J. Eur. Math. Soc. (Online first) DOI 10.4171/JEMS/1283

Pablo Shmerkin

A non-linear version of Bourgain's projection theorem

Dedicated to the memory of Jean Bourgain

Received July 10, 2020

Abstract. We prove a version of Bourgain's projection theorem for parametrized families of C^2 maps, which refines the original statement even in the linear case by requiring non-concentration only at a single natural scale. As one application, we show that if A is a Borel set of Hausdorff dimension close to 1 in $\mathbb{R}^2$ or close to $3/2$ in $\mathbb{R}^3$, then for $y \in A$ outside of a very sparse set, the pinned distance set $\{|x-y| : x \in A\}$ has Hausdorff dimension at least $1/2 + c$, where c is universal. Furthermore, the same holds if the distances are taken with respect to a C^2 norm of positive Gaussian curvature. As further applications, we obtain new bounds on the dimensions of spherical projections, and an improvement over the trivial estimate for incidences between δ-balls and δ-neighborhoods of curves in the plane, under fairly general assumptions. The proofs depend on a new multiscale decomposition of measures into "Frostman pieces" that may be of independent interest.

Keywords. Projection theorems, distance sets, pinned distance sets, incidences, Hausdorff dimension, patterns, Lipschitz functions

Contents

Pablo Shmerkin: Department of Mathematics, University of British Columbia, Vancouver, Canada; pshmerkin@math.ubc.ca

Mathematics Subject Classification (2020): Primary 28A75; Secondary 28A80, 05D99, 26A16, 49Q15

1. Introduction and statement of results

1.1. Distance sets

The Falconer distance set conjecture, originating in [8], asserts that if $A \subset \mathbb{R}^d$, $d \geq 2$, is a Borel set of Hausdorff dimension $d/2$, then the *distance set*

$$\Delta(A) = \{|x - y| : x, y \in A\} \subset [0, \infty)$$

has Hausdorff dimension 1. Despite efforts by many mathematicians, the conjecture remains open in all dimensions. A variant of this problem has also received much attention. From now on, all sets are assumed to be Borel. Given $y \in \mathbb{R}^d$ and $A \subset \mathbb{R}^d$, denote the *pinned distance set* by

$$\Delta^y(A) = \{|x - y| : x \in A\}.$$

Let $\dim_{\mathrm{H}}$ denote Hausdorff dimension. It is possible that the pinned version of Falconer's conjecture holds, that is, if $\dim_{\mathrm{H}}(A) \geq d/2$, then there is $y \in A$ such that $\dim_{\mathrm{H}}(\Delta^y(A)) = 1$.

Recent deep results [6, 9, 19] imply that if $\dim_{\mathrm{H}}(A) > \alpha(d)$, then there is $y \in A$ such that $\Delta^y(A)$ has positive Lebesgue measure, where

$$\alpha(d) = \begin{cases} \frac{5}{4} & \text{if } d = 2, \\ \frac{d^2}{2d-1} & \text{if } d \geq 3. \end{cases} \tag{1.1}$$

In recent years, substantial progress has been achieved in the plane under the assumption $\dim_{\mathrm{H}}(A) > 1$ [9, 18, 20, 23, 26–28]. For example, it is known [27] that if A is a planar set of equal Hausdorff and packing dimension, and this common value is > 1, then there is $y \in A$ such that $\dim_{\mathrm{H}}(\Delta^y(A)) = 1$. For general planar sets of Hausdorff dimension > 1, it is known that there is $y \in A$ such that $\dim_{\mathrm{H}}(\Delta^y(A)) \geq 29/42$ [28] and $\dim_{\mathrm{P}}(\Delta^y(A)) > 0.933$ [18], where $\dim_{\mathrm{P}}$ denotes packing dimension. In all these works it is crucial that $\dim_{\mathrm{H}}(A) > 1$; the methods break down if one only assumes that $\dim_{\mathrm{H}}(A) = 1$.

For general ambient dimension, under the hypothesis $\dim_{\mathrm{H}}(A) = d/2$, Falconer proved in his original paper [8] that $\dim_{\mathrm{H}}(\Delta(A)) \geq 1/2$. This bound turned out to be quite hard to improve upon. By combining results of N. Katz–T. Tao [15] and J. Bourgain [3], it follows that if A is a Borel planar set with $\dim_{\mathrm{H}}(A) \geq 1$, then

$$\dim_{\mathrm{H}}(\Delta(A)) \geq 1/2 + c, \tag{1.2}$$

where c is a small universal constant. We discuss the result of Katz–Tao and Bourgain in more detail below.

Less attention has been given to the problem of obtaining lower bounds for the dimension of distance sets when $\dim_{\mathrm{H}}(A) < d/2$. To our knowledge, the best such bound appearing in the literature was proved in Falconer's paper [8]: if $\dim_{\mathrm{H}}(A) \leq d/2$, then

$$\dim_{\mathrm{H}}(\Delta(A)) \geq \dim_{\mathrm{H}}(A) - \frac{d-1}{2}. \tag{1.3}$$

Note that this is vacuous if $\dim_{\mathrm{H}}(A) \leq (d-1)/2$.

In this paper we improve upon several of the previously mentioned results, particularly in dimensions 2 and 3.

Theorem 1.1. *Let $d \geq 2$. Given $d - 2 < \kappa, \alpha < d$, there is a number $c = c_d(\alpha, \kappa) > 0$, depending continuously on α, κ, such that the following holds.*

Let $A \subset \mathbb{R}^d$ be a Borel set with $\dim_{\mathrm{H}}(A) \geq \alpha$. Then

$$\dim_{\mathrm{H}}(\{y \in \mathbb{R}^d : \dim_{\mathrm{H}}(\Delta^y(A)) < \alpha/d + c\}) \leq \kappa.$$

Moreover, the same holds if the pinned distance set is defined with respect to any C^2 norm whose unit ball has everywhere positive Gaussian curvature (with c independent of the choice of norm).

In the plane, this result improves upon the bound (1.2) of Katz–Tao and Bourgain in several ways: (a) it provides a pinned version, (b) furthermore, not only does the pinned version hold for some $y \in A$, but in fact it holds for y outside of a set of dimension κ, arbitrarily small (with the gain c depending on κ) – this is new even when $\dim_{\mathrm{H}}(A) > 1$, (c) it works for more general smooth, curved norms, (d) it extends to values of $\dim_{\mathrm{H}}(A) \in (0, 1)$. Moreover, Theorem 1.1 provides an improvement over the bound $\dim_{\mathrm{H}}(\Delta(A)) \geq 1/2$ for Borel sets of dimension $3/2$ in $\mathbb{R}^3$, and also for Borel sets of dimension > 2 in $\mathbb{R}^4$. In dimensions ≥ 5, the value $\alpha(d)$ from (1.1) is smaller than $d - 2$, so Theorem 1.1 becomes far less interesting. We note that a careful inspection of the proofs in [15], combined with the discretized sum-product theorem of [3], might lead to the analogue of (1.2) also for sets of dimension in $(0, 1)$ and perhaps also in higher dimensions. We believe that the improvements (a)–(c) noted above do not follow in any direct way from previous approaches.

Remark 1.2. In dimensions $d \geq 3$, at least one of the assumptions $\kappa > d - 2, \alpha > d - 2$ in Theorem 1.1 is necessary, as can be seen from the examples $X = S^1 \times \{0\} \subset \mathbb{R}^2 \times \mathbb{R}^{d-2}$ and $X = S^{d-2} \times \{0\} \subset \mathbb{R}^{d-1} \times \mathbb{R}$. However, such assumptions may not be necessary if one only considers $y \in X$.

1.2. A non-linear version of Bourgain's projection theorem

There are well known connections between many important problems at the interface of analysis and geometric measure theory, such as the Kakeya, Furstenberg set, discretized sum-product, discretized projection and Falconer distance set problems. Indeed, all of these problems to some extent deal with incidences between tubes. However, the connections, even when they are explicit, are rarely straightforward. In [15], Katz and Tao introduced discretized versions of three conjectures which were at the time open. Two of the conjectures were: (a) $\dim_{\mathrm{H}}(\Delta(A)) \geq 1/2 + c$ if A is a planar Borel set with $\dim_{\mathrm{H}}(A) = 1$, (b) there is no Borel subring of the reals of Hausdorff dimension $1/2$. Note that (a)$\Rightarrow$(b): if R is a ring of dimension $1/2$, then $\Delta(R \times R) \subset \sqrt{R_{\geq 0}}$. Among other things, Katz and Tao proved that certain discretized versions of these conjectures are equivalent, and that the discretized version of (1.2) (which is rather involved) implies the actual bound (1.2).

In [3], Bourgain proved the discretized version of the ring conjecture, which is nowadays known as the discretized sum-product theorem. Hence, in combination with [15], this established the bound (1.2). A few years later, in [4], Bourgain refined the discretized sum-product theorem to obtain what is now known as Bourgain's (discretized) projection theorem. We recall this theorem below. (In fact, many of the ideas to go from sum-product to projections are already implicit in [3,15].) Thus, there is a known path from Bourgain's projection theorem to the estimate (1.2). In this article, we take a rather different path that we believe is more flexible, and can be used to make progress on other problems in combinatorial fractal geometry. We view the maps $\Delta^y(x) = |x - y|$ as a family of maps from $\mathbb{R}^d$ to $\mathbb{R}$, parametrized by the point y. These maps are smooth if we are careful to separate the domains of the x and the y, but they are non-linear. In [15], an important step in the overall argument is applying a projective transformation that linearizes a family of projections. This argument seems to be rather constrained. The approach of this paper is to develop a *non-linear* version of Bourgain's projection theorem. In doing so, we also obtain some new insights even in the linear case.

We state first a continuous (as opposed to single-scale) version of our main result. We denote the open r-neighborhood of H by $H^{(r)}$. If μ is a Borel measure on a metric space X and $g : X \to Y$ is a Borel map, then we denote the push-forward measure to Y by $g\mu$, that is, $g\mu(\cdot) = \mu(g^{-1}\cdot)$. If $\mu(A) > 0$, then μ_A denotes the normalized restriction $\mu(A)^{-1}\mu|_A$. Given $x \in \mathbb{R}^d, x \neq 0$, we let $\mathrm{dir}(x) = x/|x| \in S^{d-1}$. Finally, we denote the Grassmannian of linear k-planes in $\mathbb{R}^d$ by $\mathbb{G}(d,k)$.

Theorem 1.3. *Given $\kappa > 0$ and $0 < \alpha < d$ there is $\eta = \eta_d(\kappa,\alpha) > 0$ (that can be taken continuous in κ,α) such that the following holds.*

Let (Λ, ν) be a Borel probability space and let $U \subset \mathbb{R}^d$ be an open domain. Let $F : \Lambda \times U \to \mathbb{R}$ be a Borel function such that, for each $\lambda \in \Lambda$, the map $F_\lambda(x) = F(\lambda, x) : U \to \mathbb{R}$ is C^2 without singular points. For each $x \in \mathbb{R}^d$ define the map

$$\theta_x(\lambda) = \mathrm{dir}(\nabla F_\lambda(x)) \in S^{d-1}.$$

Let $A \subset U$ be a Borel set of dimension $\geq \alpha$ such that for all $x \in A$ there are a set Λ_x with $\nu(\Lambda_x) > 0$ and a number $C_x > 0$ satisfying

$$\theta_x \nu_{\Lambda_x}(H^{(r)}) \leq C_x r^{\kappa} \quad \text{for all } H \in \mathbb{G}(d, d-1),\ r \in (0,1]. \tag{1.4}$$

Then there is $\lambda \in \operatorname{supp}(\nu)$ such that

$$\dim_{\mathrm{H}}(F_\lambda A) \geq \alpha/d + \eta.$$

We make some remarks on this statement.

Remark 1.4. Bourgain's (continuous) projection theorem corresponds precisely to the special case in which $\Lambda = \Lambda_x = S^{d-1}$ and $F_\lambda(x) = \langle \lambda, x \rangle$ is orthogonal projection in direction λ. In this case, θ_x is the identity map for all x, and so the decay condition (1.4) has to be satisfied by the measure ν on S^{d-1}.

Remark 1.5. Allowing the set Λ_x to depend on x is important in our applications, such as Theorem 1.1. By a formal argument, in the case $\Lambda_x \equiv \Lambda$ for all x, the conclusion holds for ν-almost all λ.

Remark 1.6. It is enough that (1.4) holds for all $x \in A$ outside of a set E of dimension $< \alpha$, since we can then apply the theorem to $A \setminus E$.

Bourgain's projection theorem described above is deduced from a discretized version, which is fully stated as Theorem 3.1 below. It is often this discretized version that gets used in applications, such as [5, 16]. Correspondingly, we have the following discretized version of Theorem 1.3. The number of dyadic cubes of side length 2^{-m} hitting a set X is denoted by $\mathcal{N}(X, m)$, and $|\cdot|$ refers to Lebesgue measure.

Theorem 1.7. *Fix $d \geq 2$. Given $\kappa > 0$, there is $\eta = \eta_d(\kappa) > 0$ such that the following holds for $\varepsilon < \varepsilon_d(\kappa)$.*

Let $X \subset [0,1]^d$ be a union of 2^{-m}-dyadic cubes. Let U be a neighborhood of X and let (Λ, ν) be a Borel probability space. Let $F : \Lambda \times U \to \mathbb{R}$ be a Borel function such that, for each $\lambda \in \Lambda$, the map $x \mapsto F_\lambda(x) = F(\lambda, x) : U \to \mathbb{R}$ is C^2 without singular points. For each $x \in \mathbb{R}^d$ define the map

$$\theta_x(\lambda) = \operatorname{dir}(\nabla F_\lambda(x)) \in S^{d-1}.$$

Assume that $C_\Lambda = \sup_\lambda \|F_\lambda\|_{C^2} < \infty$ and $c_\Lambda = \inf_\lambda \inf_{x \in X} |\nabla F_\lambda(x)| > 0$.

Let m be large enough in terms of c_Λ, C_Λ and all the previous parameters. Suppose X satisfies the single-scale non-concentration condition

$$|X \cap B(x, |X|^{1/d})| \leq 2^{-\kappa m}|X| \quad \text{for all } x. \tag{1.5}$$

Suppose that for every $x \in X$ there is a set $\Lambda_x \subset \Lambda$ with $\nu(\Lambda_x) \geq 2^{-\varepsilon m}$ such that

$$\theta_x \nu_{\Lambda_x}(H^{(r)}) \leq 2^{\varepsilon m} r^{\kappa} \tag{1.6}$$

for all hyperplanes $H \in \mathbb{G}(d, d-1)$ and all $r \in [2^{-m}, 1]$.

Then there exist $\lambda \in \operatorname{supp}(\nu)$ *and a set* $X' \subset X$ *with* $|X'| \geq 2^{-2\varepsilon m}|X|$ *such that*

$$\mathcal{N}(F_\lambda X'', m) \geq 2^{\eta m}\mathcal{N}(X, m)^{1/d}$$

for all sets $X'' \subset X'$ *with* $|X''| \geq 2^{-\varepsilon m}|X'|$.

If $\Lambda_x = \Lambda$ *for all* x*, then one can take* $X' = X$*, and the conclusion holds for all* λ *outside of a set of* ν*-measure* $\leq 2^{-\varepsilon m}$.

Some remarks are in order.

Remark 1.8. Again, Bourgain's original theorem corresponds to $F_\lambda(x) = \langle \lambda, x\rangle$ and $\Lambda_x \equiv \Lambda$; see Theorem 3.1 below. However, hypothesis (1.5) is weaker than the non-concentration assumption in Theorem 3.1, since it is required at the single scale $|X|^{1/d}$. Furthermore, this is the natural scale that makes this assumption sharp: if X is a cube of side length $|X|^{1/d}$ (or dense in such a cube, or the union of very few such cubes), then $|F(X)| \approx |X|^{1/d}$ for all smooth maps $F : \mathbb{R}^d \to \mathbb{R}$ without singular points (including projections). In short, (1.5) is saying that X is not concentrated in a few cubes, with the "measure gain" exponent η depending on the quality of this non-concentration, given by κ. We remark that the theorem clearly fails if X has measure close either to 1 or to 2^{-dm}; however, in both these cases X is highly concentrated in a cube (of size 1 or 2^{-m}) and therefore these extremes are ruled out by (1.5).

Remark 1.9. In the case $d = 2$, the non-concentration assumption (1.6) is a standard Frostman (power law) assumption on the $\theta_x \nu_{\Lambda_x}$ measures of balls. For $d \geq 3$, the intersection $S^{d-1} \cap H$ is a maximal $(d-2)$-sphere in S^{d-1}, and (1.6) says that $\theta_x \nu_{\Lambda_x}$ is not concentrated near such subspheres. The example of a segment shows that, already for linear projections, such a condition is necessary.

Remark 1.10. The "gain" η in Bourgain's theorem is effective in principle, but extremely small. The value of η in Theorem 1.7 is even smaller than that of Bourgain's projection theorem (as a matter of fact, it is equal for sets with some uniform decay such as Ahlfors-regular sets). We do not make the connection explicit, although it is not hard to extract it from the proofs, and we make no attempt at optimization since the values involved are in any event tiny. In the work in progress [13], the authors provide explicit estimates for a closely related result involving entropy gain. Even though this does not automatically translate into an explicit value in the context of Theorem 3.1, it is plausible that with some additional work this will yield explicit estimates in Theorem 1.7.

1.3. Strategy of proof and a multiscale decomposition into Frostman pieces

We now explain the overall strategy of the proof of Theorems 1.3 and 1.7, and discuss informally a new multiscale decomposition of measures that is at the heart of the proof (see Section 4 and in particular Theorem 4.1 for precise statements).

Bourgain's original proof of Theorem 3.1 appears to be intrinsically restricted to the linear setting. Rather than modifying the proof, we perform a regularization and multiscale decomposition of a Frostman measure μ on the given set X, and then linearize F

at every scale in the multiscale decomposition. We then apply Bourgain's theorem (as a black box) to every small piece of μ and every scale in the multiscale decomposition.

The overall strategy is not new. In particular, we make use of a variant of a formula for estimating the entropy of smooth projections in terms of multiscale decompositions that goes back, in various forms, to [12, 14, 18, 23]. See Proposition 5.1. Because the precise formulation we need does not appear in the literature, we include a full proof in Appendix A. Our approach is closest to that of [18] which, in addition to a multiscale decomposition, required an initial decomposition of the measure μ into "regular Moran constructions". The same decomposition plays a key role in this paper.

In order to apply Bourgain's projection theorem to small pieces of the measure, we need to verify that these small pieces satisfy the required non-concentration assumption. The main technical innovation of this paper is a new multiscale decomposition of a "regular Moran measure" so that the conditional measures on small cubes of certain sizes are Frostman measures up to a small exponential error (so they satisfy essentially the strongest possible non-concentration decay). By itself, this is not enough to deduce any useful consequences, because it may well happen that nearly all of these conditional measures look like either Lebesgue measure or an atom so that, even though they are trivially Frostman measures, there is no gain to be achieved. To deal with this, we show that if the original measure satisfies the (minimal) non-concentration assumption (1.5), then the scales in the multiscale decomposition can be chosen so that, for a "positive density" set of scales, the conditional measures have "intermediate size", i.e. they are quantitatively separated from both Lebesgue and atomic measures. See Theorem 4.1 for details. We hope this new multiscale decomposition of measures (and related ones) will have further applications.

Even with Theorem 4.1 in hand, it is still possible that for many scales the conditional measures do look close to Lebesgue or atomic. In this setting, Bourgain's projection theorem does not apply. For such scales, we rely on quantitative versions of more classical projection theorems, essentially going back to R. Kaufman [17] and to K. Falconer [7]. See Section 3.

Incidentally, as the above sketch indicates, we only need to apply Theorem 3.1 when κ (nearly) matches the size of the set being projected, i.e. when there is (near) optimal non-concentration. Potentially it may be easier to obtain quantitative estimates under this stronger assumption, which would then translate into quantitative estimates under the weaker assumptions of Theorem 1.7.

To deduce Theorem 1.1 from Theorem 1.3, we appeal to a result of T. Orponen [24] on spherical projections. In fact, we need a quantitative version and an extension to higher dimensions of Orponen's argument. Since we follow Orponen's ideas quite closely, the proofs of these facts are deferred to Appendix B.

1.4. Structure of the paper and further results

In Section 2 we introduce some general notation, and some preliminary lemmas. In particular, in §2.2 we review the important concept of regular measures, and their properties,

and in §2.3 we introduce the convenient notions of robust measures and robust entropy, and relate them to Hausdorff dimension.

Section 3 deals with discretized projection theorems. The main result of this section is Theorem 3.4 which, in the language of robust measures, combines and unifies Bourgain's discretized projection theorem with quantitative versions of more classical projection theorems going back to R. Kaufman and K. Falconer.

Section 4 contains two new multiscale decompositions of (regular) measures, Theorems 4.1 and 4.9. The proofs are reduced to combinatorial statements about Lipschitz functions, which take up most of the section.

Theorems 1.3 and 1.7 are proved in Section 5.

In Section 6 we derive several applications and generalizations of Theorems 1.3 and 1.7. We begin in §6.1 with a straightforward application to one-parameter real-analytic families. In §6.2, we complete the proof of Theorem 1.1. In §6.3, we apply Theorem 1.7 to obtain incidence bounds for discretized families of curves; see Theorem 6.5. This result is new even for lines; in this particular case, it complements results of M. Bateman and V. Lie [1], and of L. Guth, N. Solomon and H. Wang [10]. Note that Theorem 1.3 only applies to projections onto the real line. In §6.4, we indicate how to use W. He's extension of Bourgain's projection theorem to higher rank projections in order to derive a higher rank extension of Theorem 1.3; see Theorem 6.13. Finally, we use this higher rank version to study spherical projections in §6.5, in particular extending recent results of T. Orponen, and of B. Liu and C.-Y. Shen.

