i \in S_l, k \in S_m$, then the distance between $C_\delta(x_i, r_i) \cap C_\delta(x_j, r_j)$ and $C_\delta(x_k, r_k) \cap C_\delta(x_j, r_j)$ is $\geq C_3^{-1} \bar{\lambda}$. We conclude that if $i_l \in S_l$ for $l = 1, 2, 3$ then $(j, i_1, i_2, i_3) \in Q$. So

$$\text{card } Q \gtrsim \bar{M} \left(\delta^{-1} \bar{\mu} \bar{\lambda} \sqrt{t\epsilon} \right)^3.$$

If we compare this equation with (57) we obtain

$$\bar{\mu}^3 \lesssim \frac{\delta^2}{t^{\frac{3}{2}} \epsilon^{\frac{1}{2}}} \bar{\lambda}^{-5} \min \left(M, \frac{t}{\delta} \right)^2 \frac{M}{\bar{M}},$$

or equivalently

$$\mu^3 \lesssim a(t, \epsilon)^{-9} \frac{\delta^2}{t^{\frac{3}{2}} \epsilon^{\frac{1}{2}}} \bar{\lambda}^{-5} \min \left(M, \frac{t}{\delta} \right)^2.$$

⁵The hypotheses on ϵ, t and λ in Lemma 3.1(ii) are satisfied provided C_2 is large enough.

This can be rewritten as

$$\mu^3 \lesssim M^{\frac{1}{2}} \lambda^{-5} \cdot \begin{cases} a(t, \epsilon)^{-9} \left(\frac{\delta}{\epsilon}\right)^{\frac{1}{2}} \left(\frac{t}{\delta M}\right)^{\frac{1}{2}} & \text{if } M \geq \frac{t}{\delta} \\ a(t, \epsilon)^{-9} \left(\frac{\delta}{\epsilon}\right)^{\frac{1}{2}} \left(\frac{M\delta}{t}\right)^{\frac{3}{2}} & \text{if } M \leq \frac{t}{\delta}. \end{cases}$$

The expression inside the brace is bounded by a constant by the definition of $a(t, \epsilon)$, provided the constant α is less than $\frac{1}{18}$. So we have proved (53) in case (i).

In case (ii), we fix j with $C(x_j, r_j) \in \bar{S}$ and make the trivial estimate card $S_{t\epsilon}(x_j, r_j) \lesssim \min(M, \frac{t}{\delta})$. It follows that

$$\begin{aligned} \bar{\mu} \bar{\lambda} \delta &\lesssim \sum_{i \in S_{t\epsilon}(x_j, r_j)} |C_\delta(x_j, r_j) \cap C_\delta(x_i, r_i)| \\ &\lesssim \min\left(M, \frac{t}{\delta}\right) \frac{\delta^2}{\sqrt{t\epsilon}}, \end{aligned}$$

where we used (16). Thus $\bar{\mu} \lesssim \bar{\lambda}^{-1} \sqrt{\frac{t}{\epsilon}} \min(\frac{M\delta}{t}, 1)$. Using the hypothesis (ii) we therefore have

$$\bar{\mu} \lesssim \bar{\lambda}^{-\frac{5}{3}} \left(\frac{t}{\epsilon}\right)^{\frac{1}{6}} \min\left(\frac{M\delta}{t}, 1\right)$$

i.e.,

$$\begin{aligned} \mu &\lesssim a(t, \epsilon)^{-\frac{8}{3}} \lambda^{-\frac{5}{3}} \left(\frac{t}{\epsilon}\right)^{\frac{1}{6}} \min\left(\frac{M\delta}{t}, 1\right) \\ &= \lambda^{-\frac{5}{3}} M^{\frac{1}{6}} \cdot \begin{cases} a(t, \epsilon)^{-\frac{8}{3}} \left(\frac{\delta}{\epsilon}\right)^{\frac{1}{6}} \left(\frac{t}{\delta M}\right)^{\frac{1}{6}} & \text{if } M \geq \frac{t}{\delta} \\ a(t, \epsilon)^{-\frac{8}{3}} \left(\frac{\delta}{\epsilon}\right)^{\frac{1}{6}} \left(\frac{M\delta}{t}\right)^{\frac{5}{6}} & \text{if } M \leq \frac{t}{\delta}. \end{cases} \end{aligned}$$

The expression in the brace is bounded by a constant provided α has been chosen less than $\frac{1}{16}$, so we have proved (53). □

Completion of proof of Theorem 4.1. With notation as above we have

$$\begin{aligned} |E| &\geq |\{x \in E : \text{card}(\{i : x \in C_\delta(x_i, r_i) \leq \mu\})\}| \\ &\geq \mu^{-1} \sum_j |\{x \in E \cap C_\delta(x_j, r_j) : \text{card}(\{i : x \in C_\delta(x_i, r_i)\}) \leq \mu\}| \\ &\gtrsim \mu^{-1} M \lambda \delta \\ &\gtrsim \lambda^{\frac{8}{3}} M^{\frac{5}{6}} \delta \end{aligned}$$

by (53). Consequently $(M\delta)^{\frac{5}{6}} \lesssim \frac{|E|}{\delta^{\frac{1}{6}} \lambda^{\frac{8}{3}}}$ and Theorem 4.1 is proved. □

Proof of Theorem 2'. It suffices to prove Theorem 2' for functions f with support in the unit disc. Therefore, we need only observe (as is done for example in [10]) that the function

$$\Phi(x, y) = |x - y|$$

satisfies the conditions (14), (15) at all points where $x \neq y$, and then use a partition of unity to reduce to Theorem 4.1. \square

Remark. One can consider various other cases besides Theorem 2' and Lemma 5.1 below. For example, one can consider circles with one of the coordinates of the center taken as the parameter r , instead of the radius. One gets the following result: If a set in \mathbb{R}^2 contains circles with centers at all points of a smooth curve, then the Hausdorff dimension of E is at least $\frac{11}{6}$, and the estimate (49) holds for the restriction of the circular maximal function to the curve. The proof is done most easily by using the remark at the beginning of Section 5 below. We note that E can have measure zero; cf. [12] (this reference was pointed out by W. Schlag).

