Finite Euclidean configurations in sets of positive density

Vjekoslav Kovač

University of Zagreb

Simons summer school
Harmonic analysis and applications to Ramsey theory

Budapest, June 2026

Last update: June 5, 2026.

Abstract

Euclidean density theorems ask which finite point configurations must occur in measurable subsets of Euclidean space that are large in the sense of positive measure or positive upper density. Motivated by the Ramsey-theoretic principle that large structures contain patterns, the central problems are to prove either that all sufficiently large scales of a configuration appear or, in compact settings, that the set of realized scales contains an interval whose length can be bounded quantitatively in terms of the density. Beginning with Székely's distance problem and Bourgain's simplex theorem, the area has developed into a meeting point of geometric measure theory, arithmetic combinatorics, and real multilinear harmonic analysis. The basic analytic strategy is to encode configurations using counting forms and decompose them into structured, error, and uniform components, with the latter two controlled through Fourier decay, oscillatory integrals, Littlewood–Paley theory, and multilinear singular or Brascamp–Lieb-type inequalities. Conversely, negative results are often proved by constructing funny colorings.

Contents

Chapter 1

What are Euclidean density theorems?

1.1 The guiding problems

A typical Euclidean density theorem has one of the following forms. Let P be a finite configuration in $\mathbb{R}^d$, and let $A \subseteq \mathbb{R}^d$ be measurable and "large." Must A contain a certain copy of P? Or copies of P at all sufficiently large scales? Or copies of P at all scales forming an interval? There are several other choices hidden in all these formulations:

- the meaning of "large": positive fraction of a cube or a ball, positive upper density, or positive upper Banach density;
- the allowed "copies": translates, similar copies, congruent copies, or isometric copies but in another norm;
- the "rigidity" of the configuration: fixed edge lengths, flexible graph embeddings, prescribed volume, etc.;
- the definition of a "scale": a certain size of a configuration, an eccentricity, etc.

Positive results are usually obtained when the configuration has enough degrees of freedom relative to the ambient dimension. Negative results exploit algebraic invariants, modular obstructions, or appropriate colorings of the ambient space.

Throughout the notes, d denotes the ambient Euclidean dimension. The letter A usually denotes the measurable set in which configurations are being sought, whether $A \subseteq \mathbb{R}^d$ is unbounded or A lies in a fixed cube. When a positive-density set is localized to a large cube, the localized set is denoted by $A_R \subseteq [0,R]^d$ and its indicator by $f = \mathbb{1}_{A_R}$.

Euclidean balls are denoted by $\mathrm{B}(a,R) = \{x \in \mathbb{R}^d : |x-a| < R\}$, and $\mathrm{B}_R = \mathrm{B}(0,R)$, while cubes are rather written explicitly as $x + [0,R]^d$. The Lebesgue measure of $A \subseteq \mathbb{R}^d$ will be written simply as $|A|$. This will cause no confusion with the absolute value or the Euclidean length or the cardinality, and the (implicit) ambient dimension d will be understood.

See Section A.1 for the explanation of the notation $\lesssim$, $\gtrsim$, O, and Ω.

1.2 Density notions

Definition 1.1 (Asymptotic densities). For a measurable set $A \subseteq \mathbb{R}^d$, define

$$\overline{\mathrm{d}}(A) := \limsup_{R\to\infty} \frac{|A \cap [-R/2, R/2]^d|}{R^d}, \qquad \underline{\mathrm{d}}(A) := \liminf_{R\to\infty} \frac{|A \cap [-R/2, R/2]^d|}{R^d},$$

called the *upper density* and the *lower density* of A, respectively. If the actual limit exists, it is denoted $\mathrm{d}(A)$ and called the *density* of A.