As indicated earlier, Appendix A contains the proof of Proposition 5.1, a crucial estimate of the entropy of a smooth image of a measure in terms of linearized images of small pieces of the measure in a multiscale decomposition. These are a key ingredient in the proofs of Theorems 1.3 and 1.7. The ideas in this appendix are not new, but the statements differ enough from existing ones in the literature that we have chosen to include a complete proof. Likewise, Appendix B contains the proof of Theorem 6.3, a higher-dimensional variant of a spherical projection result of T. Orponen [24]. The results in this appendix are used in the proofs of Theorem 1.1 in §6.2 and of Theorem 6.15 in §6.5. Although we follow Orponen's ideas closely, there are several changes in the details, so again we include full details for completeness.

2. Notation and preliminary results

2.1. Notation

Closed balls in $\mathbb{R}^d$ will be denoted by $B(x,r)$, or by $B^d(x,r)$ when the ambient dimension is not clear from context.

We use Landau's $O(\cdot)$ notation: given $X > 0$, $O(X)$ denotes a positive quantity bounded above by CX for some constant $C > 0$. If C is allowed to depend on some other parameters, these are denoted by subscripts. We sometimes write $X \lesssim Y$ in place of $X = O(Y)$ and likewise with subscripts. We write $X \gtrsim Y$, $X \approx Y$ to denote $Y \lesssim X$, $X \lesssim Y \lesssim X$ respectively.

The family of Borel probability measures on a metric space X is denoted by $\mathcal{P}(X)$, and the family of Borel finite measures by $\mathcal{M}(X)$.

Logarithms are always to base 2.

We let $\mathcal{D}_j$ be the family of half-open 2^{-j}-dyadic cubes in $\mathbb{R}^d$ (where d is understood from context), and let $\mathcal{D}_j(x)$ be the only cube in $\mathcal{D}_j$ containing $x \in \mathbb{R}^d$. Given a measure $\mu \in \mathcal{P}(\mathbb{R}^d)$, we also let $\mathcal{D}_j(\mu)$ be the cubes in $\mathcal{D}_j$ with positive μ-measure. We recall that, given $A \subset \mathbb{R}^d$, we also denote by $\mathcal{N}(A, j)$ the number of cubes in $\mathcal{D}_j$ that intersect A.

A 2^{-m}-*measure* is a measure μ in $\mathcal{P}([0,1)^d)$ such that μ_Q is a multiple of Lebesgue measure on Q for $Q \in \mathcal{D}_m$. Hence, 2^{-m}-measures are defined down to resolution 2^{-m}. The set of 2^{-m}-measures on $\mathbb{R}^d$ will be denoted $\mathcal{P}_m^d$. Likewise, a 2^{-m}-*set* is a union of cubes in $\mathcal{D}_m$.

Lebesgue measure on $\mathbb{R}^d$ will be denoted by either $|\cdot|$ or $\mathcal{L}$; the ambient dimension d will be clear from context.

Due to our use of dyadic cubes, sometimes we will need to deal with supports in the dyadic metric, i.e. given $\mu \in \mathcal{P}([0,1)^d)$ we let

$$\operatorname{supp}_{\mathrm{d}}(\mu) = \{x : \mu(\mathcal{D}_j(x)) > 0 \text{ for all } j \in \mathbb{N}\}.$$

Note that $\mu(\operatorname{supp}_{\mathrm{d}}(\mu)) = 1$ and $\operatorname{supp}_{\mathrm{d}}(\mu) \subset \operatorname{supp}(\mu)$.

If a measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ has a density in L^p, then the density is sometimes also denoted by μ, and in particular $\|\mu\|_p$ stands for the L^p norm of the density.

Let $\mu \in \mathcal{P}([0,1)^d)$. Recall that if $\mu(A) > 0$, then $\mu_A = \frac{1}{\mu(A)}\mu|_A$ is the normalized restriction of μ to A. If Q is a dyadic cube and $\mu(Q) > 0$, then we denote $\mu^Q = \operatorname{Hom}_Q \mu_Q$, where Hom_Q is the homothety renormalizing Q to $[0,1)^d$. Thus, μ^Q is a magnified and renormalized copy of the restriction of μ to Q; we sometimes refer to such measures as *conditional measures* on Q.

Recall that the Grassmannian of linear k-planes in $\mathbb{R}^d$ is denoted by $\mathbb{G}(d,k)$. When $k = 1$, we often identify $\mathbb{G}(d,1)$ with S^{d-1}; the fact that the identification is two-to-one does not cause any issues in practice. We denote the manifold of *affine* k-planes in $\mathbb{R}^d$ by $\mathbb{A}(d,k)$.

2.2. Regular measures

A key role in the paper is played by measures with a uniform tree (or Moran) structure when represented in base 2^T. This notion is made precise in the next definition. Recall that $\mathcal{P}_m^d$ stands for the family of 2^{-m}-measures in $\mathbb{R}^d$.

Definition 2.1. Given a sequence $\sigma = (\sigma_1, \ldots, \sigma_\ell) \in [0,d]^\ell$ and $T \in \mathbb{N}$, we say that $\mu \in \mathcal{P}_{T\ell}^d$ is $(\sigma; T)$-*regular* if for any $Q \in \mathcal{D}_{jT}(\mu)$, $1 \le j \le \ell$, we have

$$\mu(Q) \le 2^{-T\sigma_j}\mu(\widehat{Q}) \le 2\mu(Q),$$

where $\widehat{Q}$ is the only cube in $\mathcal{D}_{(j-1)T}$ containing Q. When T is understood from context we will simply write that μ is σ-*regular*.

We note that the same notion appears in [18], but with a different normalization for the parameters σ_j.

Lemma 2.2. *Let $\nu \in \mathcal{P}([0,1)^d)$ be $(\sigma;T)$-regular for some $\sigma \in [0,d]^\ell$, $T \in \mathbb{N}$. Write $m = T\ell$ and $X = \operatorname{supp}_{\mathtt{d}}(\nu)$.*

(i) $2^{-j}2^{-T(\sigma_1+\cdots+\sigma_j)} \le \nu(Q) \le 2^{-T(\sigma_1+\cdots+\sigma_j)}$ *for all* $Q \in \mathcal{D}_{jT}(X)$.

(ii) $2^{T(\sigma_1+\cdots+\sigma_\ell)} \le \mathcal{N}(X,m) \le 2^{T(\sigma_1+\cdots+\sigma_\ell)+\ell}$.

(iii) $2^{-\ell}\nu \le \mathbf{1}_X/|X| \le 2^\ell \nu$.

Proof. From the definition it is clear that if $Q \in \mathcal{D}_{jT}(\nu) = \mathcal{D}_{jT}(X)$, then

$$2^{-j}2^{-T\sigma_1}\cdots 2^{-T\sigma_j} \le \nu(Q) \le 2^{-T\sigma_1}\cdots 2^{-T\sigma_j}, \tag{2.1}$$

which is (i). Claim (ii) follows easily from (2.1) applied with $j = \ell$. For the final claim, it is enough to establish the inequality for $Q \in \mathcal{D}_m$, and this follows from (2.1) (which implies $\nu(Q) \le 2^{m/T}\nu(Q')$ for $Q, Q' \in \mathcal{D}_m$) and (ii). ∎

Starting with an arbitrary 2^{-m}-measure, a pigeonholing argument (which we learned from Bourgain's work) allows us to find a set X with "large" measure such that μ_X is regular:

Lemma 2.3. *Fix $T, \ell \ge 1$. Write $m = T\ell$, and let $\mu \in \mathcal{P}_m^d$. Then there is a set $X \subset \operatorname{supp}_{\mathtt{d}}(\mu)$ with $\mu(X) \ge (2dT+2)^{-\ell}$ such that μ_X is $(\sigma;T)$-regular for some $\sigma \in [0,d]^\ell$.*

See [18, Lemma 3.4] for the proof of this particular statement. Iterating the above lemma, we can decompose an arbitrary measure $\mu \in \mathcal{P}_m^d$ into regular measures, plus a negligible error. This is the content of the next crucial lemma whose proof can be found in [18, Corollary 3.5].

Lemma 2.4. *Fix $T, \ell \ge 1$ and $\varepsilon > 0$. Write $m = T\ell$, and let $\mu \in \mathcal{P}_m^d$. There exists a family of pairwise disjoint 2^{-m}-sets $X_1, \ldots, X_N$ with $X_k \subset \operatorname{supp}_{\mathtt{d}}(\mu)$ such that:*

(i) $\mu(\bigcup_{k=1}^N X_k) \ge 1 - 2^{-\varepsilon m}$. *In particular, if $\mu(A) > 2^{-\varepsilon m}$, then there exists k such that $\mu_{X_k}(A) \ge \mu(A) - 2^{-\varepsilon m}$.*

(ii) $\mu(X_k) \ge 2^{-\delta m}$, *where $\delta = \varepsilon + \log(2dT+2)/T$.*

(iii) *Each μ_{X_k} is $(\sigma(k);T)$-regular for some $\sigma(k) \in [0,d]^\ell$.*

2.3. Robust measures and robust entropy

We will need to deal with various different notions of "largeness" of a measure. The next definition captures the core property shared by all of these notions.

Definition 2.5. A a measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ is called (α,δ,m)-*robust* if, for any set A with $\mu(A) > 2^{-\delta m}$, one has $\mathcal{N}(A,m) > 2^{\alpha m}$.

Lemma 2.6. *If $\mu \in \mathcal{P}(\mathbb{R}^d)$ is (α,δ,m)-robust for all $m \ge m_0$, then $\dim_{\mathrm{H}}(A) \ge \alpha$ for all Borel sets A with $\mu(A) > 0$. Furthermore, if $\{B_j\}$ is a union of disjoint balls of radius 2^{-m}*

for some $m \geq m_0$ such that $\mu(B_j) \geq 2^{-\alpha m}$, then

$$\mu\Big(\bigcup_j B_j\Big) \leq 2^{-\delta m}.$$

Proof. Since there can be at most $2^{\alpha m}$ disjoint balls of measure $\geq 2^{-\alpha m}$, the second claim is immediate from the definition of robustness. It follows by a standard covering lemma argument and Borel–Cantelli that for μ-almost all x, $\mu(B(x,r)) \leq r^{\alpha}$ for all sufficiently small r (depending on x). The first claim is now a consequence of the mass distribution principle. ∎

Recall that the *entropy* of $\mu \in \mathcal{P}(X)$ with respect to a finite partition $\mathcal{A}$ of X (or of a set of full μ-measure in X) is defined by

$$H(\mu, \mathcal{A}) = \sum_{A \in \mathcal{A}} \mu(A) \log(1/\mu(A)).$$

We will sometimes write $H_m(\mu)$ in place of $H(\mu, \mathcal{D}_m)$. Note that this quantity is *not* normalized. For more about entropy, see Appendix A.

Given two measures μ, ν on the same space X and a number $\Delta > 0$, we write $\nu \leq \Delta\mu$ if $\nu(A) \leq \Delta\mu(A)$ for all Borel sets A. In particular, $\mu_B \leq \mu(B)^{-1}\mu$.

Definition 2.7. Let $\mu \in \mathcal{P}(\mathbb{R}^d)$, fix $\Delta \geq 1$, and let $\mathcal{A}$ be a finite partition of $\mathrm{supp}(\mu)$. We define the Δ-*robust entropy* $H^{\Delta}(\mu, \mathcal{A})$ as

$$\inf\{H(\nu, \mathcal{A}) : \nu \in \mathcal{P}(\mathbb{R}^d),\ \nu \leq \Delta\mu\}.$$

In the case $\mathcal{A} = \mathcal{D}_m$, we sometimes write $H_m^{\Delta}(\mu)$ in place of $H^{\Delta}(\mu, \mathcal{D}_m)$.

The next lemma asserts that robust measures have large robust entropy.

Lemma 2.8. *Given $\alpha, \delta, \varepsilon > 0$, the following holds for all $m \geq m_0(\delta, \varepsilon)$: if $\mu \in \mathcal{P}(\mathbb{R}^d)$ is (α, δ, m)-robust, then*

$$H_m^{2^{\delta m/2}}(\mu) \geq (\alpha - \varepsilon)m.$$

Proof. It follows from the definition that if $\nu \leq 2^{\delta m/2}\mu$, then ν is $(\alpha, \delta/2, m)$-robust. So it is enough to show that if ν is (α, δ', m)-robust and m is sufficiently large (depending on $\alpha, \delta', \varepsilon$), then

$$H(\nu, \mathcal{D}_m) \geq (\alpha - \varepsilon)m.$$

Now, if ν is (α, δ', m)-robust, then the ν-mass of the union of all the cubes in $\mathcal{D}_m$ of ν-measure $> 2^{-\alpha m}$ is $\leq 2^{-\delta' m} \leq 1/2$ (as there are $< 2^{\alpha m}$ such cubes). If A denotes the union of all the cubes in $\mathcal{D}_m$ of ν-mass $\leq 2^{-\alpha m}$, then $\nu_A(Q) \leq 2^{1-\alpha m}$ for $Q \in \mathcal{D}_m$, and therefore, by the concavity of entropy,

$$H(\nu, \mathcal{D}_m) \geq \nu(A) H(\nu_A, \mathcal{D}_m) \geq (1 - 2^{-\delta' m})(\alpha m - 1) \geq (\alpha - \varepsilon)m,$$

provided m is large enough in terms of δ', ε. ∎

3. Discretized projection theorems

Given $\theta \in S^{d-1}$, we denote the orthogonal projection $x \mapsto \langle \theta, x \rangle$, $\mathbb{R}^d \to \mathbb{R}$, by P_θ. If μ is a measure on $\mathbb{R}^d$, we also write $\mu_\theta = P_\theta \mu$.

We begin by recalling Bourgain's discretized projection theorem. There are several equivalent variants of the statement; the one we state is the special case $m = 1$ from [11, Theorem 1].

Theorem 3.1. *Given $0 < \alpha < d$ and $0 < \kappa < 1$ there exist $\delta, \eta > 0$ such that the following holds for all sufficiently large m. Let $X \subset B^d(0,1)$, and let $\rho \in \mathcal{P}(S^{d-1})$ satisfy*

$$\begin{aligned} \mathcal{N}(X, m) &\geq 2^{m(\alpha-\delta)}, \\ \mathcal{N}(X \cap B(x,r), m) &\leq 2^{\delta m} r^\kappa \mathcal{N}(X, m) \quad \text{for all } r \in [2^{-m}, 1],\ x \in B^d(0,1), \\ \rho(H^{(r)} \cap S^{d-1}) &\leq 2^{\delta m} r^\kappa \quad \text{for all } r \in [2^{-m}, 1],\ H \in \mathbb{G}(d, d-1). \end{aligned}$$

Then there is a set $E \subset S^{d-1}$ with $\rho(E) \leq 2^{-\delta m}$ such that if $\theta \in S^{d-1} \setminus E$ and $X' \subset X$ satisfies $\mathcal{N}(X', m) \geq 2^{-\delta m} \mathcal{N}(X, m)$, then

$$\log \mathcal{N}(P_\theta X', m) \geq m(\alpha/d + \eta).$$

Note that if we make η and δ slightly smaller, then they also work for nearby values of κ and α. Hence there is no loss of generality in assuming that η and δ are continuous functions of (κ, α).

While Bourgain's projection theorem will be our main tool, we will also need to consider the case in which $\log \mathcal{N}(A, m)/m$ is close to either 0 or d. Crucially, we need δ to be independent of $\log \mathcal{N}(A, m)$ in the estimates; clearly, in Theorem 3.1 the value of δ must depend on α if we allow values of α close to 0 or d. Hence we need to revisit other (more classical) projection theorems, and combine them into a single result that deals with the three regimes which $\log \mathcal{N}(A, m)/m$ is close to 0, close to d or far from both 0 and d. We start with a result that essentially goes back to R. Kaufman [17] in the 1960s. Before stating it, recall that the σ-energy of $\nu \in \mathcal{M}(\mathbb{R}^d)$ is defined as

$$\mathcal{E}_\sigma(\nu) = \iint |x - y|^{-\sigma} \, d\nu(x) \, d\nu(y) \in (0, \infty].$$

Theorem 3.2. *Fix $0 < \sigma < \kappa < 1$. Let $\rho \in \mathcal{P}(S^{d-1})$ satisfy*

$$\rho(H^{(r)} \cap S^{d-1}) \leq C r^\kappa \quad (H \in \mathbb{G}(d, d-1),\ r > 0),$$

and fix $\mu \in \mathcal{P}(\mathbb{R}^d)$. Then

$$\int_{S^{d-1}} \mathcal{E}_\sigma(\mu_\theta) \, d\rho(\theta) \leq \left(1 + \frac{C\sigma}{\kappa - \sigma}\right) \mathcal{E}_\sigma(\mu).$$

Proof. The claim is implicit in the proof of [22, Theorem 5.1]; since it is not explicitly stated in this form, we repeat the short argument for completeness. Fix $x \in \mathbb{R}^d \setminus \{0\}$, and note that

$$\{\theta \in \mathbb{R}^d : |P_\theta(x)| < \delta\} = x^\perp + B(0, \delta/|x|),$$

and hence, using the assumption on ρ,

$$\rho\{\theta \in S^{d-1} : |P_\theta(x)| < \delta\} \leq C(\delta/|x|)^\kappa.$$

Using this and Fubini, we estimate

$$\begin{aligned}\int_{S^{d-1}} |P_\theta(x)|^{-\sigma}\, d\rho(\theta) &= \int_0^\infty \rho\{\theta : |P_\theta(x)|^{-\sigma} > r\}\, dr \\ &= \int_0^{|x|^{-\sigma}} 1\, dr + \int_{|x|^{-\sigma}}^\infty \rho\{\theta : |P_\theta(x)| < r^{-1/\sigma}\}\, dr \\ &\leq \left(1 + \frac{C\sigma}{\kappa - \sigma}\right)|x|^{-\sigma}.\end{aligned}$$

Using Fubini, we conclude

$$\begin{aligned}\int_{S^{d-1}} \mathcal{E}_\sigma(\mu_\theta)\, d\rho(\theta) &= \int_{S^{d-1}} \iint_{\mathbb{R}^d\times\mathbb{R}^d} |P_\theta(x-y)|^{-\sigma}\, d\mu(x)\, d\mu(y)\, d\rho(\theta) \\ &\leq \left(1 + \frac{C\sigma}{\kappa - \sigma}\right) \iint_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^{-\sigma}\, d\mu(x)\, d\mu(y),\end{aligned}$$

as claimed. ∎

The following is a quantitative form of K. Falconer's classical bound on the dimension of exceptional projections [7].

Theorem 3.3. *Let $\rho \in \mathcal{M}(S^{d-1})$ satisfy*

$$\rho(B(\theta, r)) \leq r^\kappa \quad (\theta \in S^{d-1},\ r > 0),$$

and fix $\mu \in \mathcal{P}(\mathbb{R}^d)$ such that $\mathcal{E}_{d-\kappa}(\mu) < \infty$. Then $\mu_\theta \in L^2$ for ρ-almost all θ, and

$$\int_{S^{d-1}} \|\mu_\theta\|_2^2\, d\rho(\theta) \lesssim_d \mathcal{E}_{d-\kappa}(\mu).$$

Proof. In the course of the proof of [22, Theorem 5.6] it is shown that

$$\int_{S^{d-1}} \int_{\mathbb{R}^d} |\widehat{\mu}_\theta(x)|^2\, dx\, d\rho(\theta) \lesssim_d \int_{\mathbb{R}^d} |\widehat{\mu}(x)|^2 (1 + |x|)^{-\kappa}\, dx.$$

Note that our κ corresponds to τ in [22, Theorem 5.6], and $t = 1$ in our setting. The well-known expression of the energy in terms of the Fourier transform (see e.g. [22, Theorem 3.10]) together with Plancherel yields the result. ∎

To conclude this section, we combine Theorems 3.1, 3.2 and 3.3 in the language of robust measures.

Theorem 3.4. *Given $0 < \kappa < 1$ there exists $\eta = \eta(\kappa) > 0$ such that the following holds for all sufficiently small $\delta \le \delta_0(\kappa)$ and all sufficiently large $m \ge m_0(\delta)$. Fix $\alpha \in [0, d]$. Let $\mu \in \mathcal{P}_m^d$, and let $\rho \in \mathcal{P}(S^{d-1})$ satisfy*

$$\mu(B(x,r)) \le 2^{\delta m} r^\alpha \quad \text{for all } r \in [2^{-m}, 1],\ x \in [0,1)^d,$$
$$\rho(H^{(r)} \cap S^{d-1}) \le 2^{\delta m} r^\kappa \quad \text{for all } r \in [2^{-m}, 1],\ H \in \mathbb{G}(d, d-1).$$

Then there is a set $E \subset S^{d-1}$ with $\rho(E) \le 2^{-\delta m}$ such that μ_θ is $(\gamma(\alpha), \delta, m)$-robust for all $\theta \in S^{d-1} \setminus E$, where

$$\gamma(\alpha) = \begin{cases} \alpha - 6\delta & \text{if } \alpha < \kappa/2, \\ \alpha/d + \eta & \text{if } \kappa/2 \le \alpha \le d - \kappa/2, \\ 1 - 6\delta & \text{if } \alpha > d - \kappa/2. \end{cases}$$

Proof. **First case: $\alpha \in [\kappa/2, d - \kappa/2]$.** As already noted, we may assume that in Theorem 3.1 the same values of δ and η work for all $\alpha \in [\kappa/2, d - \kappa/2]$ (hence δ, η depend on κ only). We will in fact assume that $\delta = \delta(\kappa)$ is small enough that the conclusion of Theorem 3.1 holds for 4δ in place of δ.