5. Proof of Theorem 3'.

It is convenient to restate Theorem 4.1 using a different way of presenting the curve family (cf. also [10] in this connection). Suppose then that $U \subset \mathbb{R}^2 \times \mathbb{R}^3$, $\Psi : U \rightarrow \mathbb{R}$. We will denote variables in $\mathbb{R}^2 \times \mathbb{R}^3$ by (x, λ) , $x \in \mathbb{R}^2$, $\lambda \in \mathbb{R}^3$. Define $C(\lambda) = \{x \in \mathbb{R}^2 : \Psi(x, \lambda) = 0\}$ and let $C_\delta(\lambda)$ be its δ -neighborhood. Fix a point $(a, \lambda_0) \in U$, and let r_0 be the third coordinate of λ_0 . Assume the following: $\Psi(a, \lambda_0) = 0$, $\nabla_x \Psi(a, \lambda_0) \neq 0$, $\nabla_\lambda \Psi(a, \lambda_0) \neq 0$, and the cinematic curvature condition, i.e., that if x is close to a and λ, μ are close to λ_0 , and $C(\lambda)$ and $C(\mu)$ intersect at x , then either the unit tangent vectors to $C(\lambda)$ and $C(\mu)$ at x , or else the curvatures of $C(\lambda)$ and $C(\mu)$ at x , differ by $\gtrsim |\lambda - \mu|$. If f is supported in a small neighborhood of a , δ is small, $z \in (r_0 - \epsilon_0, r_0 + \epsilon_0)$ then define a maximal function $M_\delta f(z)$ via

$$M_\delta f(z) = \sup_\lambda \delta^{-1} \int_{C_\delta(\lambda)} |f|,$$

where the sup is taken over all parameter values λ with $|\lambda - \lambda_0| < \epsilon_0$ and $\lambda_{(3)} = z$. Then the estimate (49) is valid.

The proof is as follows: The maximal function is unaffected by the change of variables $\lambda \rightarrow (\lambda_{(1)}, \lambda_{(2)} + \lambda_{(3)}, \lambda_{(3)})$ or $\lambda \rightarrow (\lambda_{(1)} + \lambda_{(3)}, \lambda_{(2)}, \lambda_{(3)})$, so we may assume that $\frac{\partial \Psi}{\partial \lambda_{(3)}}(a, \lambda_0) \neq 0$. But then by the implicit function theorem the equation $\Psi(x, \lambda) = 0$ can be solved for λ_3 and we are reduced to the situation of Theorem 4.1.

We will now estimate the maximal function defined by (7) by applying this form of Theorem 4.1 with

$$(59) \quad \Psi((s, t), (x, y, z)) = s^2 - \frac{(t - y)^2}{z} - x.$$

Lemma 5.1. *If $p < \frac{8}{3}$ and $q = 2p'$, and if $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ satisfies $\text{supp } f \subset \{(s, t) : \frac{1}{2} \leq s \leq 2\}$, then*

$$\|\tilde{M}_\delta f\|_{L^q(\{z: \frac{1}{2} < z < 2\})} \lesssim \delta^{-\frac{1}{2}(\frac{3}{p}-1)} \|f\|_p.$$

Proof. One can easily check the cinematic curvature condition for the family of curves $H(x, y, z)$ in the region $\frac{1}{2} \leq s \leq 2$, by representing the curves as graphs over the t axis and using the formulas from first year calculus. We omit this calculation. Using Theorem 4.1 and the above remarks it follows that Lemma 5.1 is valid locally, i.e., is valid with the additional assumption that $\text{supp } f \subset \{(s, t) : -1 \leq t \leq 1\}$. The definition of \tilde{M}_δ is translation invariant in t , so the lemma is valid with bounds uniform in N provided that $\text{supp } f \subset Q_N$ for some N , where $Q_N = [\frac{1}{2}, 2] \times [N - 1, N + 1]$. On the other hand, when $z \in [\frac{1}{2}, 2]$ it is easy to check that $H_\delta(x, y, z)$ intersects only a bounded number of Q_N 's. Hence $\tilde{M}_\delta f \lesssim \sup_N \tilde{M}_\delta(\chi_{Q_N} f)$ and then the lemma follows in a standard way, namely:

$$\begin{aligned} \|\tilde{M}_\delta f\|_q &\leq \left\| \left(\sum_N \tilde{M}_\delta(\chi_{Q_N} f)^q \right)^{\frac{1}{q}} \right\|_q \\ &= \left(\sum_N \|\tilde{M}_\delta(\chi_{Q_N} f)\|_q^q \right)^{\frac{1}{q}} \\ &\lesssim \delta^{-\frac{1}{2}(\frac{3}{p}-1)} \left(\sum \|\chi_{Q_N} f\|_p^q \right)^{\frac{1}{q}} \\ &\lesssim \delta^{-\frac{1}{2}(\frac{3}{p}-1)} \|f\|_p, \end{aligned}$$

since the Q_N have finite overlap and $q \geq p$.