Definition 1.2 (Upper Banach density). The *upper Banach density* of a measurable set $A \subseteq \mathbb{R}^d$ is

$$\overline{\delta}(A) := \lim_{R\to\infty} \sup_{x\in\mathbb{R}^d} \frac{|A \cap (x + [0,R]^d)|}{R^d}.$$

The existence of this limit, and its invariance under replacing the cube by any fixed compact convex body with nonempty interior, are proved in Appendix A.2. Thus one may equivalently use large balls, ellipsoids, simplices, or other convex averaging windows $\mathcal{B}$, as long as we are stretching them isotropically by factor R:

$$R\mathcal{B} := \{Ry : y \in \mathcal{B}\},$$

$$\overline{\delta}(A) = \lim_{R\to\infty} \sup_{x\in\mathbb{R}^d} \frac{|A \cap (x + R\mathcal{B})|}{|R\mathcal{B}|}.$$

We always have

$$\underline{\mathrm{d}}(A) \leqslant \overline{\mathrm{d}}(A) \leqslant \overline{\delta}(A).$$

Definition 1.3 (Compact density). If $A \subseteq [0,R]^d$ is measurable, its *density in the cube* is $|A|/R^d$. Compact density theorems usually assume $|A| \geqslant \delta R^d$ and produce quantitative information depending on δ. Similarly we would define density in a fixed ball.

1.3 A near-full-density translation lemma

The following elementary lemma is often the quickest way to pass from a high density of a set on a fixed ball to a translated copy of a finite pattern. It is simply the union bound applied to the complement, but it is useful enough to be recorded explicitly.

Lemma 1.4 (Near-full-density translation lemma). *Let $A \subseteq \mathbb{R}^d$ be measurable, let $\mathrm{B}(a,R)$ be the Euclidean ball of radius R centered at a, and suppose that*

$$\frac{|A \cap \mathrm{B}(a,R)|}{|\mathrm{B}(a,R)|} \geqslant 1 - \eta.$$

Let $P = \{p_1, \dots, p_n\} \subseteq \mathbb{R}^d$, and put $D = \operatorname{diam} P$. If

$$D < R \qquad \text{and} \qquad n\eta < \left(1 - \frac{D}{R}\right)^d, \tag{1.1}$$

then A contains a translate of P.

Proof. Translate P in advance so that $p_1 = 0$. Then $|p_i| \leqslant D$ for every i. Let

$$H = \mathrm{B}(a,R) \setminus A, \qquad B' = \mathrm{B}(a, R-D).$$

If $t \in B'$, then $t + p_i \in \mathrm{B}(a,R)$ for each i. Suppose, for contradiction, that no translate $t + P$ is contained in A. Then for every $t \in B'$ at least one of the points $t + p_i$ lies in H, and hence

$$B' \subseteq \bigcup_{i=1}^{n} (H - p_i).$$

Taking measures and using translation-invariance gives

$$|B'| \leqslant \sum_{i=1}^{n} |H - p_i| = n|H| \leqslant n\eta|\mathrm{B}(a,R)|.$$

On the other hand,

$$|B'| = (1 - D/R)^d |\mathrm{B}(a,R)|,$$

which contradicts (1.1). Therefore some $t \in B'$ satisfies $t + p_i \in A$ for all i. □

Corollary 1.5 (Positive measure gives small homothetic copies). *Let $A \subseteq \mathbb{R}^d$ be measurable with $|A| > 0$, and let $P \subseteq \mathbb{R}^d$ be finite. Then there is $\lambda_0 > 0$ such that, for every $0 < \lambda \leqslant \lambda_0$, the set A contains a translate of λP.*

Proof. Let $n = |P|$ and choose $\eta < 1/(n2^d)$. By the Lebesgue density theorem, there is a density point a of A and a (small) radius $R > 0$ such that

$$|A \cap \mathrm{B}(a,R)| \geqslant (1 - \eta)|\mathrm{B}(a,R)|.$$

Lemma 1.4 applies to λP whenever $\lambda \operatorname{diam} P \leqslant R/2$. □

Things are not so easy for infinite configurations P, but we do not discuss those issues in these notes.

Corollary 1.6 (The threshold $1 - 1/n$ for translated copies). *Let $P \subseteq \mathbb{R}^d$ be an n-point configuration. If $A \subseteq \mathbb{R}^d$ is measurable and*

$$\overline{\delta}(A) > 1 - \frac{1}{n},$$

then, for every $\lambda > 0$, the set A contains a translate of λP.