Theorem 3.1 applies to sets rather than measures, as in our current context. We will use a fairly standard argument to deal with this. For an integer $j \ge 0$, let

$$X_j = \bigcup \{Q \in \mathcal{D}_m : 2^{-j-1} < \mu(Q) \le 2^{-j}\},$$

and set $J = \{j : \mu(X_j) \ge 2^{-2\delta m}\}$. Note that

$$\mu\Big(\bigcup_{j=2dm}^{\infty} X_j \Big) \le 2^{dm} 2^{-2dm} = 2^{-dm} < 2^{-2\delta m}.$$

In particular, $J \subset \{0, \dots, 2dm - 1\}$, and if we set $Z = [0,1)^d \setminus \bigcup_{j \in J} X_j$, then

$$\mu(Z) \le 2dm 2^{-2\delta m} + 2^{-dm} \le 3dm 2^{-2\delta m}. \tag{3.1}$$

Now the non-concentration condition on μ implies that, for $j \in J$,

$$\mu_{X_j}(B(x,r)) \le \mu(X_j)^{-1} \mu(B(x,r)) \le 2^{3\delta m} r^\alpha \quad (x \in [0,1)^d,\ r \in [2^{-m}, 1]).$$

In particular,

$$\mu_{X_j}(Q) \lesssim 2^{(3\delta - \alpha)m} \quad \text{for all } Q \in \mathcal{D}_m,$$

so that $\mathcal{N}(X_j, m) \gtrsim 2^{(\alpha - 3\delta)m}$. Also, by definition, $\mathbf{1}_{X_j} / |X_j| \le 2\mu_{X_j}$. Hence

$$|X_j \cap B(x,r)| \lesssim 2^{3\delta m} |X_j| r^\alpha \quad (x \in [0,1)^d,\ r \in [2^{-m}, 1]),$$

and since X_j is a union of 2^{-m}-cubes, this translates into a corresponding bound for the counting numbers $\mathcal{N}(X_j \cap B(x,r), m)$.

We have checked that the hypotheses of Theorem 3.1 hold for X_j, $j \in J$ (with 4δ in place of δ). Let E_j be the exceptional set given by the theorem, and define $E = \bigcup_{j\in J} E_j$. Note that

$$\rho(E) \leq |J|2^{-4\delta m} \leq 2dm2^{-4\delta m} < 2^{-\delta m}.$$

Fix $\theta \in S^{d-1} \setminus E$, and let A be a set with $P_\theta\mu(A) \geq 2^{-\delta m}$. It follows from (3.1) and the decomposition

$$\mu = \mu(Z)\mu_Z + \sum_{j\in J} \mu(X_j)\mu_{X_j}$$

that there is $j \in J$ such that

$$\mu_{X_j}(P_\theta^{-1}A) \geq \mu(P_\theta^{-1}A) - \mu(Z) \geq \tfrac{1}{2}2^{-\delta m},$$

and hence $|X_j \cap P_\theta^{-1}A| \geq \frac{1}{4}2^{-\delta m}|X_j|$. Again using the fact that X_j is a union of cubes in $\mathcal{D}_m$, we get a corresponding estimate for counting numbers. Since $\theta \notin E_j$, Theorem 3.1 implies that

$$\log \mathcal{N}(A, m) \geq (\alpha/d + \eta)m,$$

and thus we have verified that $P_\theta\mu$ is $(\alpha/d + \eta, \delta, m)$-robust, as desired.

Second case: $\alpha < \kappa/2$**.** It follows from the non-concentration assumption on μ that

$$\begin{aligned}\mathcal{E}_\alpha(\mu) &\lesssim \sum_{p=0}^{m-1} 2^{p\alpha}(\mu\times\mu)\{(x,y) : |x-y| \leq 2^{1-p}\}\\ &\leq \sum_{p=0}^{m-1} 2^{p\alpha} \max_x \mu(B(x, 2^{1-p})) \lesssim m2^{\delta m}.\end{aligned}$$

Hence we deduce from Theorem 3.2 (applied with $C = 2^{\delta m}$ and α in place of σ) that

$$\int_{S^{d-1}} \mathcal{E}_\alpha(\mu_\theta)\, d\rho(\theta) \lesssim_\kappa m2^{2\delta m},$$

and therefore $\rho(E) \leq 2^{-\delta m}$, where

$$E = \{\theta \in S^{d-1} : \mathcal{E}_\alpha(\mu_\theta) \geq C_\kappa m2^{3\delta m}\},$$

for a suitable $C_\kappa > 0$. Fix $\theta \in S^{d-1} \setminus E$ and suppose $\mu_\theta(A) \geq 2^{-\delta m}$. Then, writing $\nu = (\mu_\theta)_A$,

$$\mathcal{E}_\alpha(\nu) \lesssim \mu_\theta(A)^{-2}\mathcal{E}_\alpha(\mu_\theta) \leq C_\kappa m2^{5\delta m}.$$

On the other hand, it follows e.g. from [18, Lemma 3.1] and Cauchy–Schwarz that

$$\mathcal{E}_\alpha(\nu) \gtrsim_\alpha 2^{\alpha m} \sum_{I\in\mathcal{D}_m} \nu(I)^2 \geq 2^{\alpha m}\mathcal{N}(A, m)^{-1}.$$

Combining the last two displayed equations we see that μ_θ is $(\alpha - 6\delta, \delta, m)$-robust if m is large enough, as claimed.

Third case: $\alpha > d - \kappa/2$**.** To begin, we note that, arguing as above and using the non-concentration assumption on μ, we get

$$\mathcal{E}_{d-\kappa}(\mu) \lesssim \sum_{p=0}^{m-1} 2^{p(d-\kappa)} \max_x \mu(B(x, 2^{1-p})) \lesssim_\kappa 2^{\delta m}.$$

Applying Theorem 3.3 to $2^{-\delta m}\rho$, we deduce that, provided m is large enough,

$$\int_{S^{d-1}} \|\mu_\theta\|_2^2 \, d\rho(\theta) \lesssim 2^{2\delta m}.$$

Let $E = \{\theta : \|\mu_\theta\|_2^2 \geq C_d 2^{3\delta m}\}$, where C_d is large enough, so that $\rho(E) \leq 2^{-\delta m}$. Fix $\theta \in S^{d-1} \setminus E$ and A such that $\mu_\theta(A) \geq 2^{-\delta m}$. Writing again $\nu = (\mu_\theta)_A$, we have $\|\nu\|_2^2 \leq \mu_\theta(A)^{-2}\|\mu_\theta\|_2^2 \leq C_d 2^{5\delta m}$, and by a well-known application of Cauchy–Schwarz (see e.g. [18, Lemma 6.5]) we conclude that $\mathcal{N}(A, m) \geq 2^m \|\nu\|_2^{-2} \geq 2^{(1-6\delta)m}$, confirming that μ_θ is $(1 - 6\delta, \delta, m)$-robust. ∎

4. Multiscale decompositions of regular measures

4.1. A new multiscale decomposition

The goal of this section is to establish Theorem 4.1, providing a new kind of multiscale decomposition of a regular measure μ. Recall that, by Lemma 2.4, one can decompose an arbitrary 2^{-m}-measure into regular pieces plus an error term so, as we will see, this decomposition is also useful to study general measures.

Roughly speaking, the conclusion of the theorem says that given a regular measure μ, one can find a sequence of scales 2^{-m_j} such that for $Q \in \mathcal{D}_{m_j}$, the conditional measures μ^Q satisfy a near-Frostman decay condition. Moreover, and crucially, for a positive density of scales (weighted according to the measure), the Frostman exponent is bounded away from 0 and 1. This last claim fails if μ is the uniform measure on a square, and the assumptions of the theorem are meant precisely to avoid this counterexample. Moreover, the scales B_j can be chosen to satisfy $B_{j+1} \leq 2B_j$, which is critical for linearization arguments.

Theorem 4.1. *For every $u > 0$ and $\varepsilon > 0$ there are $\xi = \xi(u) > 0$ and $\tau = \tau(\varepsilon) > 0$ such that the following holds for all sufficiently large $T \geq T_0(\varepsilon)$ and $\ell \geq \ell_0(T, \varepsilon)$.*

Let μ be a $(\sigma; T)$-regular measure on $[0,1)^d$ with (dyadic) support X, and write $m = \ell T$. Suppose

$$\mu(B(x, |X|^{1/d})) \leq 2^{-um} \quad \text{for all } x. \tag{4.1}$$

Then there are a collection $\{[A_j, B_j)\}$ of pairwise disjoint intervals contained in $[0, \ell)$ and numbers $\alpha_j \in [0, d]$ such that the following hold:

(i) $\tau\ell \leq B_j - A_j \leq A_j$ *for all* j.

(ii) *Write* $m_j = T(B_j - A_j)$. *For each* $Q \in \mathcal{D}_{TA_j}$,

$$\mu^Q(B(x,r)) \leq 2^{\varepsilon m_j} r^{\alpha_j} \quad \text{for all } r \in [2^{-m_j}, 1].$$

(iii) $\sum_j \alpha_j m_j \geq \log \mathcal{N}(X,m) - 2\varepsilon m$.

(iv) $\sum\{m_j : \alpha_j \in [\xi, d-\xi]\} \geq \xi m$.

Theorem 4.1 will be proved in the rest of the section. Following [18], the problem is translated into one about Lipschitz functions on the line, but both the statement of the problem and the solution differ substantially from [18].

4.2. Decompositions of Lipschitz functions into almost linear/superlinear pieces

In this section we deal with the following kind of problem: given a Lipschitz function $f : [a,b] \to \mathbb{R}$, we aim to find non-overlapping intervals I_j such that $f|_{I_j}$ is close to linear/bounded below by a linear function, and the union of the I_j exhausts most of the original interval $[a,b]$. We start by making some of these concepts precise.

Definition 4.2. Given a function $f : [a,b] \to \mathbb{R}$, we let

$$s_f(a,b) = \frac{f(b) - f(a)}{b-a}$$

be the slope of the linear function that agrees with f on a and b. We also write

$$L_{f,a,b}(x) = f(a) + s_f(a,b)(x-a)$$

for the linear function that agrees with f at a and b. We say that (f,a,b) *is* ε*-linear* if

$$|f(x) - L_{f,a,b}(x)| \leq \varepsilon |b-a| \quad \text{for all } x \in [a,b].$$

Likewise, we say that (f,a,b) is ε*-superlinear* if

$$f(x) \geq L_{f,a,b}(x) - \varepsilon |b-a| \quad \text{for all } x \in [a,b].$$

Sometimes we say that f is linear/superlinear on $[a,b]$ to mean that (f,a,b) is linear/superlinear.

The following is our basic lemma for finding intervals on which f is ε-linear. A small variant of the lemma was posed as a problem on the 20-th Annual Vojtěch Jarník International Mathematical Competition; I thank K. Héra for pointing this out. The proof below is repeated from [25].

Lemma 4.3. *For every* $\varepsilon > 0$ *there is* $\delta = \delta(\varepsilon) > 0$ *such that the following holds: for any* 1*-Lipschitz function* $f : [a,b] \to \mathbb{R}$ *there exists a subinterval* $[c,d] \subset [a,b]$ *with* $d - c \geq \delta(b-a)$ *such that* (f,c,d) *is* ε*-linear. In fact,* $\delta = \varepsilon^{\lfloor 1/\varepsilon \rfloor}$ *works.*

Proof. By replacing f with $-f$ if needed, we may assume that $f(b) \geq f(a)$. We claim that if (f, a, b) is *not* ε-linear, then there exists an interval $[a', b'] \subset [a, b]$ with $b' - a' \geq \varepsilon(b-a)$ such that $s_f(a', b') \geq s_f(a, b) + \varepsilon$. Suppose, then, that (f, a, b) is not ε-linear, which by definition means that there is $x \in [a, b]$ such that

$$|f(x) - L_{f,a,b}(x)| > \varepsilon(b-a).$$

Replacing, if needed, f by the flip $\widetilde{f}(x) = -f(a+b-x)$ and x by $a+b-x$, we may assume that $f(x) - L_{f,a,b}(x) > \varepsilon(b-a)$. We have

$$s_f(a, x) = \frac{f(x) - f(a)}{x - a} \geq \frac{s_f(a,b)(x-a) + \varepsilon(b-a)}{x-a} \geq s_f(a,b) + \varepsilon.$$

On the other hand,

$$x - a \geq f(x) - f(a) \geq s_f(a,b)(x-a) + \varepsilon(b-a) \geq \varepsilon(b-a),$$

so $[a, x]$ is the claimed interval.

Now let $[a_0, b_0] = [a, b]$, and inductively apply the claim and set $[a_{j+1}, b_{j+1}] = [a'_j, b'_j]$ so long as (f, a_j, b_j) is not ε-linear. Since $s_f(a_0, b_0) \geq 0$ and $s_f(a_j, b_j) \leq 1$, the process must stop in $j \leq \lfloor 1/\varepsilon \rfloor$ steps. Then $[a_j, b_j]$ is the desired interval. ∎

By iterating the above lemma, we can cover most of $[a, b]$ by intervals on which f is ε-linear. We use $\mathcal{L}$ to denote Lebesgue measure.

Lemma 4.4. *For every $\varepsilon > 0$ there is $\tau > 0$ such that the following holds. For any 1-Lipschitz function $f : [a, b] \to \mathbb{R}$ there exists a family $\{[c_j, d_j]\}_{j=1}^{M}$ of non-overlapping intervals such that:*

(i) *(f, c_j, d_j) is ε-linear for all j.*

(ii) *$d_j - c_j \geq \tau(b-a)$ for all j.*

(iii) *$\mathcal{L}([a, b] \setminus \bigcup_j [c_j, d_j]) \leq \varepsilon(b-a)$.*

Proof. Apply Lemma 4.3 to the interval $[a, b]$ to obtain an interval $[x_1, y_1] \subset [a, b]$. Next, apply Lemma 4.3 to the intervals $[a, x_1]$ and $[y_1, b]$ to obtain intervals $[x_2, y_2]$ and $[x_3, y_3]$ (we allow for degenerate intervals). Continue inductively. In each step, a proportion at least $\delta = \delta(\varepsilon)$ is removed from the length of the set $[a, b] \setminus \bigcup_{k=1}^{2^\ell - 1} [x_k, y_k]$, so after a number $N(\varepsilon)$ of steps the remaining length is at most $(\varepsilon/2)|b-a|$. The total number of intervals is $2^{N(\varepsilon)} - 1$. Hence, if $\tau = \varepsilon 2^{-(N(\varepsilon)+1)}$, the intervals of length $< \tau|b-a|$ contribute length at most $(\varepsilon/2)|b-a|$. Removing them from the collection of all $[x_k, y_k]$ we obtain the desired collection of intervals $[c_j, d_j]$. ∎

Corollary 4.5. *For every $\varepsilon > 0$ there is $\tau = \tau(\varepsilon) > 0$ such that the following holds. For any 1-Lipschitz function $f : [a, b] \to \mathbb{R}$ there exists a family of non-overlapping intervals $\{[c_j, d_j]\}_j$ such that:*

(i) *(f, c_j, d_j) is ε-linear for all j.*

(ii) *$\tau(b-a) \leq d_j - c_j \leq c_j$ for all j.*

(iii) *$\mathcal{L}([a, b] \setminus \bigcup_j [c_j, d_j]) \leq \varepsilon|b-a|$.*

Proof. Let $\tau = \tau(\varepsilon)$ be the number given by Lemma 4.4. Let k_0 be the largest (negative) integer such that $2^{k_0+1} \geq \varepsilon$. Apply Lemma 4.4 to the intervals

$$I_k = [a + 2^k(b-a), a + 2^{k+1}(b-a)], \quad k = k_0, \dots, -2, -1,$$

and collect all resulting intervals. It is easy to check that the conclusion holds with 2ε in place of ε and $\varepsilon\tau$ in place of τ, which is a formally equivalent statement. ∎

The next lemma is similar to Lemma 4.4, but we deduce the additional information that the slopes of f on the subintervals are increasing; the price to pay is that f becomes ε-superlinear on the subintervals (instead of ε-linear).

Lemma 4.6. *Given $\varepsilon > 0$ there is $\tau = \tau(\varepsilon) > 0$ such that the following holds. Let $f : [a,b] \to \mathbb{R}$ be a 1-Lipschitz function. Then there exists a collection $\{[a_k, b_k]\}_k$ of non-overlapping intervals such that $b_{k+1} \leq a_k$ and:*

(i) *(f, a_k, b_k) is ε-superlinear for all k.*

(ii) *$b_k - a_k \geq \tau(b-a)$ for all k.*

(iii) *$\mathcal{L}([a,b] \setminus \bigcup_k [a_k, b_k]) \leq \varepsilon|b-a|$.*

(iv) *The sequence $s_f(a_k, b_k)$ is increasing.*

Proof. The idea (once b_k is given) is to define a_k as the number $a < b_k$ that maximizes $s_f(a, b_k)$. Unfortunately, for a general Lipschitz function such a number may not exist and, more importantly, it can be too close to b_k. We deal with this issue by restricting a to the set of endpoints of intervals given by Lemma 4.4. We proceed to the details.

Without loss of generality, $f(a) = 0$. Let $\tau = \tau(\varepsilon^2/4)$ be the number given by Lemma 4.4 applied with $\varepsilon^2/4$ in place of ε, and let $\{[c_i, d_i]\}_i$ be the intervals given by the lemma. Write $E = [a,b] \setminus \bigcup_i (c_i, d_i)$, so that $\mathcal{L}(E) \leq \varepsilon^2(b-a)/4$. Let $\mathcal{C} = \{c_i\} \cup \{d_i\} \cup \{a\}$. For each $y \in (a,b]$ let $x = P(y)$ be the element of $\mathcal{C} \cap [a, y)$ that maximizes $s_f(x, y)$ (if there are several such elements, pick the largest one). Let $y_0 = b$ and so long as $y_j > a$ inductively set $y_{j+1} = P(y_j)$. By construction, the sequence $s_f(y_{j+1}, y_j)$ is increasing. Let

$$\mathcal{J}_1 = \{j : |[y_{j+1}, y_j] \cap E| \geq (\varepsilon/2)(y_j - y_{j+1})\},$$
$$\mathcal{J}_2 = \{j : y_j - y_{j+1} < \tau(b-a)\}.$$

We let $\{[a_k, b_k]\}$ be the collection $\{[y_{j+1}, y_j] : j \notin \mathcal{J}_1 \cup \mathcal{J}_2\}$, ordered so that $b_k \leq a_{k+1}$. The claims that $s_f(a_k, b_k)$ is increasing and $b_k - a_k \geq \tau(b-a)$ are clear. Also, since $\mathcal{L}(E) \leq \varepsilon^2(b-a)/4$,

$$\sum_{j \in \mathcal{J}_1} (y_j - y_{j+1}) \leq 2\mathcal{L}(E)/\varepsilon \leq (\varepsilon/2)(b-a).$$

Likewise, any interval $[y_{j+1}, y_j]$ with $y_j - y_{j+1} < \tau(b-a)$ is contained in E, and therefore

$$\sum_{j \in \mathcal{J}_2} (y_j - y_{j+1}) \leq \mathcal{L}(E) \leq (\varepsilon/2)(b-a).$$

It follows that

$$\mathcal{L}\Big([a,b] \setminus \bigcup_k [a_k, b_k]\Big) \le \varepsilon(b-a).$$

It remains to prove that if $j \notin \mathcal{J}_1 \cup \mathcal{J}_2$, then (f, y_{j+1}, y_j) is ε-superlinear. For simplicity, we write $y = y_{j+1}$, $y' = y_j$. By the definition of $\mathcal{J}_1$, the interval $[y, y']$ can be split into the union of some intervals $([c_i, d_i])_{i\in I}$ plus a remainder set of measure at most $(\varepsilon/2)(y'-y)$. Since $y = P(y')$, we know that

$$s_f(x, y') \le s_f(y, y') \quad (x \in \{c_i, d_i\},\ i \in I),$$

which in turn implies that

$$L_{f,c_i,d_i}(x) \ge L_{f,y,y'}(x) \quad (x \in [c_i, d_i],\ i \in I).$$

Since f is $(\varepsilon^2/4)$-linear on $[c_i, d_i]$, we deduce that, for $x \in [c_i, d_i]$,

$$f(x) \ge L_{f,c_i,d_i}(x) - (\varepsilon^2/4)(d_i - c_i) \ge L_{f,y,y'}(x) - (\varepsilon^2/4)(y'-y).$$

Finally, if $x \in [y, y'] \cap E$, we can find $x' = c_i$ or $x' = d_i$ with $|x' - x| \le (\varepsilon/2)(y'-y)$. Applying what we already know to x' and the 1-Lipschitz property of f, we conclude

$$f(x) \ge f(x') - (\varepsilon/2)(y'-y) \ge L_{f,y,y'}(x) - (\varepsilon^2/4 + \varepsilon/2)(y'-y),$$

completing the proof. ∎

We are now able to prove the main result of this section. The statement is similar to that of the previous lemmas, but the crucial new element is that the slopes of $f : [0,1] \to \mathbb{R}$ on many of the subintervals are bounded away from 0 and 1. Note that this cannot hold for the function f that has slope 0 on $[0, 1-s]$ and slope 1 on $[1-s, 1]$, where $s = f(1)$. In other words, we must have $f(1-s) > f(0)$. It turns out that this assumption is also sufficient, with the parameters measuring "many" and "bounded away" unsurprisingly depending on the difference $f(1-s) - f(0)$.

Proposition 4.7. *Given $s, t \in (0,1)$ there is $\xi = \xi(s,t) > 0$ such that the following holds for all sufficiently small $\varepsilon \le \varepsilon_1(s,t)$ and $\tau = \tau(\varepsilon) > 0$.*

Let $f : [0, B] \to \mathbb{R}$ be a non-decreasing, 1-Lipschitz function with

$$f(0) = 0, \quad f((1-s)B) \ge tB, \quad f(B) = sB.$$

Then there is a collection $\{[a_j, b_j]\}_j$ of intervals such that $b_{j+1} \le a_j$ and:

(i) *(f, a_j, b_j) is ε-superlinear for all j.*

(ii) *$\tau B \le b_j - a_j \le a_j$ for all j.*

(iii) *$\mathcal{L}([0,B] \setminus \bigcup_j [a_j, b_j]) \le \varepsilon B$.*

(iv) *$\sum\{b_j - a_j : s_f(a_j, b_j) \in [\xi, 1-\xi]\} \ge \xi B$.*

Moreover, the values of ξ and ε_1 can be chosen to be uniform over s, t varying in any compact subset of $(0,1)$.