We will relate the Kakeya maximal function of a rotation invariant function to the maximal function \tilde{M}_δ . In order to do so we first recall the standard definitions of the x-ray transform RF and maximal x-ray transform R^*F , for functions $F : \mathbb{R}^3 \rightarrow \mathbb{R}$. Namely

$$RF(\ell) = \int_\ell F d\sigma.$$

Here ℓ denotes a line in \mathbb{R}^3 and $d\sigma$ is arc length. Also, $R^*F : \mathbb{P}^2 \rightarrow \mathbb{R}$,

$$R^*F(e) = \sup_{\ell: e \in \ell} RF(\ell),$$

where e_ℓ is the direction of the line ℓ .

Similarly, if f is a function on \mathbb{R}^2 then we define $\mathcal{H}f : \{(x, y, z) \in \mathbb{R}^3 : x > 0, \frac{1}{2} \leq z \leq 2\} \rightarrow \mathbb{R}$ via

$$\mathcal{H}f(x, y, z) = \int_{H(x,y,z)} f(s, t) dt$$

and $\mathcal{H}^* f : \{z \in \mathbb{R} : \frac{1}{2} \leq z \leq 2\} \rightarrow \mathbb{R}$ via

$$\mathcal{H}^* f(z) = \sup_{\substack{x > 0 \\ y \in \mathbb{R}}} \mathcal{H}f(x, y, z).$$

We will use complex notation for the first two variables in \mathbb{R}^3 when convenient, i.e., will denote the point (x_1, x_2, x_3) by $(x_1 + ix_2, x_3)$. If ℓ is a line in \mathbb{R}^3 which is not parallel to the x_3 axis then ℓ can be parametrized by its direction $(e^{i\theta}, \sqrt{z}) \in \mathbb{P}^2, z \geq 0$, and its closest point to the e_3 axis, which will be of the form $(i\alpha e^{i\theta}, y), \alpha \in \mathbb{R}, y \in \mathbb{R}$. In this notation we have:

Lemma 5.2. *If F is a rotation invariant function in \mathbb{R}^3 , $F(x) = f(\sqrt{x_1^2 + x_2^2}, x_3)$, then for any line ℓ which is not parallel to the x_3 axis and which satisfies $z \neq 0$,*

$$RF(\ell) = \sqrt{\frac{1+z}{z}} \mathcal{H}f(\alpha^2, y, z).$$

Proof. We have

$$RF(\ell) = \int F((i\alpha e^{i\theta}, y) + \tau(e^{i\theta}, \sqrt{z})) \sqrt{1+z} d\tau,$$

and since F is rotation invariant this implies

$$\begin{aligned} RF(\ell) &= \int_0^{2\pi} \int F(e^{i\phi}(i\alpha e^{i\theta} + \tau e^{i\theta}), y + \tau\sqrt{z}) \sqrt{1+z} d\tau \frac{d\phi}{2\pi} \\ &= \int_0^{2\pi} F((\tau + i\alpha)e^{i(\theta+\phi)}, y + \tau\sqrt{z}) \sqrt{1+z} d\tau \frac{d\phi}{2\pi} \\ &= \int_0^{2\pi} F\left(\left(\frac{t-y}{\sqrt{z}} + i\alpha\right) e^{i(\theta+\phi)}, t\right) \sqrt{\frac{1+z}{z}} dt \frac{d\phi}{2\pi} \\ &= \int f\left(\left|\frac{t-y}{\sqrt{z}} + i\alpha\right|, t\right) dt \sqrt{\frac{1+z}{z}} \\ &= \int_{H(\alpha^2, y, z)} f(s, t) dt \sqrt{\frac{1+z}{z}} \end{aligned}$$

as claimed.

We define $\Omega_\rho \subset \mathbb{P}^2$ via $\Omega_\rho = \{e : \sqrt{z} \in [\frac{\rho}{\sqrt{2}}, \sqrt{2}\rho]\}$, and $\tilde{\Omega} = \{e \in \mathbb{P}^2 : \sqrt{z} \in [\delta, \frac{1}{\delta}]\}$. We also define a variant of F_δ^* , as follows:

$$\overline{F}_\delta(e) \stackrel{\text{def}}{=} \delta^{-2} \sup_t \int_t |F|,$$

where t runs over δ -neighborhoods of lines ℓ with $e_\ell = e$. Thus $\overline{F}_\delta(e) \gtrsim F_\delta^*(e)$, and $\overline{F}_\delta(e) \approx F_\delta^*(e)$ if $z \leq 1$ and f is supported in $\{(s, t) : s \leq 2\}$. \square