Proof. Replacing $f(x)$ by $g(x) = f(Bx)/B : [0,1] \to \mathbb{R}$ we may and do assume that $B = 1$.

Let $\sigma = \sigma(s,t) > 0$ be a small enough number to be chosen later. We split $[0,1]$ as a union of non-overlapping intervals $\{I_n = [c_n, c_{n+1}]\}_{n=0}^{N}$ with c_n increasing such that:

(1) $|I_0| = 4\sigma$.

(2) For each $n \in [1,N]$, $\sigma \le |I_n| \le 2\sigma$.

(3) There is $n_0 \in [1,N]$ such that $1 - s = c_{n_0}$.

This can be easily arranged if σ is small enough in terms of s. For example, we can take $I_0 = [0, 4\sigma]$, and then split each of $[4\sigma, 1-s]$ and $[1-s, 1]$ into intervals of length equal to or slightly larger than σ. We note the following consequence of (1) and (2) that will be used later: for each $n \in [1,N]$,

$$|I_n| + |I_{n+1}| \le c_1 \le c_n.$$

Fix ε_0; we will later require it to be small enough in terms of σ. We will eventually choose $\varepsilon = 4\varepsilon_0/\sigma$. Let $\tau_0 = \tau_0(\varepsilon_0)$ be the smaller of the values of τ arising from Corollary 4.5 and Lemma 4.6. We apply Corollary 4.5 to the interval I_0, and Lemma 4.6 to the intervals I_n for each $n \in [1,N]$, in both cases with ε_0 in place of ε, to obtain intervals $\{[a_{n,k}, b_{n,k}]\}_k$. We write $s_{n,k} = s_f(a_{n,k}, b_{n,k})$ for simplicity (the function f is fixed throughout the proof).

We let $\zeta \in (0, \sigma)$ be a small parameter that will ultimately be chosen small enough in terms of σ (hence in terms of s, t only); in fact, $\zeta = \sigma/11$ works. We subdivide the indices $n \in [1,N]$ into various disjoint classes:

$$\mathcal{I}_1 = \Big\{ n \in [1,N] : \sum_k \{ b_{n,k} - a_{n,k} : s_{n,k} \le \zeta \} \ge (1-\sigma)|I_n| \Big\},$$

$$\mathcal{I}_2 = \Big\{ n \in [1,N] : \sum_k \{ b_{n,k} - a_{n,k} : s_{n,k} \ge 1-\zeta \} \ge (1-\sigma)|I_n| \Big\},$$

$$\mathcal{I}_3 = \Big\{ n \in [1,N] \setminus (\mathcal{I}_1 \cup \mathcal{I}_2) : \sum_k \{ b_{n,k} - a_{n,k} : s_{n,k} \notin [\zeta, 1-\zeta] \} \ge (1-\sigma/2)|I_n| \Big\},$$

$$\mathcal{I}_4 = [1,N] \setminus (\mathcal{I}_1 \cup \mathcal{I}_2 \cup \mathcal{I}_3).$$

Roughly speaking, if $n \in \mathcal{I}_1$ then $f(c_{n+1}) \approx f(c_n)$ (so that f must be roughly constant on $\mathcal{I}_n$); if $n \in \mathcal{I}_2$, then f is close to a linear function with slope 1 on I_n; if $n \in \mathcal{I}_3$, then f is close to a piecewise linear function that has slope 0 in an initial interval $[c_n, z_n]$ and then slope 1 on $[z_n, c_{n+1}]$ (with z_n not too close to either c_n or c_{n+1}).

Assume first that $\mathcal{I}_4 \ne \emptyset$. Note that, for any $n \in [1,N]$,

$$\sum_k b_{n,k} - a_{n,k} \ge (1-\varepsilon_0)|I_n| \ge (1-\sigma/4)|I_n|,$$

provided $\varepsilon_0 \le \sigma/4$. If $n \in \mathcal{I}_4$, then

$$\sum_k \{ b_{n,k} - a_{n,k} : s_{n,k} \notin [\zeta, 1-\zeta] \} \le (1-\sigma/2)|I_n|,$$

and hence

$$\sum\{b_{n,k} - a_{n,k} : s_{n,k} \in [\zeta, 1-\zeta]\} \geq \frac{\sigma}{4}|I_n| \geq \frac{\sigma^2}{4}.$$

On the other hand, for each $n \in [0, N]$,

$$b_{n,k} - a_{n,k} \leq |I_n| \leq c_n \leq a_{n,k}, \quad b_{n,k} - a_{n,k} \geq \tau_0 |I_n| \geq \sigma\tau_0,$$

so the collection $\{[a_{n,k}, b_{n,k}]\}_{n,k}$ already works (we take $\varepsilon = \varepsilon_0$ in this case). We hence assume that $\mathcal{I}_4$ is empty from now on.

Claim 1. *There is $n \in [n_0, N]$ such that $n \notin \mathcal{I}_2$, provided σ is chosen small enough in terms of s, t; in fact, $\sigma < 1 - \sqrt{1 - t/s}$ is enough (any value of σ works if $t/s > 1$).*

Indeed, suppose $n \in \mathcal{I}_2$ for all $n \in [n_0, N]$. Since

$$f(b_{n,k}) = f(a_{n,k}) + s_{n,k}(b_{n,k} - a_{n,k}),$$

and f is non-decreasing, it easily follows from the definition of $\mathcal{I}_2$ and the inequality $\zeta \leq \sigma$ that

$$f(c_{n+1}) \geq f(c_n) + (1-\sigma)^2(c_{n+1} - c_n)$$

for all $n \in \mathcal{I}_2$. Adding from n_0 to N, and using $f(c_{n_0}) = f(1-s) \geq t$, we get

$$s = f(1) \geq t + (1-\sigma)^2 \cdot s,$$

which contradicts the choice of σ.

Claim 2. *There is $n \in [1, n_0]$ such that $n \notin \mathcal{I}_1$, provided $\sigma < t/6$.*

To see this, assume on the contrary that $n \in \mathcal{I}_1$ for all $n \in [0, n_0]$ and note that since f is 1-Lipschitz and $\zeta \leq \sigma$,

$$\begin{aligned} f(c_1) &\leq |I_0| = 4\sigma, \\ f(c_{n+1}) &\leq f(c_n) + 2\sigma|I_n| \quad \text{for all } n \in [1, n_0], \end{aligned}$$

and hence, telescoping,

$$f(1-s) = f(c_{n_0}) \leq f(c_1) + 2\sigma \leq 6\sigma.$$

This contradicts the assumption $f(1-s) > t$ if $\sigma < t/6$.

In conclusion, since we are assuming that $\mathcal{I}_4 = \emptyset$, there must exist $n \in [2, N]$ such that $n - 1 \in \mathcal{I}_2 \cup \mathcal{I}_3$ and $n \in \mathcal{I}_1 \cup \mathcal{I}_3$. We work with this fixed value of n for the rest of the proof. Unpacking the definitions, this implies that

$$\begin{aligned} \sum\{b_{n-1,k} - a_{n-1,k} : s_{n-1,k} \geq 1-\zeta\} &\geq \frac{\sigma}{2}|I_{n-1}|, \\ \sum\{b_{n,k} - a_{n,k} : s_{n,k} \leq \zeta\} &\geq \frac{\sigma}{2}|I_n|. \end{aligned}$$

Let

$$k' = \min\{k : s_{n-1,k} \geq 1 - \zeta\}.$$

Write $\widetilde{a} = a_{n-1,k'}$, and note that $\widetilde{a} \le c_n - \frac{\sigma}{2}|I_{n-1}|$. Recall that $s_{n-1,k}$ is increasing in k, f is ε_0-superlinear on $[a_{n-1,k}, b_{n-1,k}]$, and $\sum_k b_{n-1,k} - a_{n-1,k} \ge (1-\varepsilon_0)|I_{n-1}|$. It then follows that

$$f(x) \ge f(\widetilde{a}) + (1-\zeta)(x-\widetilde{a}) - 2\varepsilon_0|I_{n-1}| \quad \text{for all } x \in [\widetilde{a}, c_n]. \tag{4.2}$$

Likewise, if we set

$$k'' = \max\{k : s_{n,k} \le \zeta\},$$

and write $\widetilde{b} = b_{n,k''}$, then $\widetilde{b} \ge c_n + \frac{\sigma}{2}|I_n|$ and

$$f(c_n) \le f(x) \le f(\widetilde{b}) \le f(x) + \zeta(\widetilde{b}-x) + \varepsilon_0|I_n| \quad \text{for all } x \in [c_n, \widetilde{b}]. \tag{4.3}$$

We will show that

$$\zeta \le s_f(\widetilde{a}, \widetilde{b}) \le 1-\zeta. \tag{4.4}$$

provided $\zeta \le \sigma/11$ and $\varepsilon_0 \le \zeta$. Indeed,

$$\begin{aligned} s_f(\widetilde{a}, \widetilde{b}) &\le \frac{(c_n - \widetilde{a}) + (\zeta + \varepsilon_0)|I_n|}{\widetilde{b} - \widetilde{a}} && ((4.3) \text{ for } x = c_n) \\ &\le 1 - \frac{|I_n|(\sigma/2 - 2\zeta)}{\widetilde{b} - \widetilde{a}} && (\widetilde{b} \ge c_n + \tfrac{\sigma}{2}|I_n|) \\ &\le 1 - \tfrac{1}{3}(\sigma/2 - 2\zeta) \le 1 - \zeta && ((2), \zeta \le \sigma/10). \end{aligned}$$

On the other hand,

$$\begin{aligned} s_f(\widetilde{a}, \widetilde{b}) &\ge \frac{f(c_n) - f(\widetilde{a})}{\widetilde{b} - \widetilde{a}} && (f \text{ non-decreasing}) \\ &\ge \frac{(c_n - \widetilde{a})(1-\zeta) - 2\varepsilon_0|I_{n-1}|}{\widetilde{b} - \widetilde{a}} && ((4.2) \text{ for } x = c_n) \\ &\ge \frac{(\sigma/2)|I_{n-1}|(1-\zeta) - 2\varepsilon_0|I_{n-1}|}{|I_{n-1}| + |I_n|} && (\widetilde{a} \le c_n - \tfrac{\sigma}{2}|I_{n-1}|) \\ &\ge \tfrac{1}{3}(\sigma/2(1-\zeta) - 2\zeta) \ge \zeta && ((2),\ \zeta \le \sigma/11,\ \sigma \le 1). \end{aligned}$$

We have verified that (4.4) holds. We will now use this to show that $(f, \widetilde{a}, \widetilde{b})$ is ε-superlinear, where we define $\varepsilon = 4\varepsilon_0/\sigma$. It follows from (4.2) and the right-hand side inequality in (4.4) that, for x in the interval $[\widetilde{a}, c_n]$,

$$f(x) \ge L_{f,\widetilde{a},\widetilde{b}}(x) - 2\varepsilon_0|I_{n-1}|.$$

Now, for $x \in [c_n, \widetilde{b}]$ we know from (4.3) that

$$\begin{aligned} f(x) &\ge f(\widetilde{b}) - \zeta(\widetilde{b}-x) - \varepsilon_0|I_n| \\ &= f(\widetilde{a}) + s_f(\widetilde{a}, \widetilde{b})(\widetilde{b}-\widetilde{a}) - \zeta(\widetilde{b}-x) - \varepsilon_0|I_n| \\ &\ge f(\widetilde{a}) + s_f(\widetilde{a}, \widetilde{b})(x-\widetilde{a}) - \varepsilon_0|I_n| = L_{f,\widetilde{a},\widetilde{b}}(x) - \varepsilon_0|I_n|, \end{aligned}$$

using $s_f(\widetilde{a},\widetilde{b}) \geq \zeta$ in the last line. Since $\max(|I_{n-1}|,|I_n|) \leq (\sigma/2)^{-1}(\widetilde{b}-\widetilde{a})$, we have shown that $(f,\widetilde{a},\widetilde{b})$ is $(4\varepsilon_0/\sigma)$-superlinear, as claimed.

Now the collection consisting of $[\widetilde{a},\widetilde{b}]$ together with all the intervals $\{[a_{\widetilde{n},k},b_{\widetilde{n},k}]\}_{\widetilde{n}=0}^{N}$ disjoint from $[\widetilde{a},\widetilde{b}]$ is easily seen to satisfy the claims in the proposition: the intervals $[a_{\widetilde{n},k},b_{\widetilde{n},k}]$ satisfy (i)–(ii) by the same reasoning as in the case $\mathcal{I}_4 \neq \emptyset$, while we have verified that $(f,\widetilde{a},\widetilde{b})$ is ε-superlinear,

$$\sigma^2/2 \leq \sigma|I_{n-1}|/2 \leq |\widetilde{b}-\widetilde{a}| \leq |I_{n-1}| + |I_n| \leq c_{n-1} \leq \widetilde{a},$$

and (iv) follows with $\xi = \min(\sigma^2/2,\zeta)$ from $|\widetilde{b}-\widetilde{a}| \geq \sigma^2/2$ and (4.4). Finally, (iii) also follows since the total length of the intervals is at least the sum of the lengths of $[a_{\widetilde{n},k},b_{\widetilde{n},k}]$, which is at least $1-\varepsilon_0 \geq 1-\varepsilon$ by construction.

Finally, it is clear that all the parameters appearing in the proof can be taken to work in a neighborhood of s and t, and this completes the proof of the proposition. ∎

4.3. The proof of Theorem 4.1, and a variant

We now convert the combinatorial decompositions from the previous section into suitable multiscale decompositions of regular measures, in particular proving Theorem 4.1. We start with a variant of Proposition 4.7 in which the endpoints of the intervals $[a_j,b_j]$ lie on a lattice.

Corollary 4.8. *Let $f:[0,B]\to\mathbb{R}$ be a 1-Lipschitz function satisfying the assumptions of Proposition 4.7. Then in Proposition 4.7 we may further require that the numbers a_j and b_j lie in $(B/\ell)\mathbb{N}_0$ provided $\ell \geq \ell_0(\varepsilon)$ is sufficiently large (after changing the values of τ and ξ slightly).*

Proof. We may assume $B=1$. Let $\{[a_j,b_j]\}$ be the intervals given by Proposition 4.7, and set

$$\widetilde{a}_j = \lceil \ell a_j \rceil/\ell, \quad \widetilde{b}_j = \lfloor \ell a_j \rfloor/\ell.$$

Then $a_j \leq \widetilde{a}_j \leq a_j + 1/\ell$ and $b_j - 1/\ell \leq \widetilde{b}_j \leq b_j$. Since f is 1-Lipschitz and non-decreasing, the same inequalities are preserved when applying f. If ℓ is taken large enough in terms of τ (hence in terms of ε), then $\widetilde{b}_j - \widetilde{a}_j \geq b_j - a_j - \tau^2$ (say) so that (ii) and (iii) continue to hold with $\tau/2$ in place of τ. Likewise, a short calculation shows that if ℓ is large enough then

$$|s_{f,\widetilde{a}_j,\widetilde{b}_j} - s_{f,a_j,b_j}| \leq \varepsilon,$$

and if (f,a_j,b_j) is ε-superlinear, then $(f,\widetilde{a}_j,\widetilde{b}_j)$ is (2ε)-superlinear. Using these facts it is easy to check that (i) and (iv) continue to hold for the intervals $[\widetilde{a}_j,\widetilde{b}_j]$, with 2ε in place of ε and, say, $\xi/2$ in place of ξ. ∎

We can now complete the proof of Theorem 4.1.

Proof of Theorem 4.1. Let

$$\xi = \inf_{s\in[u/4,1-u/2]} \xi(s,u/(2d)) > 0, \quad \varepsilon_1 = \inf_{s\in[u/4,1-u/2]} \varepsilon_1(s,u/(2d)) > 0$$

be the values given by Proposition 4.7, and note that they depend only on u and d. Fix $\varepsilon \in (0,\varepsilon_1)$, and let $\tau(\varepsilon), \ell_0(\varepsilon)$ be the numbers given by Proposition 4.7. Take $\ell \geq \ell_0(\varepsilon)$.

Let $f : [0,\ell] \to [0,\ell]$ be the function such that $f(0) = 0$,

$$f(j) = \frac{1}{d}(\sigma_1 + \cdots + \sigma_j) \quad (j = 1,\ldots,\ell),$$

and interpolates linearly between j and $j+1$. Since $\sigma_i \in [0,d]$, the function f is 1-Lipschitz and non-decreasing. Let $\{[A_j,B_j]\}_{j=1}^M$ be the intervals provided by Proposition 4.7. In light of Corollary 4.8, we may and do assume that $A_j, B_j \in \mathbb{N}_0$.

Let

$$s = \frac{f(\ell)}{\ell} = \frac{\sigma_1 + \cdots + \sigma_\ell}{\ell}.$$

Since $|X| \geq 2^{-dm}$, it follows from (4.1) that $\mu(Q) \lesssim 2^{-um}$ for all $Q \in \mathcal{D}_m$, and therefore $\mathcal{N}(X,m) \gtrsim 2^{um}$. On the other hand, we may cover $[0,1]^d$ by $\lesssim |X|^{-1}$ balls of radius $|X|^{1/d}$, whence $1 \lesssim |X|^{-1}2^{-um}$. Hence, if T is large enough in terms of u, we may assume

$$2^{(u/2-d)m} \leq |X| \leq 2^{-um/2}. \tag{4.5}$$

Now, from this and Lemma 2.2 (ii), we deduce that, provided $1/T \leq \varepsilon \leq u/2$,

$$1 - s - \varepsilon \leq \frac{\log|X|}{-md} \leq 1 - s \implies s \in [u/4, 1-u/2].$$

Let $j = \lfloor (1-s)\ell \rfloor$, and note that any cube Q of side length 2^{-jT} can be covered by $C_{d,T} = O_{d,T}(1)$ balls of radius $2^{-(1-s)m} \leq |X|^{1/d}$. It follows from our assumption (4.1) that

$$\log\mu(Q) \leq \log C_{d,T} - um \leq (\varepsilon - u)m,$$

provided ℓ is large enough in terms of ε, d, T. On the other hand, we see from Lemma 2.2 (i) that for all 2^{-jT}-dyadic cubes Q hitting X, we have the lower bound

$$\log\mu(Q) \geq -\varepsilon m - (m/\ell)(\sigma_1 + \cdots + \sigma_{\lfloor(1-s)\ell\rfloor}),$$

provided T is large enough in terms of ε. Comparing the upper and lower bounds on $\log\mu(Q)$, we conclude

$$f((1-s)\ell) \geq \frac{1}{d}(\sigma_1 + \cdots + \sigma_{\lfloor(1-s)\ell\rfloor}) \geq \frac{1}{d}(u-2\varepsilon)\ell \geq \frac{u\ell}{2d},$$

provided ε is small enough in terms of u. We have therefore checked that the hypotheses of Proposition 4.7 are satisfied with our choice of parameters.

Claim (i) is clear. Let $\alpha_j = ds_f(A_j,B_j)$. By the definition of f,

$$s_f(A_j,B_j) = \frac{\sigma_{A_j+1} + \cdots + \sigma_{B_j}}{d(B_j - A_j)}.$$

Since (f, A_j, B_j) is ε-superlinear,

$$d(f(A_j + k) - f(A_j)) = \sigma_{A_j+1} + \cdots + \sigma_{A_j+k} \geq k\alpha_j - d\varepsilon(B_j - A_j)$$

for any $k \in [0, B_j - A_j]$. Fix $Q \in \mathcal{D}_{TA_j}$. It follows from Lemma 2.2 (i) that

$$\mu^Q(R) \leq 2^{d\varepsilon T(B_j - A_j)} 2^{-kT\alpha_j}$$

for all $R \in \mathcal{D}_{kT}(\operatorname{supp}_{\mathrm{d}}(\mu^Q))$. Any ball $B = B(x, r)$ with $x \in \operatorname{supp}(\mu^Q)$, $r \in [2^{-m_j}, 1]$ can be covered by $O_T(1)$ 2^T-adic cubes of side length $2^{-kT} \leq r$. Therefore, provided ℓ is taken large enough in terms of ε, T,

$$\mu^Q(B(x, r)) \leq 2^{2d\varepsilon m_j} r^{\alpha_j}$$

for all $r \in [2^{-m_j}, 1]$ and all $x \in \operatorname{supp}(\mu^Q)$. This gives (ii), with $2d\varepsilon$ in place of ε.

For the third claim, write $[0, \ell] \setminus \bigcup_j [A_j, B_j] = \bigcup_i [C_i, D_i]$ with the intervals $[C_i, D_i]$ non-overlapping. We know from Corollary 4.5 that $\sum_i (D_i - C_i) \leq \varepsilon\ell$, and $s_f(C_i, D_i) \leq 1$ simply because f is 1-Lipschitz. On the other hand, telescoping,

$$\ell s_f(0, \ell) = \sum_j (B_j - A_j) s_f(A_j, B_j) + \sum_i (D_i - C_i) s_f(C_i, D_i).$$

Recalling that $\alpha_j = d s_f(A_j, B_j)$, we deduce that

$$\sum_j \alpha_j m_j \geq \left(\frac{1}{\ell}(\sigma_1 + \cdots + \sigma_\ell) - \varepsilon \right) m.$$

Applying Lemma 2.2 (i) and taking $T \geq 1/\epsilon$, we obtain (iii).