Remark. At this point we are in a position to prove the Hausdorff dimension statement, Theorem 3. We only sketch the argument since of course Theorem 3 is also a corollary of Theorem 3'. Suppose F and f are as in Lemma 5.2, $F \geq 0$. By applying Lemma 5.2 to the convolution of F with the characteristic function of a disc of radius δ , we obtain

$$F_\delta^*(e) \lesssim \tilde{M}_\delta f(z) \quad \forall e \in \Omega_1.$$

Consequently by Lemma 5.1,

$$(60) \quad \|F_\delta^*\|_{L^q(\Omega_1)} \lesssim \delta^{-\frac{1}{2}(\frac{3}{p}-1)} \|F\|_p, \quad p < \frac{8}{3}, q = 2p'$$

for rotation invariant functions supported in $\{x \in \mathbb{R}^3 : \sqrt{x_1^2 + x_2^2} \in [\frac{1}{2}, 2]\}$. It is clear that the interval $[\frac{1}{2}, 2]$ here could be replaced by any other closed bounded interval not containing the origin. If E is a rotation invariant set which contains a line segment in every direction, then its intersection with $\sqrt{x_1^2 + x_2^2} \in [a, \frac{1}{a}]$ contains a segment in a positive measure set of Ω_1 -directions for suitable $a > 0$. We may then obtain the dimension statement by applying (60) to the characteristic function of the latter set.

The purpose of the contortions below is to pass from estimates like (60) to the scale invariant $L^p \rightarrow L^q$ estimates of Theorem 3'. As motivation, we note that (by the preceding remark) Theorem 3' and (60) have essentially the same geometric content, and on the other hand the values of p in Theorem 3' are larger than those in (60). It is therefore natural to interpolate between an estimate like (60) and an $L^\infty \rightarrow L^\infty$ estimate. This is in fact what we will do; cf. (77) below.

Lemma 5.3. *Assume that $p < \frac{17}{6}$ and that p is sufficiently close to $\frac{17}{6}$.⁶ Let $q = 2p'$. Then there is $\epsilon = \epsilon_p > 0$ such that if F and f are as in Lemma 5.2 and $\text{supp } f \subset \{(s, t) : \frac{1}{2} \leq s \leq 2\}$ then*

$$\|\overline{F}_\delta\|_{L^q(\tilde{\Omega})} \lesssim \delta^{-(\frac{3}{p}-1)+\epsilon} \|F\|_p.$$

Proof. We always assume $F \geq 0$. It suffices to prove

$$(61) \quad \|\overline{F}_\delta\|_{L^q(\Omega_\rho)} \lesssim \delta^{-(\frac{3}{p}-1)+\epsilon} \|F\|_p$$

⁶“sufficiently close to $\frac{17}{6}$ ” can be taken to mean that $p > \frac{5}{3}$.

with $\epsilon > 0$ independent of ρ , since then we obtain the estimate of the lemma with a slightly smaller value of ϵ by summing over dyadic ρ from δ to $\frac{1}{\delta}$. To prove (61) we consider two cases, (i) $\rho \leq 1$ and (ii) $\rho \geq 1$.

In case (i) we fix ρ and define a map $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ via $T(x_1, x_2, x_3) = (x_1, x_2, \rho x_3)$. Then the action of T on \mathbb{P}^2 is given by

$$(e^{i\theta}, \sqrt{z}) = e \Rightarrow e_T \stackrel{\text{def}}{=} (e^{i\theta}, \rho\sqrt{z})$$

i.e.,

$$(62) \quad (1+z)^{-\frac{1}{2}} R^*(F \circ T)(e) = (1+\rho^2 z)^{-\frac{1}{2}} R^* F(e_T).$$

Note that $e \in \Omega_1 \leftrightarrow e_T \in \Omega_\rho$. Also let $\Phi_{\delta\sigma}$ be the characteristic function of the set $\{x \in \mathbb{R}^3 : \sqrt{x_1^2 + x_2^2} < \delta, |x_3| < \sigma\}$, normalized to have L^1 norm 1, and similarly let $\phi_{\delta\sigma}$ be the L^1 normalized characteristic function of the set $\{(s, t) \in \mathbb{R}^2 : |s| < 2\delta, |t| < \sigma\}$. If $G = F \circ T$ and g is defined via $G(x) = g(\sqrt{x_1^2 + x_2^2}, x_3)$, and if $e \in \Omega_1$, then standard arguments together with (62) and Lemma 5.2 justify the following string of inequalities:

$$\begin{aligned} F_\delta^*(e_T) &\lesssim R^*(F * \Phi_{\delta\delta})(e_T) \\ &\approx R^*((F * \Phi_{\delta\delta}) \circ T)(e) \\ &= R^*\left(G * \Phi_{\delta \frac{\delta}{\rho}}\right)(e) \\ &\lesssim \mathcal{H}^*\left(g * \phi_{\delta \frac{\delta}{\rho}}\right)(z). \end{aligned}$$

Taking into account that the map $e \rightarrow e_T$ distorts areas by a factor of roughly ρ we therefore have

$$(63) \quad \|F_\delta^*\|_{L^q(\Omega_\rho)} \lesssim \rho^{\frac{1}{q}} \left\| \mathcal{H}^*\left(g * \phi_{\delta \frac{\delta}{\rho}}\right) \right\|_{L^q(\{z: \frac{1}{2} < z < 2\})},$$

and it remains to estimate the latter expression. □

Claim. Let $\sigma \leq 1$, $\delta \leq C_0\sigma$, and assume also that $\delta < \frac{1}{10}$. Fix $x > 0, y \in \mathbb{R}, z \in [\frac{1}{2}, 2]$, let $H = H(x, y, z)$ and let μ_H be the measure defined via $\int f d\mu_H = \int_{H(x,y,z)} f(s, t) dt$. Let $H_\alpha = H_\alpha(x, y, z)$, $S(y, \beta) = \{(s, t) \in \mathbb{R}^2 : \frac{1}{2} \leq s \leq 2, |t - y| \leq C_1\beta\}$ for an appropriate constant C_1 . Then when $\frac{1}{2} \leq s \leq t$,

$$(64) \quad \phi_{\delta\sigma} * \mu_H(s, t) \lesssim \sum_\alpha \alpha^{-1} \chi_{H_\alpha \cap S(y, \frac{\alpha}{\sigma})}(s, t) + R.$$