Finally, from the last part of Proposition 4.7 we obtain

$$\sum \{m_j : \alpha_j \in [\xi, d - \xi]\} \geq T \sum \{B_j - A_j : s_f(A_j, B_j) \in [\xi/d, 1 - \xi/d]\} \geq \xi m.$$

Hence (iv) holds, and this completes the proof. ∎

In a very similar (but simpler) way, we obtain the following:

Theorem 4.9. *For every $\varepsilon > 0$ there is $\tau = \tau(\varepsilon) > 0$ such that the following holds for all sufficiently large $T \geq T_0(\varepsilon)$ and for all large enough $\ell \geq \ell_0(T, \varepsilon)$.*

Let μ be a $((\sigma_1, \ldots, \sigma_\ell); T)$-regular measure on $[0, 1)^d$ with support X, and write $m = \ell T$. Then there are a collection $\{[A_j, B_{j+1})\}$ of pairwise disjoint intervals contained in $[0, \ell)$ and numbers α_j such that the following hold:

(i) $\tau\ell \leq B_j - A_j \leq A_j$ *for all* j.

(ii) *Write $m_j = T(B_j - A_j)$. For each $Q \in \mathcal{D}_{TA_j}$,*

$$2^{-\varepsilon m_j} r^{\alpha_j} \leq \mu^Q(B(x, r)) \leq 2^{\varepsilon m_j} r^{\alpha_j} \quad \text{for all } r \in [2^{-m_j}, 1], x \in \operatorname{supp}(\mu^Q).$$

(iii) $\sum_j \alpha_j m_j \geq (\frac{\sigma_1 + \cdots + \sigma_\ell}{\ell} - \varepsilon) m$.

The proof is the same as that of Theorem 4.1, except that we rely on Corollary 4.5 (providing a decomposition into ε-linear, rather than superlinear, pieces) instead of Proposition 4.7.

5. Proof of main results

5.1. A multiscale lower bound for entropy

In this section we prove Theorems 1.3 and 1.7. The proofs rely on a multiscale bound for the entropy of (non-linear) projections that goes back in various forms to [12, 18, 23, 27]. We denote the orthogonal projection onto a plane $V \in \mathbb{G}(d,k)$ by P_V.

Proposition 5.1. *Fix $1 \le k < d$. Let $\mu, \nu \in \mathcal{P}([0,1)^d)$ and let $[A_i, B_i)_{i=1}^q$ be disjoint subintervals of $(0,m]$ such that $B_i \le 2A_i$. Let $F : U \to \mathbb{R}^k$ be a C^2 map defined in a neighborhood of $\operatorname{supp}(\mu)$ such that $DF(x)$ has full rank k for all $x \in \operatorname{supp}(\mu)$. Denote*

$$V(x) = \ker(DF(x))^{\perp} \in \mathbb{G}(d,k).$$

Assume furthermore that $\nu \le \Delta\mu$. Then

$$H_m(F\nu) \ge -O_{F,d}(q) + \sum_{i=0}^{q-1} \sum_{Q \in \mathcal{D}_{A_i}} \nu(Q) H^{m\Delta}_{B_i - A_i}(P_{V(x_Q)}\mu^Q),$$

where x_Q is an arbitrary point in Q. The constant implicit in $O_{F,d}(q)$ depends in a continuous manner on the C^2 norm of F.

Since this particular statement does not appear in the literature, but the proof is a minor modification of arguments in [12, 18, 23], we defer the proof to Appendix A.

Note that, in the case $k = 1$, $V(x)$ is the line generated by $\operatorname{dir} F'(x)$, and $P_{V(x)}(\cdot) = \langle \operatorname{dir} F'(x), \cdot \rangle$. This is the only case needed for the proofs of Theorems 1.3 and 1.7.

A key feature of Proposition 5.1 is that even though the map F is (possibly) non-linear, the lower bound on $H_m(F\nu)$ involves only linear projections. This is why we will ultimately be able to deduce statements about families of smooth maps from the (linear) projection theorems discussed in Section 3.

5.2. Proof of Theorem 1.7

The number κ and the family $\mathcal{F} = \{F_\lambda : \lambda \in \Lambda\}$ are given. The proof will involve a number of other small parameters, whose dependencies are as follows: $\varepsilon_1 = \varepsilon_1(\kappa)$, $\xi = \xi(\kappa)$, $\eta = \eta(\kappa,\xi)$, $\eta' = \eta'(\eta,\xi)$, $\delta_0 = \delta_0(\kappa)$, $\zeta = \zeta(\kappa,\xi,\eta,\varepsilon_1,\delta_0)$, $\tau = \tau(\zeta)$. All the parameters are also allowed to depend on the ambient dimension d.

We will show that the claim holds with $\eta'/2$ in place of η, provided $\varepsilon = \varepsilon(\kappa,\zeta,\tau,\eta')$ is taken small enough, and $m = T\ell$, where $T = T(\varepsilon,\zeta)$, $\ell = \ell(\zeta, T, \mathcal{F})$ are sufficiently large integers. If m is not of the form $T\ell$, we consider instead $m' = T\lfloor m/T \rfloor$.

Let

$$\mu = \mathbf{1}_X / |X|,$$

and let Y be the 2^{-m}-set given by Lemma 2.3 applied to μ. Taking T large enough in terms of ε, we may assume

$$|Y| \ge 2^{-(\varepsilon/2)m}|X|. \tag{5.1}$$

The non-concentration hypothesis (1.5) yields

$$\mu_Y(B(x,|Y|^{1/d})) \le \mu(Y)^{-1}\mu(B(x,|X|^{1/d})) \le 2^{(\varepsilon/2-\kappa)m} \le 2^{-\kappa m/2},$$

taking ε small enough in terms of κ. We can therefore apply Theorem 4.1 to the measure μ_Y, with $u = \kappa/2$. Let $\varepsilon_1 = \varepsilon_1(\kappa/2)$ and $\xi = \xi(\kappa/2)$ be the numbers given by the theorem. Pick $0 < \zeta < \varepsilon_1$, let $\tau = \tau(\zeta)$, and suppose that T is chosen large enough in terms of ζ, and ℓ is chosen large enough in terms of ζ and T, that the conclusions of Theorem 4.1 hold (with ζ in place of ε).

Let $\{[A_j, B_j)\}_j$ and $\{\alpha_j\}_j$ be the intervals and exponents obtained from Theorem 4.1 applied to μ_Y. Write

$$m_j = T(B_j - A_j).$$

Let $\eta = \eta(\kappa), \delta_0 = \delta_0(\kappa)$ be the numbers provided by Theorem 3.4. Our aim is to apply Theorem 3.4 to the measures μ_Y^Q, $Q \in \mathcal{D}_{TA_j}$ and $\rho_x := \theta_x \nu_{\Lambda_x}$, and the scales 2^{-m_j}. Since $m_j \ge \tau m$ by the first part of Theorem 4.1, the non-concentration assumption (1.6) on ρ_x implies that

$$\rho_x(H^{(r)}) \le 2^{(\varepsilon/\tau)m_j} r^\kappa \quad (H \in \mathbb{G}(d, d-1),\ r \in [2^{-m_j}, 1]). \tag{5.2}$$

Given $Q \in \mathcal{D}_{TA_j}$, we know from Theorem 4.1 (ii) that

$$\mu_Y^Q(B(x,r)) \le 2^{\zeta m_j} r^{\alpha_j} \quad (r \in [2^{-m_j}, 1]).$$

Thus, if ζ and then ε are chosen so that $\varepsilon/\tau \le \zeta < \delta_0(\kappa)$, the hypotheses of Theorem 3.4 are met with ζ in place of δ. Note that by making ℓ large enough in terms of all other parameters, we can make τm_j large so that the theorem is applicable.

Recall that given $x \in [0,1)^d$, the only element of $\mathcal{D}_{TA_j}$ containing x is denoted by $Q_{TA_j}(x)$. We apply Theorem 3.4 to the measures $\mu_Y^{Q_{TA_j}(x)}$ and ρ_x, to obtain sets $E(x,j)$ with

$$\rho_x(E(x,j)) \le 2^{-\zeta m_j} \le 2^{-\zeta\tau m}$$

such that

$$e \notin E(x,j) \implies P_e \mu_Y^{Q_{TA_j}(x)} \text{ is } (\gamma(\alpha_j), \zeta, m_j)\text{-robust},$$

where γ is the function from Theorem 3.4. Note that $\gamma(a) \ge a/d - 6\zeta$ for all $a \in [0,d]$. A little algebra shows that, provided $\eta < \kappa/(2d)$ (which we may assume) and ζ is small enough in terms of κ, then

$$\gamma(a) \ge a/d + \eta \quad \text{if } a \in [(\eta + 6\zeta)d/(d-1), d - d\eta - 6d\zeta].$$

Making η and ζ smaller in terms of ξ and d only, we may further assume that

$$\gamma(a) \ge \begin{cases} a/d - 6\zeta & \text{if } a \notin [\xi, d-\xi], \\ a/d + \eta & \text{if } a \in [\xi, d-\xi]. \end{cases} \tag{5.3}$$

Let $E(x) = \bigcup_j E(x, j)$. Then

$$\rho_x(E(x)) \leq (\ell/\tau)2^{-\xi\tau m} \leq 1/2, \tag{5.4}$$

assuming ℓ is large enough in terms of ζ. Recalling that $\rho_x = \theta_x \nu_{\Lambda_x}$ and that $\nu(\Lambda_x) \geq 2^{-\varepsilon m}$ by assumption, this shows that there exists a set $G_x \subset \Lambda$ such that $\nu(G_x) \geq 2^{-\varepsilon m}/2$ and $\theta_x(\lambda) \notin E(x)$ for all $\lambda \in G_x$. A standard argument shows that the set $\{(\lambda, x) : \theta_x(\lambda) \notin E(x)\}$ is Borel. By Fubini, we can find $\lambda \in \operatorname{supp}(\nu)$ and a set $Z \subset Y$ with

$$|Z| \geq \tfrac{1}{2}2^{-\varepsilon m}|Y| \geq 2^{-2\varepsilon m}|X|, \tag{5.5}$$

and such that $\lambda \in G_x$ for all $x \in Z$. We work with this value of λ for the rest of the proof, and note that

$$P_{\theta_x(\lambda)}\mu_Y^{Q_{TA_j}(x)} \text{ is } (\gamma(\alpha_j), \zeta, m_j)\text{-robust for all } j \text{ and } x \in Z. \tag{5.6}$$

On the other hand, we estimate

$$\begin{aligned}
\sum_j \gamma(\alpha_j)m_j &\geq \sum_{j:\alpha_j \notin [\xi, d-\xi]} (\alpha_j/d - 6\zeta)m_j + \sum_{j:\alpha_j \in [\xi, d-\xi]} (\alpha_j/d + \eta)m_j \quad \text{(Eq. (5.3))} \\
&\geq -6\zeta m + \frac{1}{d}\sum_j \alpha_j m_j + \eta \sum_{j:\alpha_j \in [\xi, d-\xi]} m_j \\
&\geq -6\zeta m + (\log \mathcal{N}(Y, m) - 2\zeta)/d + \eta\xi m \quad \text{(Thm. 4.1 (iii)–(iv))} \\
&\geq (\log \mathcal{N}(X, m))/d + (\eta\xi - 9\zeta)m \quad \text{(Eq. (5.1)}, \varepsilon \leq \zeta).
\end{aligned}$$

Hence, making ζ small enough in terms of ξ and η, and writing $\eta' = \xi\eta/2$, we get

$$\frac{1}{m}\sum_j \gamma(\alpha_j)m_j \geq \frac{\log \mathcal{N}(X, m)}{d\, m} + \eta'. \tag{5.7}$$

Note that the number q of intervals $[A_j, B_j]$ is at most $1/\tau$. For each $Q \in \mathcal{D}_{TA_j}$ such that $Q \cap Z \neq \emptyset$, pick some $x_Q \in Q \cap Z$. We are now ready to apply Proposition 5.1. Suppose $Z' \subset Z$, $|Z'| \geq 2^{-\varepsilon m}|Z|$. Recalling (5.5), we get $|Z'| \geq 2^{-2\varepsilon m - 1}|Y|$. Applying Proposition 5.1 to $\nu = \mu_{Z'}$, $\mu = \mu_Y$, $\Delta = 2^{2\varepsilon m+1}$, and the sequence $([TA_j, TB_j])_{j=1}^q$, we get

$$H_m(F_\lambda \mu_{Z'}) \geq -O_{\mathcal{F}}(1/\tau) + \sum_{j=1}^{q} \sum_{Q \in \mathcal{D}_{TA_j} : Q \cap Z' \neq \emptyset} \mu_{Z'}(Q) H_{m_j}^{m2^{2\varepsilon m+1}}(P_{\theta_{x_Q}(\lambda)}\mu_Y^Q). \tag{5.8}$$

Recall from (5.6) that $P_{\theta_{x_Q}(\lambda)}\mu_Y^Q$ is $(\gamma(\alpha_j), \zeta, m_j)$-robust. Lemma 2.8 then yields (provided ℓ is large enough in terms of ζ, ε and τ, which makes m_j large enough in terms of ζ, ε)

$$H_{m_j}^{2^{\zeta m_j/2}}(P_{\theta_{x_Q}(\lambda)}\mu_Y^Q) \geq (\gamma(\alpha_j) - \varepsilon)m_j.$$

If ε is small enough in terms of ζ of τ, then

$$2^{\zeta m_j/2} \geq 2^{\zeta\tau m/2} \geq m2^{2\varepsilon m+1}.$$

Recalling (5.7) and (5.8), we conclude that, provided m is large enough in terms of τ, ε and $\mathcal{F}$,

$$\begin{aligned} H_m(F_\lambda \mu_{Z'}) &\geq -\varepsilon m + \sum_{j=1}^{q} \sum_{Q\in\mathcal{D}_{TA_j}\colon Q\cap Z'\neq\emptyset} \mu_{Z'}(Q)(\gamma(\alpha_j)-\varepsilon)m_j \\ &\geq (\log \mathcal{N}(X,m))/d + (\eta'-2\varepsilon)m. \end{aligned}$$

Thus, assuming $\varepsilon < \eta'/4$,

$$\log \mathcal{N}(F_\lambda(Z'),m) \geq (\log \mathcal{N}(X,m))/d + \eta'/2,$$

giving the claim with $\eta'/2$ in place of η and Z in place of X'.

5.3. *The case $\Lambda_x = \Lambda$ for all x*

We indicate what changes are needed in the proof to obtain the stronger conclusions when $\Lambda_x = \Lambda$. As before, let $\mu = \mathbf{1}_X/|X|$. We need to apply Lemma 2.4 to μ (with the parameter 1.5ε in place of ε) instead of Lemma 2.3, to obtain the sets $(X_i)_i$. Taking T large enough in terms of ε, we have $\mu(X_i) \geq 2^{-2\varepsilon m}$. Other than the X_i having slightly smaller measure, depending on ε, the analysis we did for the set Y in the proof of Theorem 1.7 carries over to each of the X_i verbatim. In particular, the bound (5.4) holds for any X_i and $x \in X_i$. We take ε small enough in terms of ζ, τ so that

$$\rho_x(E(x)) = \nu\{\lambda : \theta_x(\lambda) \in E(x)\} \leq 2^{-5\varepsilon m}.$$

Write $\widetilde{X} = \bigcup_i X_i$. By Fubini, denoting Lebesgue measure by $\mathcal{L}$,

$$(\nu\times\mathcal{L})\{(\lambda,x) \in \Lambda\times\widetilde{X} : \theta_x(\lambda)\in E(x)\} \leq 2^{-5\varepsilon m}|\widetilde{X}|.$$

It follows that the set

$$\Lambda' = \{\lambda\in\Lambda : |x\in\widetilde{X} : \theta_x(\lambda)\in E(x)| \geq 2^{-4\varepsilon m}|\widetilde{X}|\}$$

has ν-measure at most $2^{-\varepsilon m}$.

Fix $\lambda \in \Lambda\setminus\Lambda'$ and $X'\subset X$ with $|X'|\geq 2^{-\varepsilon m}|X|$. By the first part of Lemma 2.4 (which we are applying with 1.5ε in place of ε), we can find i with

$$\frac{|X'\cap X_i|}{|X_i|} \geq \frac{|X'|}{|X|} - 2^{-1.5\varepsilon m} \geq 2^{-\varepsilon m}/2.$$

Since $\lambda\notin\Lambda'$,

$$|x\in\widetilde{X} : \theta_x(\lambda)\in E(x)| < 2^{-4\varepsilon m}|X| < \tfrac{1}{2}|X'\cap X_i|.$$

Let $Z = \{x \in X' \cap X_i : \theta_x(\lambda) \notin E(x)\} \subset X'$. We have seen that $\mu_{X_i}(Z) \geq 2^{-\varepsilon m}/4$. Now starting with (5.6), the argument from the previous section (with X_i in place of Y) yields the desired conclusion

$$\mathcal{N}(F_\lambda(X'), m) \geq \mathcal{N}(F_\lambda(Z), m) \geq 2^{m\eta'/2}\mathcal{N}(X, m)^{1/d}.$$

5.4. Proof of Theorem 1.3

Our next goal is to prove Theorem 1.3. We will in fact establish the following stronger fact.

Theorem 5.2. *Given $\kappa > 0, 0 < \alpha < d$ there is $\eta = \eta_d(\kappa, \alpha) > 0$ (that can be taken continuous in κ, α) such that the following holds.*

Let (Λ, ν) be a Borel probability space, let $U \subset \mathbb{R}^d$ be an open domain, and let μ be a Borel probability measure on U such that

$$\mu(B(x, r)) \leq C r^\alpha \quad (x \in U,\ r > 0).$$

Let $F : \Lambda \times U \to \mathbb{R}$ be a Borel function such that, for each $\lambda \in \Lambda$, the map $x \mapsto F_\lambda(x) = F(\lambda, x) : U \to \mathbb{R}$ is C^2 without singular points. For each $x \in \mathbb{R}^d$ define

$$\theta_x(\lambda) = \operatorname{dir}(\nabla F_\lambda(x)) \in S^{d-1}.$$

Suppose further that for μ-almost all x there are a set Λ_x with $\nu(\Lambda_x) \geq c$ and a number $C_x > 0$ satisfying the decay condition (1.4).

Then there are a constant $c' = c'(c) > 0$ and a set Λ_0 with $\nu(\Lambda_0) \geq c'$ such that the following holds. For all $\lambda \in \Lambda_0$ there is a set A_λ with $\mu(A_\lambda) \geq c'$ such that, for all sufficiently large m, the measure $F_\lambda(\mu_{A_\lambda})$ is $(\alpha/d + \eta, \eta, m)$-robust.

Moreover, $c' \to 1$ as $c \to 1$.

Note that Theorem 1.3 follows by considering a Frostman measure μ on A of exponent arbitrarily close to α, and applying Lemma 2.6. The proof of Theorem 5.2 is very similar to that of Theorem 1.7. We indicate the required changes.

By passing to a subset of nearly full μ-measure, we may assume that $C_x \leq C$ for all x. Denote $\rho_x = \theta_x \nu_{\Lambda_x}$. We have the following (stronger) analog of (5.2):

$$\rho_x(H^{(r)}) \leq C r^\kappa \quad (r \in (0, 1], H \in \mathbb{G}(d, d-1)). \tag{5.9}$$

Unlike the proof of Theorem 1.7, we need to consider all small scales at once. We can define the parameters $\zeta, \tau, \xi, \eta, \delta_0$ arising from Theorems 4.1 and 3.4; these are independent of the scale. Fix a small parameter ε, a large integer T and an even larger integer ℓ. We continue to write $m = T\ell$. As before, ε, T, ℓ can depend on all the previous parameters, T can depend on ε, and ℓ can depend on T and ε.

For each large ℓ, we apply Lemma 2.4 to μ, T and ε, to obtain a family of sets $(X_{\ell,i})_i$. For each i and $x \in \bigcup_i X_{\ell,i}$, we define the set $E(x)$ as in the proof of Theorem 1.7, but

now denote it by $E(x,\ell)$. For completeness, we define $E(x,\ell)=\emptyset$ if $x\notin\bigcup_i X_{\ell,i}$. As before, for each $x\in X$ we have

$$\rho_x(E(x,\ell))\le (m/\tau)2^{-\xi\tau m}.$$

Hence, provided ℓ_1 is large enough in terms of ζ, T, ε only, we can ensure $\rho_x(E(x))\le\varepsilon$ for all $x\in X$, where

$$E(x)=\bigcup_{\ell=\ell_1}^{\infty} E(x,\ell).$$

Unwrapping the definitions, this means that

$$\nu\{\lambda\in\Lambda_x:\theta_x(\lambda)\notin E(x)\}\ge\nu(\Lambda_x)(1-\varepsilon)\ge c(1-\varepsilon).$$

By Fubini, if we let $c'=1-\sqrt{1-c(1-\varepsilon)}$, there is a set Λ_0 with $\nu(\Lambda_0)\ge c'$ such that if $\lambda\in\Lambda_0$ then

$$\mu(Z_\lambda)\ge c',\quad \text{where}\quad Z_\lambda=\{x:\theta_x(\lambda)\notin E(x)\}. \tag{5.10}$$

We fix $\lambda\in\Lambda_0$ from now on and write $Z=Z_\lambda$.

Let $\varepsilon=\varepsilon(\kappa,\alpha)>0$, $\eta'=\eta'(\kappa,\alpha)>0$ be numbers to be determined. Suppose that $F_\lambda\mu_Z(A)\ge 2^{-\varepsilon m/2}$; our goal is to show that, provided m is large enough,

$$\log\mathcal{N}(A,m)\ge m(\alpha/d+\eta'). \tag{5.11}$$

It is enough to consider the case $m=\ell T$ with $\ell\ge\ell_1$ sufficiently large.

Let $Y=F_\lambda^{-1}A$. Since $\mu(Y)\ge 2^{-\varepsilon m/2}$, the first part of Lemma 2.4 guarantees that there is a set $X_{\ell,i}$ with $\nu(Y)\ge 2^{-\varepsilon m}$, where $\nu=\mu_{X_{\ell,i}}$. At this point we are working at a fixed (small) scale, with a regular measure ν and a set Y of sufficiently large measure, which is a subset of Z given in (5.10). Then essentially the argument from the proof of Theorem 1.7 shows that

$$H_m(F_\lambda\nu_Y)\ge m(\alpha/d+\eta'),$$

which, since $F_\lambda\nu_Y$ is supported on A, gives (5.11). The only difference is that ν might be very uniform, so that the upper bound in (4.5) need not hold. However, this case is even better for us. Formally, let $\overline{\alpha}=(1+\alpha)/2$. Then we consider the cases $\sigma_1+\cdots+\sigma_\ell\le\overline{\alpha}\ell$ and $\sigma_1+\cdots+\sigma_\ell>\overline{\alpha}\ell$ separately. In the first case, (4.5) does hold, with $u=u(\alpha)$, so we can proceed exactly as in the proof of Theorem 1.7. Otherwise, we apply Theorem 4.9 in place of Theorem 4.1, and use the fact that the function γ from Theorem 3.4 satisfies $\gamma(\alpha)\ge\alpha/d-7\delta$. In the end, the lower estimate we get on $H_m(\nu_Y)$ in this case is $m(\overline{\alpha}-$ error term), where we can make the error term arbitrarily small, completing the proof.

6. Applications and generalizations

6.1. Real-analytic families

In this section we derive several applications and generalizations of Theorems 1.3 and 1.7. We start with a statement about one-parameter real-analytic families.

Theorem 6.1. *Given $d \geq 2$, $\alpha \in (0, d)$ and $\kappa \in (d-2, 2)$ there is $\eta = \eta_d(\alpha, \kappa) > 0$ (depending continuously on the parameters) such that the following holds.*

Let $F_y(x) = F(x, y) : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ be a real-analytic function. Let $X \subset \mathbb{R}^d$ be a Borel set of Hausdorff dimension $> \alpha$, and suppose $y \mapsto \operatorname{dir} F_y'(x)$ is non-constant for all $x \in X$ outside of a set of dimension $< \dim_{\mathrm{H}}(X)$. Then

$$\dim_{\mathrm{H}}\{y : \dim_{\mathrm{H}}(F_y X) \leq \alpha/d + \eta\} \leq \kappa.$$

Proof. Let $\eta(\kappa, \alpha, d)$ be the number provided by Theorem 1.3. Let μ be a Frostman measure of exponent α supported on the $x \in X$ for which $y \mapsto \theta_x(y) := \operatorname{dir} F_y'(x)$ is non-constant. It is enough to show that if Y is a compact set with $\dim_{\mathrm{H}}(Y) > \kappa$ then there is $y \in Y$ such that $\dim_{\mathrm{H}}(F_y X) > \alpha/d + \eta$.

Let ν be a κ-Frostman measure on Y. By analyticity and the fact that $\theta_x(y)$ is non-constant, we can find open balls $U \subset \mathbb{R}^d$ and $V \subset \mathbb{R}$ meeting $\operatorname{supp}(\mu)$ and $\operatorname{supp}(\nu)$ respectively, and an index $j \in \{1, \ldots, d-1\}$ such that

$$\left| \frac{d}{dy} \theta_{x,j}(y) \right| \geq c > 0 \quad (x \in U,\ y \in V),$$

where $\theta_{x,j}$ is the j-th coordinate of θ_x. (To be more precise, this holds in a local chart of S^{d-1}.) Then $\theta_x \nu_V$ satisfies a Frostman condition of exponent κ, uniformly in $x \in U$. Indeed, it is enough to prove this for the projection $\theta_{x,j} \nu_V$ of $\theta_x \nu_V$ to its j-th coordinate, but this holds since $\theta_{x,j}$ is bi-Lipschitz on V, uniformly in $x \in U$.

Since $\kappa > d - 2$, the assumptions of Theorem 1.3 are satisfied with $\Lambda = U \cap \operatorname{supp}(\nu)$ and $V \cap \operatorname{supp}(\mu)$ in place of X (and $\Lambda_x \equiv \Lambda$). The conclusion of Theorem 1.3 is exactly what we were trying to prove. ∎

Remark 6.2. The proof shows that, instead of analyticity, it is enough to assume that F is C^2 and the set of zeros of $\frac{d}{dy}\theta_x(y)$ has Hausdorff dimension $< \kappa$ for all x outside of a set of Hausdorff dimension $< \dim_{\mathrm{H}}(X)$.

6.2. Distance sets

In this section we apply Theorem 1.3 to prove Theorem 1.1. In order to verify that the assumptions of Theorem 1.3, we rely on the following spherical projection estimate. Let

$$e(x, y) = \frac{y - x}{|y - x|}$$

be the direction spanned by two different points in $\mathbb{R}^d$, and write $e_x(y) = e(x, y)$ for the spherical projection with center x.

Theorem 6.3. *Fix $1 \leq k < d$. Given $\kappa, \alpha > 0$ there are $c, \eta > 0$ (depending continuously on α, κ) such that the following holds.*

Let $\mu, \nu \in \mathcal{P}(B^d(0, 1))$ be measures satisfying the decay conditions

$$\mu(V^{(r)}) \leq C_\mu r^\kappa, \quad \nu(V^{(r)}) \leq C_\nu r^\alpha,$$

for all $V \in \mathbb{A}(d, k-1)$ and $0 < r \leq 1$. (If $k = 1$, $V^{(r)}$ is a ball of radius r.) Suppose ν gives zero mass to every affine k-plane. Then for all x in a set E of μ-measure $\geq 1/2$ there is a set $K = K(x)$ with $\nu(K(x)) \geq 1/2$ such that

$$e_x(\nu_{K(x)})(W^{(r)}) \leq r^{\eta} \quad (r \in (0, r_0],\ W \in \mathbb{G}(d,k)), \tag{6.1}$$

where $r_0 > 0$ depends only on d, μ, C_ν, α.

Finally, the set $\{(x, y) : x \in E,\ y \in K(x)\}$ is compact.

In dimension $d = 2$, this is due to T. Orponen [24] . The argument in higher dimensions is similar; for completeness, we include the details of the proof in Appendix B.

Proof of Theorem 1.1. We first prove the theorem for the Euclidean metric, and comment on the changes required to handle other norms at the end.

We may assume that $\dim_{\mathrm{H}}(X) > \alpha$, provided we can show the value of η is continuous in the parameters (we can then apply the claim to $\alpha - \varepsilon$ in place of α for a suitably small ε to deduce that the statement also holds for $\dim_{\mathrm{H}}(X) = \alpha$). It is enough to show that there is $\eta = \eta_d(\alpha, \kappa) > 0$ such that, for all compact sets Y with $\dim_{\mathrm{H}}(Y) > \kappa$, there is $y \in Y$ satisfying $\dim_{\mathrm{H}}(\Delta^y X) \geq \alpha/d + \eta$. We may assume that Y is disjoint from X and both X and Y are contained in the unit ball. Let μ be an α-Frostman measure on X.

Let ν be a Frostman measure on Y with exponent κ. Our goal is to apply Theorem 1.3 with $\Lambda = Y$ and $F_y(x) = |x - y|$. Note that

$$F_y'(x) = y - x/|y - x| = e(x, y) \in S^{d-1}.$$

Thus the maps θ_x from Theorem 1.3 are the spherical projections $e_x(y) = e(x, y)$.

Suppose first that there is a hyperplane P with $\nu(P) > 0$; we can then assume that $Y = P$. We consider two subcases. If $\dim_{\mathrm{H}}(X \cap P) \geq \alpha$, then we are in the $(d-1)$-dimensional setting and we can argue by induction to get an even better estimate (note that the base case $d = 1$ is trivial). Otherwise, $\dim_{\mathrm{H}}(X \setminus P) = \dim_{\mathrm{H}}(X)$, so we may assume that $X \cap P = \emptyset$.

Now, for $x \notin P$, the map $y \mapsto e_x(y) : P \cap B(0,1) \to S^{d-1}$ is bi-Lipschitz onto its image, with a constant that depends on $\operatorname{dist}(x, P)$. Hence $e_x\nu(B_r) \lesssim_x r^{\kappa}$ for all r-balls B_r and $x \in X$. For every linear hyperplane $H \in \mathbb{G}(d, d-1)$ we can cover $S^{d-1} \cap H^{(r)}$ by $\lesssim r^{-(d-2)}$ balls of radius r. Since $\kappa > d - 2$, this shows that the assumptions of Theorem 1.3 hold (with $\kappa - (d-2)$ in place of κ). Theorem 1.3 then gives the desired conclusion.

Suppose now that ν gives zero mass to all hyperplanes. In this case we apply Theorem 6.3 with $k = d - 1$. As before, the hypotheses hold because $\kappa, \alpha > d - 2$. Let $\kappa' = \kappa'(\alpha, \kappa) > 0$ be the value given by Theorem 6.3. The theorem provides us with a set X' with $\mu(X') \geq 1/2$, and sets $K(x)$ for each $x \in X'$ with $\nu(K(x)) \geq 1/2$, such that (6.1) holds. The claim now follows from Theorem 1.3 applied to the set X' (with parameters α and κ').

We now consider the case of a C^2 norm N with unit ball of everywhere positive Gaussian curvature. Note that $N'(x) = N'(x/|x|)$ for all non-zero x. Hence, defining

$\varphi(e) = N'(e) : S^{d-1} \to \mathbb{R}^d$ and $\psi = \mathrm{dir} \circ \varphi : S^{d-1} \to S^{d-1}$, we have

$$\theta_x(y) = \mathrm{dir}\, \frac{d}{dx} N(x-y) = \mathrm{dir}\, \varphi(e_x(y)) = \psi(e_x(y)),$$

whence $\theta_x = \psi \circ e_x$. Now the hypothesis of everywhere positive Gaussian curvature translates to the map $\psi(e)$ having non-vanishing Jacobian (this can be seen e.g. by expressing the unit sphere of N as a graph $y = f(x)$ in local coordinates, and noting that the Jacobian of ψ is essentially the Hessian of f). Therefore we can cover S^{d-1} by finitely many patches on which ψ is a diffeomorphism (onto its image) and hence bi-Lipschitz. It follows that any decay condition enjoyed by $e_x \nu$ is still valid for $\psi e_x \nu = \theta_x \nu$, with different constants but the same exponent. With this observation, the proof in the Euclidean case goes through unchanged. ∎

Remark 6.4. The hypotheses on the norm N can be substantially weakened. What the argument above actually uses is that the Gauss map of the unit ball of N locally has a Lipschitz inverse. This can be relaxed further; for example, if the Gauss map locally has a Hölder inverse, then the argument still goes through, except that the dimension gain c will depend also on the Hölder exponent of these local inverses. In particular, Theorem 1.1 extends, for example, to all ℓ^p norms for $p \in (1, \infty)$.

6.3. Discretized incidences

As noted in the introduction, the circle of problems studied in this paper are related to incidence counting in the discretized setting. In this section we show how Theorem 1.7 can be used to deduce non-trivial incidence counting bounds, under suitable assumptions. Let $\{C_a : a \in B^2(0,1)\}$ be a parametrized family of planar curves. Fix δ-separated sets $E, A \subset B^2(0,1)$. We are interested in bounding the size of the (discretized) incidence set

$$\mathcal{I}(E, A) = \{(p, a) \in E \times A : \mathrm{dist}(p, C_a) < \delta\}.$$

Let us make the further assumptions that $E \subset X \times \mathbb{R}$ for some δ-separated set $X \subset [-1, 1]$, and that each curve meets vertical lines in a uniformly bounded number of points. In this case, we have the trivial bound

$$|\mathcal{I}(E, A)| \lesssim |X|\,|A|.$$

In general, this bound is sharp, for example if $E^{(2\delta)}$ is a small square and all curves parametrized by A cross this square from side to side. It is also sharp in cases where both E and A satisfy strong non-concentration assumptions. The examples are given by "train tracks", a well known object that many problems in discretized geometry have to grapple with (see, for example, [15, Figure 1]). For concreteness, let $X \subset [-1, 1]$ be a well separated set of size $\approx \delta^{-1/2}$ (for example, it could be an arithmetic progression of gap $\delta^{-1/2}$), and let E be a maximal δ-separated subset of $X \times [0, \delta^{1/2}]$ (so $E^{(2\delta)}$ resembles a train track). Then if A is any set of δ-separated lines that cross the rectangle $[0, 1] \times [0, \delta^{1/2}]$ from side to side, the incidence count $|\mathcal{I}(E, A)|$ equals $|X|\,|A|$. The family of all such

lines is essentially (parametrized by) a ball of radius $\delta^{1/2}$, and is therefore maximally concentrated, but if we take a spread-out subset, such as horizontal lines $y = \delta j$ with $j \in [0, \delta^{-1/2}]$, then A is highly non-concentrated, and we still have $|\mathcal{I}(E, A)| \approx |X|\,|A|$. The next theorem shows that, under non-concentration assumptions on X and A, and a further hypothesis that rules out train-track examples, an improvement over the trivial bound can be achieved.

Theorem 6.5. *Given $\kappa \in (0,1)$ and $c > 0$ there is $\varepsilon(\kappa) > 0$ such that the following holds for all δ small enough in terms of κ, c.*

Let $G : B^2(0,1) \times [-1,1] \to \mathbb{R}$ be a C^2 map of C^2 norm $\leq c^{-1}$, and such that

$$|\partial G/\partial a(\cdot)| \geq c \quad \text{on } B^2(0,1) \times [-1,1].$$

Assume further that

$$\left| \frac{d}{dx} \operatorname{dir}\left(\frac{\partial G}{\partial a}(\cdot) \right) \right| \geq c \quad \text{on } B^2(0,1) \times [-1,1]. \tag{6.2}$$

Let $E, A \subset B^2(0,1)$, $X \subset [-1,1]$ be δ-separated sets such that $E \subset X \times \mathbb{R}$. Further, assume that

$$\begin{aligned} &|E| \leq \delta^{-\varepsilon} |X|\,|A|^{1/2}, \\ &|A \cap B(a, \delta|A|^{1/2})| \leq \delta^{\kappa}|A| \quad (a \in B^2(0,1)), \\ &|X \cap B(x,r)| \leq \delta^{-\varepsilon} r^{\kappa} |X| \quad (x \in [-1,1],\ r \in [\delta, 1]). \end{aligned} \tag{6.3}$$

Then

$$|\{(p,a) \in E \times A : \operatorname{dist}(p, \operatorname{Graph}(G_a)) < \delta\}| \leq \delta^{\varepsilon} |X|\,|A|.$$

Before presenting the deduction of this theorem from Theorem 1.7, we make some remarks and then give some examples.

Remark 6.6. The assumption (6.3) rules out train-track examples. Consider again lines $y = ax + b$ parametrized by (a,b). Let $X \subset [-1,1]$ be a well separated set with $|X| = \delta^{-t}$, let $E = X \times \{j\delta\}_{j=0}^{\delta^{-t}}$, and let $A_0 = \{j\delta\}_{j=0}^{\delta^{-t}} \times \{j\delta\}_{j=0}^{\delta^{-t}}$. Then $|E| \approx |X|\,|A_0|^{1/2}$ and $|\mathcal{I}(E, A_0)| \approx |X|\,|A_0|$. Of course, in this case A_0 is maximally concentrated. But if we allowed $|E|$ to be much larger than $|X|\,|A|^{1/2}$, we could take A to be a large non-concentrated subset of A_0. Then all the hypotheses of the theorem would hold, except for (6.3), and the conclusion would clearly fail. There is nothing special about lines here – for more general graphs, one has to consider "curved train tracks" of the form $C^{(\delta^s)} \cap (X \times \mathbb{R})$ for a fixed graph C in the family.

Remark 6.7. In the case of lines, Theorem 6.5 follows from Bourgain's discretized projection theorem, except that our non-concentration requirement on the lines is weaker (and in some sense the weakest possible one). We note that (for lines), M. Bateman and V. Lie [1] have previously obtained a related result, but under different assumptions. Roughly speaking, they assume more on the family of lines, and on the structure of the set E, but on the other hand, they have a single-scale non-concentration hypothesis

on X (while we have a single-scale non-concentration hypothesis on the set A of lines). Recently, L. Guth, N. Solomon and H. Wang [10] studied incidences between tubes satisfying the strongest possible non-concentration assumption, and obtained sharp incidence bounds in that regime.

Remark 6.8. We have tried to give an indication of the relationship between our results and incidence counting, but Theorem 6.5 is not the strongest or most general formulation possible. The hypothesis (6.2) can be weakened, although of course some non-degeneracy condition on the family of curves is needed. Concretely, the number on the left-hand side of (6.2) can be allowed to be zero on a set small enough that, if we remove a suitable neighborhood of it, the incidence counting does not change. Using the results in §6.4 below, one can also deduce versions in higher dimensions, and for higher-dimensional families of curves, although the required assumptions become more restrictive and cumbersome to verify.

We now give some examples of families of curves to which Theorem 6.5 applies. The verifications are straightforward calculations.

Corollary 6.9. *Let $\kappa, \varepsilon, E, A$ and X be as in Theorem 6.5. For each of the following families of curves, provided δ is small enough in terms of κ and the family, we have*

$$|\mathcal{I}(E, A)| \leq \delta^{\varepsilon}|X||A|.$$

(i) *The family of lines $y = ax + b$ with (a, b) in some fixed bounded set.*

(ii) *The family of parabolas $y = ax^2 + bx$ with (a, b) in some fixed bounded set.*

(iii) *The family of circles with center $(a, 0)$ and radius r, with (a, r) in some bounded set.*

(iv) *The family of circles of unit radius and center in a bounded set S, provided the projection of S onto the x-axis is separated from $(X - 1) \cup (X + 1)$.*

Proof of Theorem 6.5. Suppose $\delta = 2^{-m}$. Assume on the contrary that

$$|\{(p, a) \in E \times A : \operatorname{dist}(p, \operatorname{Graph}(G_a)) < 2^{-m}\}| > 2^{-\varepsilon m}|X||A|.$$

Let $E_x = \{y : (x, y) \in E\}$. Then

$$\sum_{x \in X} |\{a \in A : \operatorname{dist}(G_a(x), E_x) < 2^{-m}\}| > 2^{-\varepsilon m}|X||A|.$$

Hence, if we set $A_x = \{a \in A : \operatorname{dist}(G_a(x), E_x) < 2^{-m}\}$, we have

$$\left|\left\{x \in X : |A_x| \geq \tfrac{1}{2}2^{-\varepsilon m}|A|\right\}\right| \geq \tfrac{1}{2}2^{-\varepsilon m}|X|.$$

Let $X_0 \subset X^{(2^{-m})}$ be the 2^{-m}-neighborhood of the set appearing on the left-hand side, and write $\nu = \mathbf{1}_{X_0}/|X_0|$. Using $|X_0| \gtrsim 2^{-\varepsilon m}|X^{(2^{-m})}|$ and the non-concentration assumption on X, we see that

$$\nu(B(x, r)) \lesssim 2^{2\varepsilon m} r^{\kappa} \quad (r \in [2^{-m}, 1]).$$

Our goal is to apply Theorem 1.7 to the family $\{a \mapsto G_a(x)\}$ and the set $A^{(2^{-m})}$. Note that X plays the role of Λ. Using (6.2), the same argument in the proof of Theorem 6.1 shows that

$$\theta_a \nu(B(y,r)) \lesssim_c 2^{2\varepsilon m} r^\kappa,$$

uniformly over $a \in A$. Using the assumptions, we see that if ε is small enough in terms of κ, then the hypotheses of Theorem 1.7 are met (with $A^{(2^{-m})}$ in place of X). By definition of X_0, we have $|A_x| \gtrsim 2^{-\varepsilon m}|A|$ for $x \in X_0$, so again assuming ε is small enough only in terms of κ, Theorem 1.7 (applied to the 2^{-m}-neighborhoods of A, A_x) guarantees that, provided m is large enough in terms of c and κ,

$$|E_x| \gtrsim \mathcal{N}(\{G_a(x) : a \in A_x\}, m) \gtrsim 2^{\eta m}|A|^{1/2} \quad \text{for all } x \in X_0,$$

where $\eta = \eta(\kappa) > 0$ is the value provided by Theorem 1.7. We conclude

$$|E| \gtrsim 2^{\eta m}|X_0|\,|A|^{1/2} \gtrsim 2^{(\eta-\varepsilon)m}|X|\,|A|^{1/2}.$$

This contradicts (6.3) if ε is small enough compared to η (hence in terms of κ only), finishing the proof. ∎

6.4. A higher rank non-linear projection theorem

W. He [11, Theorem 1] extended Bourgain's discretized projection theorem to higher rank projections. Using his result, we can obtain a corresponding higher rank version of Theorems 1.3 and 1.7. The proofs are nearly the same; we only need to take a little care in extending Theorem 3.4 properly.

Following [11], given $V \in \mathbb{G}(d,k)$ and $W \in \mathbb{G}(d,d-k)$, we define

$$d(V,W) = \det(P_{W^\perp|V}),$$

where $P_{W^\perp|V}$ denotes the restriction of $P_{W^\perp}$ to V. See [11, (13)] for this characterization of $d(V,W)$. We remark that d is *not* a metric; to begin with, V and W live in different spaces. Also, $d(V,W) = 0$ if and only if $\dim(V \cap W) \geq 1$. However, d is symmetric. For $W \in \mathbb{G}(d,d-k)$, we will denote

$$\mathcal{V}(W,r) = \{V \in \mathbb{G}(d,k) : d(V,W) < r\}.$$

Given $W \in \mathbb{G}(d,d-k)$, the set $\{V \in \mathbb{G}(d,k) : d(V,W) = 0\}$ is a smooth hypersurface in $\mathbb{G}(d,k)$. Hence, $\mathcal{V}(W,r)$ can be seen as a neighborhood of this hypersurface.