In the sum, α runs over all numbers of the form $2^j \max(\sigma^2, \delta), j \in \mathbb{Z}^+ \cup \{0\}, \alpha \leq \sigma$, and R satisfies

$$R = 0, \text{ if } \sigma^2 \leq \delta$$

$$\text{supp } R \subset \{(s, t) : |t - y| \leq C\sigma, |s - \sqrt{x}| \leq C\sigma^2\} \quad \text{and}$$

$$(65) \quad |R(s, t)| \lesssim (\delta\sigma)^{-1} \min \left(\sqrt{\delta}, \frac{\delta}{\sqrt{|s - \sqrt{x}|}} \right), \quad \text{if } \sigma^2 > \delta.$$

Proof of the claim. This is a standard type of estimation using (implicitly) the curvature of the curves $H(x, y, z)$.

Fix a point (s_0, t_0) with $\frac{1}{2} \leq s_0 \leq 2$; we will prove the claim at the point (s_0, t_0) . Let B be the rectangle centered at (s_0, t_0) with sides parallel to the axes, of lengths 4δ in the s direction and 2σ in the t direction. We always assume that $B \cap H \neq \emptyset$ which is evidently satisfied if $\phi_{\delta\sigma} * \mu_H(s_0, t_0) \neq 0$. We will also use the notation $(s(t), t)$ to denote points of H ; thus

$$(66) \quad s(t) = \sqrt{x} + \frac{(t - y)^2}{z}.$$

We split the proof of the claim into three cases.

(i) $|t_0 - y| \geq 2\sigma$.

Then $|t - t_0| \leq \sigma$ implies (cf. (66)) that

$$(67) \quad \left| \frac{\partial s(t)}{\partial t} \right| \approx |t_0 - y|.$$

Formula (67) immediately implies that $\mu_H(B \cap H) \lesssim \min \left(\frac{\delta}{|t_0 - y|}, \sigma \right)$, i.e.,

$$(68) \quad |\phi_{\delta\sigma} * \mu_H| \lesssim (\delta\sigma)^{-1} \min \left(\frac{\delta}{|t_0 - y|}, \sigma \right) \approx (\delta + \sigma|t_0 - y|)^{-1}.$$

Next, since $B \cap H$ is nonempty we can fix t with $|t - t_0| \leq \sigma$ and $|s(t) - s_0| \leq 2\delta$. Then $|s_0 - s(t_0)| \leq |s_0 - s(t)| + |s(t) - s(t_0)| \lesssim \delta + \sigma|t_0 - y|$, where we used (67). Thus $(s_0, t_0) \in H_\alpha \cap S(y, \frac{\alpha}{\sigma})$, with $\alpha \approx \delta + \sigma|t_0 - y|$. This and (68) imply the claim (with $R = 0$) in case (i).

(ii) $|t_0 - y| \leq 2\sigma$, and $\sigma^2 \leq \delta$.

In this case we just use that $\mu_H(B \cap H) \leq \sigma$, i.e.,

$$(69) \quad |\phi_{\delta\sigma} * \mu_H(s_0, t_0)| \lesssim \delta^{-1} \approx (\delta + \sigma^2)^{-1}.$$

Also (by (66))

$$(70) \quad \left| \frac{\partial s(t)}{\partial t} \right| \lesssim \sigma$$

when $|t - t_0| \leq \sigma$. Fix t with $|t - t_0| \leq \sigma$ and $|s(t) - s_0| \leq 2\delta$. Then

$$(71) \quad |s_0 - s(t_0)| \leq |s_0 - s(t)| + |s(t) - s(t_0)| \lesssim \delta + \sigma^2.$$

We used (70). It follows that $(s_0, t_0) \in H_\alpha \cap S(y, \frac{\alpha}{\sigma})$, $\alpha \approx \delta + \sigma^2$. The claim (with $R = 0$) follows from this and (69).

(iii) $|t_0 - y| \leq 2\sigma$, and $\sigma^2 \geq \delta$.

In this case we set $R(s_0, t_0) = \phi_{\delta\sigma} * \mu_H$. It suffices to show that (65) then holds. From (66) we easily obtain the following two estimates:

$$\begin{aligned} \left| \frac{\partial s(t)}{\partial t} \right| &\approx \sqrt{s(t) - \sqrt{x}} \\ \left| \frac{\partial^2 s(t)}{\partial t^2} \right| &\approx 1 \end{aligned}$$

when $|t - t_0|$ is small. The first estimate implies that if $|s_0 - \sqrt{x}|$ is large compared with δ , then $\mu_H(B \cap H) \lesssim \frac{\delta}{\sqrt{|s_0 - \sqrt{x}|}}$. The second estimate implies that in all cases $\mu_H(B \cap H) \lesssim \sqrt{\delta}$. Combining these, we get $\mu_H(B \cap H) \lesssim \min\left(\sqrt{\delta}, \frac{\delta}{\sqrt{|s_0 - \sqrt{x}|}}\right)$ in all cases, which gives the estimate for $|R(s, t)|$ in (65). On the other hand, we're assuming $|t_0 - y| \leq 2\sigma$, and the proof of (71) did not use that $\sigma^2 \leq \delta$, so (71) is valid in the present case also. Using (70) once more it follows that $|s(t_0) - \sqrt{x}| \lesssim \sigma^2$, so $|s_0 - \sqrt{x}| \lesssim \sigma^2$ and the claim is completely proved. \square

Applying the claim with $\sigma = \frac{\delta}{\rho}$, we estimate the right hand side of (63) as follows. For each $z \in [\frac{1}{2}, 2]$,

$$(72) \quad \mathcal{H}^*\left(g * \phi_{\delta \frac{\delta}{\rho}}\right)(z) \lesssim \sum_{\alpha} \sup_y \tilde{M}_{\alpha}\left(\chi_{S(y, \frac{\alpha\rho}{\delta})}g\right)(z) + E(z).$$

Here α runs over all numbers of the form $2^j \max\left(\delta, \frac{\delta^2}{\rho^2}\right)$ with $j \geq 0$ and $\alpha \leq \frac{\delta}{\rho}$, and $E(z) \leq \|gR\|_1$ where R is a function (depending on z) which satisfies the estimates (65) for a certain s_0 and t_0 , with $\sigma = \frac{\delta}{\rho}$. We first discuss the last term in (72). Estimates (65) imply that if $\delta > \rho^2$ then

$$\begin{aligned} \|R\|_{p'}^{p'} &\lesssim \left(\frac{\rho}{\delta}\right)^{p'} \int_{|t-y| \leq \frac{\delta}{\rho}} \int_{|s-\sqrt{x}| \leq \frac{\delta^2}{\rho^2}} |s - \sqrt{x}|^{-\frac{p'}{2}} ds dt \\ &\approx \left(\frac{\rho}{\delta}\right)^{2p'-3} \end{aligned}$$

i.e., $\|R\|_{p'} \lesssim \left(\frac{\rho}{\delta}\right)^{\frac{3}{p}-1}$, and of course $R = 0$ if $\delta \leq \rho^2$. Using this together with Hölder's inequality and the change of variables formula $\|g\|_p = \rho^{-\frac{1}{p}} \|f\|_p$, we obtain

$$\begin{aligned} \rho^{\frac{1}{q}} \|E\|_q &\leq \rho^{\frac{1}{q}} \|E\|_{\infty} \lesssim \begin{cases} \left(\frac{\rho}{\delta}\right)^{\frac{3}{p}-1} \rho^{\frac{1}{q}-\frac{1}{p}} \|f\|_p & \text{if } \rho \leq \delta^{\frac{1}{2}} \\ 0 & \text{if } \rho \geq \delta^{\frac{1}{2}} \end{cases} \\ &\lesssim \delta^{-\frac{3}{4}\left(\frac{3}{p}-1\right)} \|f\|_p. \end{aligned}$$

The exponent $-\frac{3}{4}(\frac{3}{p} - 1)$ here is of course of the form $-(\frac{3}{p} - 1) + \epsilon$, so, substituting (72) into (63), we see that it suffices to prove

$$(73) \quad \rho^{\frac{1}{q}} \sum_{\alpha} \left\| \sup_y \tilde{M}_{\alpha} \left(g\chi_{S(y, \frac{\alpha\rho}{\delta})} \right) \right\|_q \lesssim \delta^{-(\frac{3}{p}-1)+\epsilon} \|f\|_p.$$

In order to prove (73) we fix α and let $\{I_j\}$ be a covering of \mathbb{R} by intervals of length $3C\frac{\alpha\rho}{\delta}$, with finite overlap. Here C is the constant in the definition of $S(y, \beta)$. Then we let $S_j = [\frac{1}{2}, 2] \times I_j \subset \mathbb{R}^2$. We have

$$(74) \quad \begin{aligned} \rho^{\frac{1}{q}} \left\| \sup_y \tilde{M}_{\alpha} \left(g\chi_{S(y, \frac{\alpha\rho}{\delta})} \right) \right\|_q &\lesssim \rho^{\frac{1}{q}} \left\| \sup_j \tilde{M}_{\alpha}(g\chi_{S_j}) \right\|_q \\ &\lesssim \rho^{\frac{1}{q}} \left\| \left(\sum_j \tilde{M}_{\alpha}(g\chi_{S_j})^q \right)^{\frac{1}{q}} \right\|_q \\ &= \rho^{\frac{1}{q}} \left(\sum_j \|\tilde{M}_{\alpha}(g\chi_{S_j})\|_q^q \right)^{\frac{1}{q}} \end{aligned}$$

with all L^q norms being taken over $\frac{1}{2} < z < 2$, and we now proceed to bound the terms in the sum over j . We have an $L^\infty \rightarrow L^\infty$ estimate

$$(75) \quad \|\tilde{M}_{\alpha}h\|_{\infty} \lesssim \frac{\alpha\rho}{\delta} \|h\|_{\infty} \quad \text{if } \text{supp } h \subset S_j,$$

since each set $H_{\alpha}(x, y, z)$ intersects S_j in measure $\lesssim \alpha \cdot \frac{\alpha\rho}{\delta}$. Now fix p_0 with