Later we will need to deal with the case $k = d-1$, which is particularly simple. If $H \in \mathbb{G}(d,d-1)$ and $\ell \in \mathbb{G}(d,1)$, then

$$d(H,\ell) = |\det(P_{H^\perp|\ell})| = |\cos(\angle(H^\perp,\ell))|. \tag{6.4}$$

The following lemma will help us achieve the correct generalization of Theorem 3.4.

Lemma 6.10. *Fix* $1 \leq k \leq d-1$.

(i) *Given* $x \in S^{d-1} \cap W$ *with* $W \in \mathbb{G}(d, d-k)$,

$$d(W, V) \leq \operatorname{dist}(x, V).$$

(ii) *Given* $x \in S^{d-1} \cap W^{\perp}$ *with* $W \in \mathbb{G}(d, d-k)$,

$$d(W, V) \leq |P_V(x)|.$$

Proof. (i) Note that $\operatorname{dist}(x, V) = |P_{V^{\perp}}(x)|$. Hence if we extend x to an orthonormal basis of W, pick an orthonormal basis for $V^{\perp}$, and write the matrix of $P_{V^{\perp}|W}$ in these bases, the first row has norm $\operatorname{dist}(x, V)$, while all the rows have norm at most 1. Hence

$$d(W, V) = |\det(P_{V^{\perp}|W})| \leq \operatorname{dist}(x, V).$$

(ii) Since $|P_V(x)| = \operatorname{dist}(x, V^{\perp})$, the claim follows from the first one and the identity $d(W, V) = d(V^{\perp}, W^{\perp})$ (see [11, (14)]). ∎

Corollary 6.11. *Fix* $1 \leq k \leq d-1$. *Let* $\rho \in \mathcal{P}(\mathbb{G}(d, k))$ *be a measure satisfying the decay condition*

$$\rho(\mathcal{V}(W, r)) \leq C r^{\kappa} \quad \text{for all } 0 < r \leq 1, W \in \mathbb{G}(d, d-k). \tag{6.5}$$

Then for every $x \in \mathbb{R}^d \setminus \{0\}$,

$$\rho\{V : \operatorname{dist}(x, V) \leq r\} \leq C(r/|x|)^{\kappa}, \tag{6.6}$$

$$\rho\{V : |P_V(x)| \leq r\} \leq C(r/|x|)^{\kappa}. \tag{6.7}$$

Proof. By rescaling, we may assume x has unit norm. Choosing W such that $x \in W$ (resp. $x \in W^{\perp}$), the claim is immediate from Lemma 6.10. ∎

The non-concentration condition (6.5) is the one appearing in He's projection theorem [11, Theorem 1]. On the other hand, (6.6) is the required decay in order to get the higher rank version of Theorem 3.2, and (6.7) is the decay needed to obtain the higher rank version of Theorem 3.3. The proofs are very similar to the rank 1 case; see [22, §5.3] for details. Note that even though in [22, §5.3] there is a Frostman condition (on balls) for ρ, the exponents are chosen large enough so that (6.6)–(6.7) hold; see in particular [22, (5.11) and (5.12)]. Thus Corollary 6.11 shows that (6.5) is enough to obtain the decay required in all three projection regimes. As a consequence, the same proof of Theorem 3.4 yields

Theorem 6.12. *Fix* $1 \leq k \leq d-1$. *Given* $0 < \kappa < 1$ *there exists* $\eta = \eta(\kappa) > 0$ *such that the following holds for all sufficiently small* $\delta \leq \delta_0(\kappa)$ *and all sufficiently large* $m \geq m_0(\delta)$. *Fix* $\alpha \in [0, d]$. *Let* $\mu \in \mathcal{P}_m^d$, *and let* $\rho \in \mathcal{P}(\mathbb{G}(d, k))$ *satisfy*

$$\mu(B(x, r)) \leq 2^{\delta m} r^{\alpha} \quad \text{for all } r \in [2^{-m}, 1],\ x \in [0, 1)^d,$$

$$\rho(\mathcal{V}(W, r)) \leq 2^{\delta m} r^{\kappa} \quad \text{for all } r \in [2^{-m}, 1],\ W \in \mathbb{G}(d, d-k).$$

Then there is a set $E \subset \mathbb{G}(d,k)$ with $\rho(E) \le 2^{-\delta m}$ such that $P_V\mu$ is $(\gamma(\alpha),\delta,m)$-robust for all $V \in \mathbb{G}(d,k) \setminus E$, where

$$\gamma(\alpha) = \begin{cases} \alpha - 6\delta & \text{if } \alpha < \kappa/2, \\ \frac{k}{d}\alpha + \eta & \text{if } \kappa/2 \le \alpha \le d - \kappa/2, \\ k - 6\delta & \text{if } \alpha > d - \kappa/2. \end{cases}$$

In turn, the proofs of Theorems 1.3 and 1.7 carry over to the higher rank setting with only notation changes. Of course the decay assumption on ρ must be of the form (6.5), and the dimension/exponent of the "good" projections becomes $\frac{k}{d}\alpha + \eta$. For example, the following is the higher rank analog of Theorem 1.3.

Theorem 6.13. *Fix $1 \le k \le d-1$. Given $\kappa > 0$ and $0 < \alpha < d$ there is $\eta = \eta_{d,k}(\kappa,\alpha) > 0$ (that can be taken continuous in κ,α) such that the following holds.*

Let (Λ,ν) be a Borel probability space and let $U \subset \mathbb{R}^d$ be an open domain. Let $F : \Lambda \times U \to \mathbb{R}^k$ be a Borel map such that $x \mapsto F_\lambda(x) = F(\lambda,x)$ is C^2 for all $\lambda \in \Lambda$, and $DF_\lambda(x)$ has full rank k for all $(\lambda,x) \in \Lambda \times U$. For each $x \in U$ define the map

$$V_x(\lambda) = \ker(DF_\lambda(x))^\perp \in \mathbb{G}(d,k).$$

Let $A \subset U$ be a Borel set of Hausdorff dimension $\ge \alpha$ such that for all $x \in A$ there are a set Λ_x with $\nu(\Lambda_x) > 0$ and a number $C_x > 0$ satisfying

$$V_x\nu_{\Lambda_x}(\mathcal{V}(W,r)) \le C_x r^\kappa \quad \text{for all } W \in \mathbb{G}(d,d-k),\ r \in (0,1]. \tag{6.8}$$

Then there is $\lambda \in \operatorname{supp}(\nu)$ such that

$$\dim_{\mathrm{H}}(F_\lambda A) \ge \frac{k}{d}\alpha + \eta.$$

6.5. Dimension of spherical projections

It is well known that if $A \subset \mathbb{R}^d$ is a Borel set of Hausdorff dimension $> d-1$, then the set of directions spanned by A has full measure in S^{d-1} (this follows e.g. from the Marstrand–Mattila Intersection Theorem). This clearly fails if $\dim_{\mathrm{H}}(A) = d-1$, as A can then be contained in a hyperplane. But what if A is not contained in a hyperplane? One might conjecture that then the set of directions spanned by A has full dimension $d-1$, and maybe even there is $x \in A$ such that the spherical projection $e_x(A)$ has full dimension (recall that $e_x(y) = x - y/|x-y|$). This is wide open in all dimensions $d \ge 2$, but Orponen [24, Theorem 1.5] proved the following partial result.

Theorem 6.14. *Let $A, E \subset \mathbb{R}^2$ be Borel sets such that $\dim_{\mathrm{H}}(E) > 0$ and E is not contained in a line. Then there is $y \in E$ such that*

$$\dim_{\mathrm{H}}(e_y(A \setminus \{y\})) \ge \dim_{\mathrm{H}}(A)/2.$$

In particular, if $A \subset \mathbb{R}^2$ is not contained in a line, it spans a set of directions of dimension $\ge \dim_{\mathrm{H}}(A)/2$.

Very recently, B. Liu and C.-Y. Shen [21, Theorem 1.3] combined Orponen's approach with Bourgain discretized sum-product theorem to improve this, by showing that if $A \subset \mathbb{R}^2$ has Hausdorff dimension $\alpha \in (0, 2)$ and is not contained in a line, then it spans a set of directions of dimension $\geq \alpha/2 + c(\alpha)$, where $c(\alpha) > 0$. They also obtained a "pinned" version [21, Theorem 1.4]. See also [2, Theorem 1.6] for an earlier, closely related "single-scale" result. As a corollary of our framework, we obtain a pinned version that is slightly different from that of Liu and Shen, and partially extend it to higher dimensions. We note that Orponen's proof of Theorem 6.14 features heavily in our argument via Theorem 6.14 (see Appendix B for more details).

Theorem 6.15. *Fix $d \geq 2$. Given $\alpha, \kappa \in (d-2, d)$, there is $\eta = \eta_d(\alpha, \kappa) > 0$ such that the following holds.*

Let $A, E \subset \mathbb{R}^d$ be Borel sets such that $\dim_{\mathrm{H}}(A) = \alpha$, $\dim_{\mathrm{H}}(E) = \kappa$ and E is not contained in a hyperplane. Then there is $y \in E$ such that

$$\dim_{\mathrm{H}}(e_y A) \geq \frac{(d-1)\alpha}{d} + \eta.$$

Remark 6.16. Compared to [21, Theorem 1.4] (which only considers the planar case), the case $d = 2$ of the theorem has the natural hypothesis that E, rather than A as in [21], is not contained in a line; this allows us to avoid the dichotomy in [21, Theorem 1.4]. On the other hand, unlike in [21, Theorem 1.4], the gain η depends on the dimensions of both A and E.

Proof of Theorem 6.15. The proof is similar to that of Theorem 1.1, but (in dimension $d > 2$) we need to appeal to Theorem 6.13 instead.

As usual we may assume that A and E are disjoint compact subsets of the unit ball of positive Hausdorff measure in their dimensions. We consider the parametrized family of smooth maps $\{e_y(x) : y \in E\}$ defined on A. The range of these maps is S^{d-1}, but we can identify it with $\mathbb{R}^{d-1}$ in local coordinates. For example, by restricting and rotating A and E, we may assume that $y_d - x_d \neq 0$ for all $x \in A$, $y \in E$, and work with the map $\widetilde{e}_y(x) = (y_i - x_i/(y_d - x_d))_{i=1}^{d-1}$ instead.

Our aim is to apply Theorem 6.13 (in the case $d = 2$, we can apply Theorem 1.3). It is easy to see that the map V_x featuring in Theorem 6.13 is given by

$$V_x(y) = e_x(y)^{\perp}.$$

Recalling (6.4), this implies that, for $\theta \in S^{d-1}$,

$$d(V_x(y), \langle \theta \rangle) = |\langle e_x(y), \theta \rangle| < r \iff e_x(y) \in (\theta^{\perp})^{(r)}. \tag{6.9}$$

Let μ, ν be Frostman measures on A, E of exponents α, κ. Then

$$\mu(V^{(r)}) \lesssim r^{\alpha-(d-2)}, \quad \nu(V^{(r)}) \lesssim r^{\kappa-(d-2)} \tag{6.10}$$

for all $V \in \mathbb{A}(d, d-2)$ and $r \in (0,1]$. We may assume that μ assigns zero mass to all affine hyperplanes, for otherwise there is a hyperplane P such that $\dim_{\mathrm{H}}(A \cap P) = \dim_{\mathrm{H}}(A)$ and, by assumption, there is a point $y \in E \setminus P$. It is then clear that

$$\dim_{\mathrm{H}}(e_y(A \cap P)) = \dim_{\mathrm{H}}(A \cap P) = \dim_{\mathrm{H}}(A),$$

and we are done.

Suppose first that ν assigns positive mass to a hyperplane P. Then we can assume that $E \subset P$, and because $\mu(P) = 0$, we may also assume that A is disjoint from P. Using this and (6.9), it follows that, for $x \in A$,

$$V_x \nu(\mathcal{V}(\langle\theta\rangle, r)) = \nu\{e_x^{-1}((\theta^\perp)^{(r)})\} \le \nu(V_\theta^{(C_x r)}),$$

where $V_\theta = e_x^{-1}(\theta^\perp) \cap P \in \mathbb{A}(d, d-2)$, and the constant C_x depends on the distance from x to P. According to (6.10), the hypotheses of Theorem 6.13 are met, and we find that there is $y \in E$ such that

$$\dim_{\mathrm{H}}(e_y A) \ge \frac{(d-1)\dim_{\mathrm{H}}(A)}{d} + \eta,$$

where $\eta = \eta(\kappa - (d-2), \alpha)$ is the number given by Theorem 6.13.

Suppose now that ν gives zero mass to every hyperplane. We can then apply Theorem 6.3 with $k = d-1$ to obtain a parameter $\kappa' = \kappa'_d(\kappa, \alpha) > 0$ and a set L with $\mu(L) \ge 1/2$ such that for each $x \in L$ there is a set $K(x) \subset A$ with $\nu(K(x)) \ge 1/2$ such that

$$e_x \nu_{K(x)}(H^{(r)}) \le O_{\nu,\mu}(1) r^{\kappa'} \quad \text{for all } H \in \mathbb{G}(d, d-1),\ 0 < r \le 1. \tag{6.11}$$

Using (6.9), we deduce that

$$V_x \nu_{K(x)}(\mathcal{V}(\langle\theta\rangle, r)) = e_x \nu_{K(x)}((\theta^\perp)^{(r)}).$$

Recalling (6.11), this shows that the hypotheses of Theorem 6.13 are satisfied with κ' in place of κ. Hence, if $\eta = \eta_d(\kappa', \alpha) > 0$ is the value provided by Theorem 6.13, then there is $y \in \operatorname{supp}(\nu) \subset E$ such that

$$\dim_{\mathrm{H}}(e_y(A)) \ge \frac{(d-1)\alpha}{d} + \eta.$$

This is what we wanted to prove. ∎

Appendix A. Entropy of projections

A.1. Entropy basics

In this section we prove Proposition 5.1. We start by reviewing some basic facts about (Shannon) entropy. Recall that if $\nu \in \mathcal{P}(\mathbb{R}^d)$ and $\mathcal{A}$ is a finite partition of $\mathbb{R}^d$, up to a ν-null set, then the entropy of ν with respect to $\mathcal{A}$ is given by

$$H(\nu, \mathcal{A}) = -\sum_{A \in \mathcal{A}} \nu(A) \log(\nu(A)),$$

with the usual convention $0 \cdot \log 0 = 0$, and in the case $\mathcal{A} = \mathcal{D}_m$, we write $H_m(\nu) = H(\nu, \mathcal{D}_m)$. Likewise, we define the conditional entropy with respect to the finite measurable partition $\mathcal{G}$ by

$$H(\mu, \mathcal{A} \mid \mathcal{G}) = \sum_{G \in \mathcal{G} : \mu(G) > 0} \mu(G) H(\mu_G, \mathcal{A}).$$

It follows from the concavity of the logarithm that one always has

$$H(\nu, \mathcal{A} \mid \mathcal{G}) \leq \log |\mathcal{A}|. \tag{A.1}$$

The following are some further elementary properties of entropy that we will need to call upon:

(A) If $\mathcal{F}, \mathcal{G}$ have the property that each element of $\mathcal{F}$ hits at most N elements of $\mathcal{G}$ and vice versa, then
$$|H(\mu, \mathcal{F}) - H(\mu, \mathcal{G})| \leq \log N.$$

(B) If $\mathcal{G}$ refines $\mathcal{F}$ (that is, each element of $\mathcal{F}$ is a union of elements in $\mathcal{G}$), then
$$H(\mu, \mathcal{F} \mid \mathcal{G}) = H(\mu, \mathcal{G}) - H(\mu, \mathcal{F}).$$

(C) Conditional entropy is concave as a function of the measure: for $t \in [0, 1]$,
$$H(t\mu + (1-t)\nu, \mathcal{F} \mid \mathcal{G}) \geq tH(\mu, \mathcal{F} \mid \mathcal{G}) + (1-t)H(\nu, \mathcal{F} \mid \mathcal{G}).$$

A.2. Multiscale estimates for the entropy of projections

The following proposition provides a lower bound for the entropy of a smooth image in terms of entropies of linear images of conditional measures at certain scales. Recall that we denote orthogonal projection onto $V \in \mathbb{G}(d, k)$ by P_V. Furthermore, if $A : \mathbb{R}^d \to \mathbb{R}^k$ is a linear map of rank k, we denote by $J_k(A)$ the absolute value of the determinant of $A|_{\ker(A)^\perp} : \ker(A)^\perp \to \mathbb{R}^k$.

Proposition A.1. *Fix $1 \leq k < d$. Let $\mu \in \mathcal{P}([0,1)^d)$ and let $[A_i, B_i)_{i=1}^q$ be disjoint subintervals of $(0, m]$ such that $B_i \leq 2A_i$. Let $F : U \to \mathbb{R}^k$ be a C^2 map defined in a neighborhood of* $\operatorname{supp}(\mu)$ *such that $DF(x)$ has full rank k for all $x \in \operatorname{supp}(\mu)$. Denote*

$$V(x) = \ker(DF(x))^\perp.$$

Then

$$H(F\mu, \mathcal{D}_m) \geq -O_{F,d,k}(q) + \sum_{i=1}^{q} \sum_{Q \in \mathcal{D}_{A_i}} \mu(Q) H(P_{V(x_Q)} \mu^Q, \mathcal{D}_{B_i - A_i}), \tag{A.2}$$

where x_Q is an arbitrary point in Q. The constant implicit in $O_{F,d,k}(q)$ depends only on d, k, $\|F\|_{C^2}$ and $\inf_{x \in \operatorname{supp}(\mu)} J_k(F'(x))$*; in particular, it can be taken uniform in a C^2 neighborhood of F.*

The proof of the proposition depends on a linearization argument, which helps explain the assumption that F is C^2 and has no singular points. The hypothesis $B_j \le 2A_j$ comes from linearization, and can be dropped if F is already linear.

Lemma A.2. *Under the assumptions of Proposition A.1, if $Q \in \mathcal{D}_{A_j}$ has positive μ-measure and $x \in Q$, then*

$$|H(F(\mu_Q), \mathcal{D}_{B_j}) - H(P_{V(x)}(\mu_Q), \mathcal{D}_{B_j})| \lesssim_{F,d} 1. \tag{A.3}$$

Proof. All implicit constants are allowed to depend on d. The claim can be rewritten as

$$|H(\mu_Q, F^{-1}(\mathcal{D}_{B_j})) - H(\mu_Q, P_{V(x)}^{-1}(\mathcal{D}_{B_j}))| \lesssim_F 1.$$

Let $L_x(z) = F(x) + DF(x) \cdot (z - x)$ be the affine approximation to F centered at x. Note that

$$DF(x)|_{V(x)} P_{V(x)} = DF(x),$$

and $DF(x)|_{V(x)}$ is an isomorphism from $V(x)$ to $\mathbb{R}^k$. By the assumption that F has no singular points, the determinant of this isomorphism (given by $J_k(DF(x))$ in absolute value) is bounded away from 0 and ∞ for $x \in \operatorname{supp}(\mu)$. Hence it follows from property (A) that

$$|H(\mu_Q, L_x^{-1}(\mathcal{D}_{B_j})) - H(\mu_Q, P_{V(x)}^{-1}(\mathcal{D}_{B_j}))| \lesssim_F 1.$$

So it is enough to show that

$$|H(\mu_Q, F^{-1}(\mathcal{D}_{B_j})) - H(\mu_Q, L_x^{-1}(\mathcal{D}_{B_j}))| \lesssim_F 1. \tag{A.4}$$

This is just a consequence of Taylor's formula. Indeed, since F is C^2,

$$|F(z) - L_x(z)| \le O_{\|F\|_{C^2}}(|z - x|^2) \le O_{\|F\|_{C^2}}(2^{-B_j}) \quad \text{for } z \in Q,$$

using the assumption $B_j \le 2A_j$. This implies that each element of $F^{-1}(\mathcal{D}_{B_j})$ hitting Q intersects $\lesssim_F 1$ elements of $L_x^{-1}(\mathcal{D}_{B_j})$, and vice versa. Property (A) above implies that (A.4) is satisfied, and this establishes the claim (A.3). ∎

We can now conclude the proof of Proposition A.1.

Proof of Proposition A.1.

$$\begin{aligned} H(F\mu, \mathcal{D}_m) &\overset{(B)}{\ge} -H(F\mu, \mathcal{D}_0) + \sum_{i=1}^{q} H(F\mu, \mathcal{D}_{B_i} \mid \mathcal{D}_{A_i}) \\ &\overset{(A.1)}{\ge} -O_{F,d}(1) + \sum_{i=1}^{q} H\Big(\sum_{Q \in \mathcal{D}_{A_i}} \mu(Q) F(\mu_Q), \mathcal{D}_{B_i} \;\Big|\; \mathcal{D}_{A_i} \Big) \\ &\overset{(C)}{\ge} -O_{F,d}(1) + \sum_{i=1}^{q} \sum_{Q \in \mathcal{D}_{A_i}} \mu(Q) H(F(\mu_Q), \mathcal{D}_{B_i} \mid \mathcal{D}_{A_i}) \end{aligned}$$

$$\overset{\text{(B),(A.1)}}{\geq} -O_{F,d}(q) + \sum_{i=1}^{q} \sum_{Q \in \mathcal{D}_{A_i}} \mu(Q) H(F(\mu_Q), \mathcal{D}_{B_i})$$

$$\overset{\text{(A.3)}}{\geq} -O_{F,d}(q) + \sum_{i=1}^{q} \sum_{Q \in \mathcal{D}_{A_i}} \mu(Q) \big(H(P_{V(x_Q)}\mu_Q, \mathcal{D}_{B_i}) - O_F(1) \big)$$

$$\overset{P_{V(x_Q)} \text{ linear}}{=} -O_{F,d}(q) + \sum_{i=1}^{q} \sum_{Q \in \mathcal{D}_{A_i}} \mu(Q) H(\Pi_{V(x_Q)}\mu^Q, \mathcal{D}_{B_i - A_i}).$$

An inspection of the proof shows that the constants depending on F do so in the way described in the statement. ∎

We can now complete the proof of Proposition 5.1, which can be seen as a "robust" version of Proposition A.1; we repeat the statement for the reader's convenience.