$$(76) \quad 2p - 3 < p_0 < \frac{8}{3}.$$

This is possible since $p < \frac{17}{6}$. Let $q_0 = 2p'_0$. Interpolating the bound (75) with the bound

$$\|\tilde{M}_{\alpha}h\|_{q_0} \lesssim \alpha^{-\frac{1}{2}(\frac{3}{p_0}-1)} \|h\|_{p_0}$$

of Lemma 5.1, we obtain

$$(77) \quad \|\tilde{M}_{\alpha}h\|_{\bar{q}} \lesssim \alpha^{-\frac{1}{2}(\frac{3}{p_0}-1)\frac{p_0}{p}} \left(\frac{\alpha\rho}{\delta}\right)^{1-\frac{p_0}{p}} \|h\|_p \quad \text{if } \text{supp } h \subset S_j.$$

In (77), the number \bar{q} is equal to $q_0\frac{p}{p_0}$ and therefore larger than q . It follows that (77) remains valid with \bar{q} replaced by q . Applying (77) with this modification to the terms in the sum (74), we obtain

$$(74) \lesssim \rho^{\frac{1}{q}} \alpha^{-\frac{1}{2}(\frac{3}{p_0}-1)\frac{p_0}{p}} \left(\frac{\alpha\rho}{\delta}\right)^{1-\frac{p_0}{p}} \left(\sum_j \|g\chi_{S_j}\|_p^q \right)^{\frac{1}{q}}$$

$$\lesssim \rho^{\frac{1}{q}} \alpha^{-\frac{1}{2}} \left(\frac{3}{p_0} - 1\right)^{\frac{p_0}{p}} \left(\frac{\alpha \rho}{\delta}\right)^{1 - \frac{p_0}{p}} \|g\|_p.$$

The last line follows since $q > p$ and the sets S_j have finite overlap. Using $\|g\|_p = \rho^{-\frac{1}{p}} \|f\|_p$ and juggling some indices, we obtain

$$(74) \lesssim \rho^{\frac{3}{2} - \frac{3+2p_0}{2p}} \alpha^{1 - \frac{3+p_0}{2p}} \delta^{\frac{p_0}{p} - 1}.$$

The power of α is negative by (76), and $\alpha \gtrsim \delta$ so by summing over α we bound the left side of (73) by $\rho^{\frac{3}{2} - \frac{3+2p_0}{2p}} \delta^{\frac{p_0-3}{2p}}$. The power of ρ here is positive if p is close to $\frac{17}{6}$, so we obtain

$$\text{left hand side of (73)} \lesssim \delta^{\frac{p_0-3}{2p}} \|f\|_p.$$

Now we are done (with case (i) of (61)), since $\frac{3-p_0}{2p} < \frac{3}{p} - 1$ by (76).

Case (ii) is simply a scaling argument. We consider the same map $T(x_1, x_2, x_3) = (x_1, x_2, \rho x_3)$, where now $\rho \geq 1$. If $e \in \Omega_1$ and t is a δ -neighborhood of a line in the e_T direction, then it is not difficult to see that $T^{-1}t$ is essentially a δ -neighborhood of a line in the e direction. Consequently, since volumes are distorted by ρ ,

$$\overline{F}_\delta(e_T) \lesssim \rho \overline{(F \circ T)}_\delta(e).$$

For large ρ the map $e \rightarrow e_T$ contracts areas by a factor of roughly ρ^2 , so

$$\begin{aligned} \|\overline{F}_\delta\|_{L^q(\Omega_\rho)} &= \rho^{1-\frac{2}{q}} \left\| \overline{(F \circ T)}_\delta \right\|_{L^q(\Omega_1)} \\ &\lesssim \rho^{1-\frac{2}{q}} \delta^{-\left(\frac{3}{p}-1\right)+\epsilon} \|F \circ T\|_p \\ &= \rho^{1-\frac{2}{q}-\frac{1}{p}} \delta^{-\left(\frac{3}{p}-1\right)+\epsilon} \|F\|_p \\ &= \delta^{-\left(\frac{3}{p}-1\right)+\epsilon} \|F\|_p \end{aligned}$$

where we used the $\rho = 1$ case and the relation between p and q . This finishes the proof of (61), hence of Lemma 5.3.

Proof of Theorem 3'. We may assume that p is sufficiently close to $\frac{17}{6}$. We claim first that if $\text{supp } f \subset \{(s, t) : \frac{r}{2} \leq s \leq 2r\}$ then

$$(78) \quad \|\overline{F}_\delta\|_{L^q(\tilde{\Omega})} \lesssim r^{-\epsilon} \delta^{-\left(\frac{3}{p}-1\right)+\epsilon} \|F\|_p$$

with ϵ as in Lemma 5.3.

The case $r = 1$ is the content of Lemma 5.3. The general case will follow by scaling. If we set $G(x) = F(rx)$ then we have

$$\overline{F}_\delta(e) = r \overline{G}_{\frac{\delta}{r}}(e),$$

so by applying Lemma 5.3 to G we get

$$\|\overline{F}_\delta\|_{L^q(\tilde{\Omega})} = r \|\overline{G}_\delta\|_{L^q(\tilde{\Omega})}$$

$$\begin{aligned} &\lesssim r \left(\frac{\delta}{r}\right)^{-\left(\frac{3}{p}-1\right)+\epsilon} \|G\|_p \\ &= r \left(\frac{\delta}{r}\right)^{-\left(\frac{3}{p}-1\right)+\epsilon} r^{\frac{3}{p}} \|F\|_p, \end{aligned}$$

which is (78).