Proposition A.3. *Under the same assumptions and notation of Proposition A.1, if $\nu \in \mathcal{P}(\mathbb{R}^d)$ satisfies $\nu \leq \Delta\mu$, then*

$$H_m(F\nu) \geq -O_{F,d}(q) + \sum_{i=0}^{q-1} \sum_{Q \in \mathcal{D}_{A_i}} \nu(Q) H^{m\Delta}_{B_i - A_i}(P_{V(x_Q)}\mu^Q).$$

Proof. Fix $i \in \{0, \dots, q-1\}$, and note that

$$\sum \left\{ \nu(Q) : Q \in \mathcal{D}_{A_i}, \nu(Q) < \frac{1}{m}\mu(Q) \right\} < \frac{1}{m}. \tag{A.5}$$

Suppose $\nu(Q) \geq \frac{1}{m}\mu(Q) > 0$ for a given $Q \in \mathcal{D}_{A_i}$. Then

$$\nu_Q(S) = \frac{\nu(Q \cap S)}{\nu(Q)} \leq \frac{\Delta\mu(Q \cap S)}{\frac{1}{m}\mu(Q)} = m\Delta\mu_Q(S)$$

for any Borel set $S \subset [0,1)^2$. We deduce that $\Pi\nu_Q \leq m\Delta\Pi\mu_Q$ for any linear map Π, and hence

$$H_{B_i - A_i}(P_{V(x_Q)}\nu^Q) \geq H^{m\Delta}_{B_i - A_i}(P_{V(x_Q)}\mu^Q),$$

always assuming that $\nu(Q) \geq \frac{1}{m}\mu(Q) > 0$ and $Q \in \mathcal{D}_{A_i}$.

On the other hand, for fixed i, from (A.5) and the trivial bound $H_p(\cdot) \leq k(p + O(1))$ for measures supported on a ball of radius $O(1)$ in $\mathbb{R}^k$, we get

$$\sum_{Q \in \mathcal{D}_{A_i} : \nu(Q) < \frac{1}{m}\mu(Q)} \nu(Q) H^{m\Delta}_{B_i - A_i}(P_{V(x_Q)}\mu^Q) \leq k(B_i - A_i + O_d(1))/m.$$

Splitting (for each i) the sum $\sum_{Q \in \mathcal{D}_{A_i}}$ in Proposition A.1 into the cubes with $\nu(Q) \geq \frac{1}{m}\mu(Q)$ and $\nu(Q) < \frac{1}{m}\mu(Q)$, we get the desired result. ∎

Appendix B. Spherical projections

In this section we prove Theorem 6.3. In fact, we will prove the following more quantitative statement:

Theorem B.1. *Fix $1 \leq k < d$. Given $\kappa, \alpha > 0$ there are $c, \eta > 0$ (depending continuously on α, κ) such that the following holds.*

Let $\mu, \nu \in \mathcal{P}(B^d(0,1))$ be measures satisfying the decay conditions

$$\mu(V^{(r)}) \leq C_\mu r^\kappa, \quad \nu(V^{(r)}) \leq C_\nu r^\alpha$$

for all $V \in \mathbb{A}(d, k-1)$ and $0 < r \leq 1$. (If $k = 1$, $V^{(r)}$ is a ball of radius r.) Fix $\varepsilon \in (0,1)$ and suppose there is $\tilde{\delta} > 0$ such that

$$\nu(V^{(\tilde{\delta})}) \leq c\varepsilon \quad (V \in \mathbb{A}(d,k)). \tag{B.1}$$

Then for all x in a set E of μ-measure $\geq 1 - \varepsilon$ there is a set $K(x)$ with $\nu(K(x)) \geq 1 - \varepsilon$ such that

$$e_x(\nu_{K(x)})(W^{(r)}) \leq r^\eta \quad (r \in (0, r_0],\ W \in \mathbb{G}(d,k)), \tag{B.2}$$

where $r_0 > 0$ depends only on $d, C_\mu, C_\nu, \alpha, \tilde{\delta}, \varepsilon$.

Finally, the set $\{(x,y) : x \in E,\ y \in K(x)\}$ is compact.

Theorem 6.3 follows by taking $\varepsilon = 1/2$ and observing that, by compactness, if ν gives zero mass to all affine k-planes, then there is δ_0 such that (B.1) holds. We follow closely the ideas from [24, Section 2], although there are some differences: we work in arbitrary dimension, while Orponen works only in the plane. Also, Orponen obtains a single x such that (B.2) holds. On the other hand, some aspects of the proof are simplified; in particular, we avoid having to argue by induction in the scale, which reduces the number of parameters to keep track of in the proof (I thank Hong Wang for suggesting this simplification and allowing me to include it in this article).

The proof of the theorem depends on a single-scale variant that we state and prove first. By a *k-plate* in $\mathbb{R}^d$ we mean a set of the form

$$W^{(\delta)} \cap B^d(0,1)$$

with $W \in \mathbb{A}(d,k)$. In particular, 0-plates are balls and 1-plates are tubes (intersected with $B^d(0,1)$). We refer to δ as the *width* of the plate, and denote the family of all k-plates of width δ intersecting a set E by $\mathcal{T}_k(E,\delta)$.

Proposition B.2. *Given $\kappa, \alpha > 0$ and $1 \leq k \leq d-1$, the following holds if $\eta \leq \eta_0(\kappa,\alpha)$, and δ is small enough depending on C_μ, C_ν appearing below and all the other parameters.*

Let $\nu, \mu \in \mathcal{P}(B^d(0,1))$ be measures satisfying the decay conditions

$$\mu(V^{(r)}) \leq C_\mu r^\alpha \quad (V \in \mathbb{A}(d,k-1),\ 0 < r \leq 1),$$
$$\nu(V^{(\delta)}) \leq C_\nu \delta^\kappa \quad (V \in \mathbb{A}(d,k-1)).$$

Then there exist:

- *a set* $E \subset \operatorname{supp}(\mu)$ *with* $\mu(\mathbb{R}^d \setminus E) \le \delta^{\kappa\alpha/2}$,
- *for each* $x \in E$, *a set* $P(x)$, *which is either empty or a* k*-plate in* $\mathcal{T}_k(x, \delta^{\eta/2})$,

such that $\{(x, y) : x \in E,\ y \in \operatorname{supp} \nu \setminus P(x)\}$ *is compact, and*

$$\nu(W) \le \delta^{\eta} \quad \textit{for all } W \in \mathcal{T}_k(x, \delta) \cap \mathcal{T}_k(\operatorname{supp}(\nu) \setminus P(x), \delta).$$

In turn, the proposition relies on the following lemma stating that all δ-plates of too large measure are contained in a relatively small number of relatively thin plates.

Lemma B.3. *Let* $\nu \in \mathcal{P}(B^d(0,1))$ *satisfy* $\nu(W) \le C_\nu \delta^{\kappa}$ *for all* $(k-1)$*-plates* W *of width* δ. *Fix* $\eta > 0$. *Then there exists a family of* $\lesssim_d C_\nu \delta^{-\eta}$ k*-plates* T_j *of width* $\delta^{\kappa - 2\eta}$ *such that any* k*-plate* W *of width* δ *with* $\nu(W) \ge \delta^{\eta}$ *is contained in one of the plates* T_j.

Proof. We construct a sequence $Y_1, \dots, Y_M$ of k-plates of width δ as follows. To begin, pick (if it exists) a k-plate Y_1 of width δ such that $\nu(Y_1) \ge \delta^{\eta}$. Now suppose $Y_1, \dots, Y_m$ are k-plates of width δ such that:

(1) $\nu(Y_i) \ge \delta^{\eta}$.

(2) $\nu(Y_i \cap Y_j) \le \delta^{2\eta}/2$ for $1 \le i < j \le m$.

If there exists Y_{m+1} such that the collection $(Y_1, \dots, Y_{m+1})$ still satisfies properties (1)–(2), we add it to the list; otherwise we stop. A standard L^2 argument implies that we have to stop after $\le 2\delta^{-\eta}$ steps. Indeed, if $f = \sum_{i=1}^m \mathbf{1}_{Y_i}$, then

$$\begin{aligned}\Big(\sum_i \nu(Y_i)\Big)^2 = \Big(\int f \, d\nu\Big)^2 &\le \int f^2 \, d\nu \\ &= \sum_{1 \le i,j \le m} \nu(Y_i \cap Y_j) < \Big(\sum_i \nu(Y_i)\Big) + m^2\delta^{2\eta}/2.\end{aligned}$$

Let $S = \sum_{i=1}^m \nu(Y_i) \ge m\delta^{\eta}$. We have seen that $S^2 - S < m^2\delta^{2\eta}/2$. If $m > 2\delta^{-\eta}$, then $S^2 - S \ge S^2/2$, and we deduce that

$$m^2 \frac{1}{2}\delta^{2\eta} > \frac{1}{2}\Big(\sum_i \nu(Y_i)\Big)^2 \ge \frac{1}{2} m^2 \delta^{2\eta},$$

which is a contradiction. Let, then, $(Y_i)_{i=1}^M$ be the final family so obtained, with $M \le 2\delta^{-\eta}$.

Now fix a k-plate W of width δ such that $\nu(W) \ge \delta^{\eta}$. Hence

$$\nu(W \cap Y_j) \ge \delta^{2\eta}/2 \quad \text{for some } Y_j,$$

for otherwise we could add W to the list of Y_i. On the other hand, $W \cap Y_j$ is contained in a box of size at most a constant C_d times

$$\underbrace{1 \times \cdots \times 1}_{k-1 \text{ times}} \times \delta/\angle(W, Y_j) \times \underbrace{\delta \times \cdots \times \delta}_{d-k \text{ times}},$$

where $\angle(\cdot,\cdot)$ denotes the largest principal angle between the k-planes determining the corresponding plates. Hence the assumption on ν yields

$$\nu(W \cap Y_j) \lesssim_d \frac{C_\nu \delta^\kappa}{\angle(W, Y_j)}.$$

Comparing the upper and lower bounds on $\nu(W \cap Y_j)$, we deduce that $\angle(W, Y_j) \lesssim_d C_\nu \delta^{\kappa-2\eta}$, and therefore W is contained in the plate T_j centered at (the plane defining) Y_j but of width $\lesssim_d C_\nu \delta^{\kappa-2\eta}$. This is what we wanted to prove. ∎

Proof of Proposition B.2. Let

$$\mathbf{Bad} = \{x \in B^d(0,1) : \nu(W) \geq \delta^\eta \text{ for some } W \in \mathcal{T}_k(x,\delta)\}.$$

Let T_j be the plates provided by Lemma B.3, and define

$$\mathbf{Bad\,Bad} = \{x \in \mathbf{Bad} : x \in T_i \cap T_j \text{ for some } i, j \text{ such that } \angle(T_i, T_j) \geq \delta^\eta\}.$$

Let $r = \delta^{k-2\eta}$. As before, $T_i \cap T_j$ is contained in the union of $\lesssim_d 1$ boxes of size

$$\underbrace{1 \times \cdots \times 1}_{k-1 \text{ times}} \times r/\angle(T_i, T_j) \times \underbrace{r \times \cdots \times r}_{d-k \text{ times}},$$

and hence, now using the non-concentration assumption on μ,

$$\begin{aligned}\mu(\mathbf{Bad\,Bad}) &\leq \sum \{\mu(T_i \cap T_j) : \angle(T_i, T_j) \geq \delta^\eta\} \\ &\lesssim_d C_\mu C_\nu \delta^{-3\eta} \delta^{(\kappa-2\eta)\alpha} = C_\mu C_\nu \delta^{\kappa\alpha-\eta(3+2\kappa)}.\end{aligned}$$

Hence, if η is small enough in terms of κ, α, and δ is small enough in terms of C_μ, C_ν and d, then

$$\mu(\mathbf{Bad\,Bad}) \leq \delta^{\kappa\alpha/2}/2.$$

Let $E = \operatorname{supp}(\mu) \setminus \mathbf{Bad\,Bad}$, and fix $x \in E$. If $x \notin \mathbf{Bad}$, we can just take $P(x) = \emptyset$, so assume that $x \in \mathbf{Bad}$. By Lemma B.3, $\mathbf{Bad}$ is covered by the $2\delta^{-\eta}$ plates $T_j \in \mathcal{T}_k(E, O(C_\nu)\delta^{\kappa-2\eta})$. Let $E_j = E \cap T_j \setminus \bigcup_{i=1}^{j-1} T_i$, and for $x \in E_j$ let $P(x)$ be the $\delta^{\eta/2}$-plate centered at T_j. By passing to a subset of E, we can ensure that $E \cap \mathbf{Bad}$ and the E_j are compact, and still $\mu(\mathbb{R}^d \setminus E) \leq \delta^{\kappa\alpha/2}$. This ensures that $\{(x, y) : x \in E,\ y \notin P(x)\}$ is compact.

Now fix $x \in E_j$ and suppose $W \in \mathcal{T}_k(x,\delta) \cap \mathcal{T}_k(\operatorname{supp}(\nu) \setminus P(x), \delta)$. Then, provided η is taken small enough that $\kappa - 2\eta \geq \eta$ and δ is small, the plate W cannot be contained in any of the plates T_i making an angle smaller than δ^η with T_j. Since $x \in T_j \setminus \mathbf{Bad\,Bad}$, we conclude that the plate W cannot be contained in any of the T_i. Lemma B.3 now implies that $\nu(W) \leq \delta^\eta$, completing the proof. ∎

We can now complete the proof of Theorem B.1.

Proof of Theorem B.1. Let $\eta > 0$ be the number given by Proposition B.2. Making η slightly smaller if needed, we may assume that $\eta/2 = 2^{-N}$ for some integer N. We will

prove the claim with $\eta/2$ in place of η. We take $c = 1/(2N)$, so by assumption there is $\delta_0 > 0$ such that $\nu(W) \leq \varepsilon/(2N)$ for all k-plates W of width $\delta_0^{\eta/2}$. Making δ_0 smaller if needed, we can also ensure that

$$\sum_{n=1}^{\infty} \delta_0^{2^n \cdot \eta} < \varepsilon/2. \tag{B.3}$$

Finally, we also take δ_0 small enough that Proposition B.2 applies to all $\delta \leq \delta_0$.

Write $\delta_n = \delta_0^{2^n}$. For each n, let E_n and $P_n(x), x \in E_n$ be the sets provided by Proposition B.2 for the scales δ_n. Let

$$E = \bigcap_{n=0}^{\infty} E_n, \quad K(x) = \operatorname{supp}(\nu) \setminus \bigcup_{n=0}^{\infty} P_n(x).$$

Then $\{(x, y) : x \in E,\ y \in K(x)\}$ is compact as an intersection of compact sets, and $\nu(E) \geq 1 - \varepsilon$ by (B.3), taking $\eta < \alpha\kappa/2$. Fix $x \in E$ for the rest of the proof.

By Proposition B.2, and choosing for each $r < \delta_1$ the n such that $\delta_n^2 = \delta_{n-1} < r \leq \delta_n$, we get $\nu(K(x) \cap W) \leq r^{\eta/2}$. This yields (B.2).

It remains to show that $\nu(K(x)) \geq 1 - \varepsilon$. By construction, $\nu(P_n(x)) \leq \varepsilon/(2N)$ for all $n \leq N$. Recall that $\eta/2 = 2^{-N}$, and hence $\delta_n^{\eta/2} = \delta_{n-N} \leq \delta_1$ for all $n > N$. Applying Proposition B.2 at scale δ_{n-N} with $P_n(x)$ in place of W, we get

$$\nu(P_n(x) \setminus \bigcup_{j=1}^{n-1} P_j(x)) \leq \nu(P_n(x) \setminus P_{n-N}(x)) \leq \delta_{n-N}^{\eta} \quad \text{for all } n > N.$$

Adding up and recalling (B.3), we see that indeed $\nu(K(x)) \geq 1 - \varepsilon$, and this completes the proof. ∎

Acknowledgments. Part of this work was completed while the author was visiting the Universities of St Andrews and Cambridge. I am grateful for hospitality and a productive work environment at both places. I thank K. Héra for pointing out Problem 4 in [25]. I am grateful to V. Lie for useful comments, and for sharing the results from [1] . I also have J. Zahl to thank for useful insights, and in particular for pointing out the relevance of "train-track" examples in incidence counting. Finally, I thank an anonymous referee for helpful comments that helped improve the presentation.

Funding. PS has received funding from a University of St Andrews Global Fellowship and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 803711).

References

[1] Bateman, M., Lie, V.: Extremal configurations for Kakeya problem in 2D. Manuscript in preparation (2019)

[2] Bond, M., Łaba, I., Zahl, J.: Quantitative visibility estimates for unrectifiable sets in the plane. Trans. Amer. Math. Soc. **368**, 5475–5513 (2016) Zbl 1339.28007 MR 3458388

[3] Bourgain, J.: On the Erdős–Volkmann and Katz–Tao ring conjectures. Geom. Funct. Anal. **13**, 334–365 (2003) Zbl 1115.11049 MR 1982147

[4] Bourgain, J.: The discretized sum-product and projection theorems. J. Anal. Math. **112**, 193–236 (2010) Zbl 1234.11012 MR 2763000

[5] Bourgain, J., Furman, A., Lindenstrauss, E., Mozes, S.: Stationary measures and equidistribution for orbits of nonabelian semigroups on the torus. J. Amer. Math. Soc. **24**, 231–280 (2011) Zbl 1239.37005 MR 2726604

[6] Du, X., Zhang, R.: Sharp L^2 estimates of the Schrödinger maximal function in higher dimensions. Ann. of Math. (2) **189**, 837–861 (2019) Zbl 1433.42010 MR 3961084

[7] Falconer, K. J.: Hausdorff dimension and the exceptional set of projections. Mathematika **29**, 109–115 (1982) Zbl 0477.28004 MR 673510

[8] Falconer, K. J.: On the Hausdorff dimensions of distance sets. Mathematika **32**, 206–212 (1986) (1985) Zbl 0605.28005 MR 834490

[9] Guth, L., Iosevich, A., Ou, Y., Wang, H.: On Falconer's distance set problem in the plane. Invent. Math. **219**, 779–830 (2020) Zbl 1430.28001 MR 4055179

[10] Guth, L., Solomon, N., Wang, H.: Incidence estimates for well spaced tubes. Geom. Funct. Anal. **29**, 1844–1863 (2019) Zbl 1460.52017 MR 4034922

[11] He, W.: Orthogonal projections of discretized sets. J. Fractal Geom. **7**, 271–317 (2020) Zbl 1455.28007 MR 4148151

[12] Hochman, M.: On self-similar sets with overlaps and inverse theorems for entropy. Ann. of Math. (2) **180**, 773–822 (2014) Zbl 1337.28015 MR 3224722

[13] Hochman, M., Lindenstrauss, E., Varjú, P.: Manuscript in preparation

[14] Hochman, M., Shmerkin, P.: Local entropy averages and projections of fractal measures. Ann. of Math. (2) **175**, 1001–1059 (2012) Zbl 1251.28008 MR 2912701

[15] Katz, N. H., Tao, T.: Some connections between Falconer's distance set conjecture and sets of Furstenburg type. New York J. Math. **7**, 149–187 (2001) Zbl 0991.28006 MR 1856956

[16] Katz, N. H., Zahl, J.: An improved bound on the Hausdorff dimension of Besicovitch sets in $\mathbb{R}^3$. J. Amer. Math. Soc. **32**, 195–259 (2019) Zbl 1403.42020 MR 3868003

[17] Kaufman, R.: On Hausdorff dimension of projections. Mathematika **15**, 153–155 (1968) Zbl 0165.37404 MR 248779

[18] Keleti, T., Shmerkin, P.: New bounds on the dimensions of planar distance sets. Geom. Funct. Anal. **29**, 1886–1948 (2019) Zbl 1428.28005 MR 4034924

[19] Liu, B.: An L^2-identity and pinned distance problem. Geom. Funct. Anal. **29**, 283–294 (2019) Zbl 1416.28005 MR 3925111

[20] Liu, B.: Hausdorff dimension of pinned distance sets and the L^2-method. Proc. Amer. Math. Soc. **148**, 333–341 (2020) Zbl 1444.28002 MR 4042855

[21] Liu, B., Shen, C.-Y.: Intersection between pencils of tubes, discretized sum-product, and radial projections. arXiv:2001.02551 (2020)

[22] Mattila, P.: Fourier Analysis and Hausdorff Dimension. Cambridge Stud. Adv. Math. 150, Cambridge Univ. Press, Cambridge (2015) Zbl 1332.28001 MR 3617376

[23] Orponen, T.: On the distance sets of Ahlfors–David regular sets. Adv. Math. **307**, 1029–1045 (2017) Zbl 1355.28018 MR 3590535

[24] Orponen, T.: On the dimension and smoothness of radial projections. Anal. PDE **12**, 1273–1294 (2019) Zbl 1405.28011 MR 3892404

[25] Problems and solutions. The 20-th annual Vojtěch Jarník international mathematical competition. https://vjimc.osu.cz/storage/uploads/j20solutions2.pdf (2010)

[26] Shmerkin, P.: On distance sets, box-counting and Ahlfors regular sets. Discrete Anal. **2017**, art. 9, 22 pp. Zbl 1404.28019 MR 3667933

[27] Shmerkin, P.: On the Hausdorff dimension of pinned distance sets. Israel J. Math. **230**, 949–972 (2019) Zbl 1414.28011 MR 3940442

[28] Shmerkin, P.: Improved bounds for the dimensions of planar distance sets. J. Fractal Geom. **8**, 27–51 (2021) Zbl 1467.28004 MR 4226184