An easy fact about F_δ^* is that $F_\delta^*(e) \lesssim F_\delta^*(\tilde{e})$ if $|e - \tilde{e}| \leq \delta$. Consequently $\|F_\delta^*\|_q$ and $\|F_\delta^*\|_{L^q(\tilde{\Omega})}$ are comparable, so (78) implies that

$$\|F_\delta^*\|_q \lesssim r^{-\epsilon} \delta^{-\left(\frac{3}{p}-1\right)+\epsilon} \|F\|_p, \quad \text{supp } f \subset \left\{ (s, t) : \frac{r}{2} \leq s \leq 2r \right\}.$$

It follows on summing a geometric series that

$$(79) \quad \|F_\delta^*\|_q \lesssim \delta^{-\left(\frac{3}{p}-1\right)} \|F\|_p$$

provided $\text{supp } f \subset \{(s, t) : s \geq \delta\}$.

We still have to prove estimate (79) when $\text{supp } f \subset \{(s, t) : s \leq \delta\}$. In order to do so we first consider a function $\psi : \mathbb{R} \rightarrow \mathbb{R}$, and define

$$\mathcal{A}\psi(t) = \sup_I \frac{1}{|I|} \int_I |\psi|,$$

where I runs over all intervals of length t . We can dominate $\mathcal{A}\psi(t)$ by the average over $(0, t)$ of the nonincreasing rearrangement of $|\psi|$ so by Hardy's inequality

$$\int_0^\infty t^{\frac{q}{p}-1} \mathcal{A}\psi(t)^q dt \lesssim \|\psi\|_p^q, \quad 1 < p \leq q < \infty.$$

Define a function ψ of one variable via $\psi(x) = \delta^{-2} \int F(x_1, x_2, x) dx_1 dx_2$. Because of the support restriction on F , it is easy to see that

$$F_\delta^*(e) \lesssim \frac{\delta}{\theta} \mathcal{A}\psi\left(\frac{\delta}{\theta}\right),$$

where θ is the angle between e and the x_3 axis. Consequently,

$$\begin{aligned} \|F_\delta^*\|_q^q &\lesssim \int_0^{\frac{\pi}{2}} \left(\frac{\delta}{\theta} \mathcal{A}\psi\left(\frac{\delta}{\theta}\right)\right)^q \theta d\theta \\ &\lesssim \int_0^\infty (t \mathcal{A}\psi(t))^q \delta^2 t^{-3} dt \\ &\lesssim \delta^2 \|\psi\|_p^q \end{aligned}$$

since $q - 3 = \frac{q}{p} - 1$.

On the other hand, it follows using Hölder's inequality that $\|\psi\|_p \lesssim \delta^{-\frac{2}{p}} \|F\|_p$. We conclude that $\|F_\delta^*\|_q \lesssim \delta^{\frac{2}{q}-\frac{2}{p}} \|F\|_p = \delta^{-\left(\frac{3}{p}-1\right)} \|F\|_p$ and the proof of Theorem 3' is complete. \square

References

- [1] A.S. Besicovitch and R. Rado, *A plane set of measure zero containing circumferences of every radius*, J. London Math. Soc., **43** (1968), 717-719.
- [2] J. Bourgain, *Besicovitch type maximal operators and applications to Fourier analysis*, Geometric and Functional Analysis, **1** (1990), 147-187.
- [3] A. Cordoba, *The Keakeya maximal function and spherical summation multipliers*, Amer. J. Math., **99** (1977), 1-22.
- [4] R.L. Graham, B.L. Rothschild and J. H. Spencer, *Ramsey Theory*, 2nd edition, Wiley-Interscience, 1990.
- [5] J.R. Kinney, *A thin set of circles*, Amer. Math. Monthly, **75** (1968), 1077-1081.
- [6] J.E. Marsden, *Elementary Classical Analysis*, W.H. Freeman, 1974.
- [7] J.M. Marstrand, *Packing circles in the plane*, Proc. London Math. Soc., **55** (1987), 37-58.
- [8] G. Mockenhaupt, A. Seeger and C. Sogge, *Wave front sets and Bourgain's circular maximal theorem*, Ann. Math., **134** (1992), 207-218.
- [9] H. Pecher, *Nonlinear small data scattering for the wave and Klein-Gordon equation*, Math. Z., **185** (1984), 261-270.
- [10] C. Sogge, *Propagation of singularities and maximal functions in the plane*, Inv. Math., **104** (1991), 349-376.
- [11] ———, *Fourier Integrals in Classical Analysis*, Cambridge University Press, 1993.
- [12] M. Talagrand, *Sur la mesure de la projection d'un compact et certaines familles de cercles*, Bull. Sci. Math., **104(2)** (1980), 225-231.
- [13] T. Wolff, *An improved bound for Keakeya type maximal functions*, Revista Math. Iberoamericana, **11** (1995), 651-674.

Received January 22, 1996 and revised May 20, 1997. The second author was supported by DMS 93-07872.

RYERSON POLYTECHNIC UNIVERSITY
TORONTO, ONTARIO M5B 2K3
CANADA
E-mail address: lkolasa@acs.ryerson.ca

253-37 CALTECH
PASADENA, CA 91125
E-mail address: wolff@cco.caltech.edu