Proof. Choose η with

$$1 - \overline{\delta}(A) < \eta < \frac{1}{n}.$$

By the ball formulation of upper Banach density proved in Appendix A.2, there are arbitrarily large balls $\mathrm{B}(a,R)$ satisfying

$$|A \cap \mathrm{B}(a,R)| \geqslant (1 - \eta)|\mathrm{B}(a,R)|.$$

For the fixed configuration λP, let $D = \operatorname{diam}(\lambda P)$. Since $\eta < 1/n$, we can choose such a ball with R so large that (1.1) holds. Lemma 1.4 then gives a translate of λP inside A. □

Density theorems for simplices, boxes, and distance graphs

This chapter collects positive model results for isotropic configurations, i.e., those that will be scaled equally in "all directions." We will principally study all-large-scale theorems: every measurable set of positive upper Banach density contains the prescribed configuration at every sufficiently large scale. We will also briefly discuss their "interval of scales" variants, which are interesting mainly because they have quantitative formulations.

The examples gathered here will also provide the basic prototypes for the later chapters. The order of problems is deliberately methodological, and only approximately historically faithful. After a general "abstract nonsense" discussion, we begin with two-point distances, then move to simplices, rectangular boxes, and hypercube graphs. Historically, the first large-distance theorem (for a pair of points) was motivated by a question of Székely [49], and it is due to Furstenberg–Katznelson–Weiss [22] and also (independently) to Falconer–Marstrand [20]. Bourgain introduced the analytic large-scale strategy for simplices [3], and Lyall–Magyar developed product and graph versions for simplices, boxes, and distance graphs [40, 41, 42]. The proof language used below is the so-called *largeness–smoothness method*, close in spirit to Bourgain's original argument and to later multilinear formulations of Cook–Magyar–Pramanik and their heat-flow variants of Durcik and the author [8, 10, 31].

2.1 The largeness–smoothness template

Much of the following is probably incomprehensible until one sees concrete examples, so either proceed with caution, or skip to the next section (but be sure to come back).

For a geometric family $\mathcal{C}_\lambda$, let $\mathcal{N}_\lambda^0$ denote the exact *counting form*. Positivity of $\mathcal{N}_\lambda^0(A)$ means that A contains a copy of $\mathcal{C}_\lambda$. The exact measure (or measures) involved in the definition of the counting form is usually singular: it is a spherical measure, an iterated product of spherical measures, or a measure constrained by several distance equations. One therefore introduces a *smoothed count* $\mathcal{N}_\lambda^\varepsilon(A)$, $0 < \varepsilon \leqslant 1$, and writes

$$\mathcal{N}_\lambda^0 = \mathcal{N}_\lambda^1 + \big(\mathcal{N}_\lambda^\varepsilon - \mathcal{N}_\lambda^1\big) + \big(\mathcal{N}_\lambda^0 - \mathcal{N}_\lambda^\varepsilon\big). \tag{2.1}$$

The terms in (2.1) have different meanings.

- The *structured part* $\mathcal{N}_\lambda^1$ is blurred at the same scale as the configuration. At this resolution singularity/curvature is no longer important, and density alone forces many configurations (usually via certain combinatorial "counting").
- The *error part* $\mathcal{N}_\lambda^\varepsilon - \mathcal{N}_\lambda^1$ compares two positive smoothing scales. It is rarely small at every scale, but heat-flow cancellation and lacunarity make it small on average over a long list of scales.

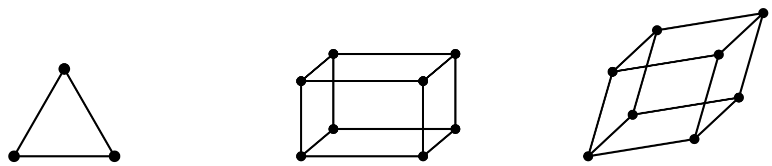

Figure 2.1: Typical configurations in the theory: a simplex, a rectangular box, and a distance (hypercube) graph.

- The *uniform part* $\mathcal{N}_\lambda^0 - \mathcal{N}_\lambda^\varepsilon$ removes the final smoothing. It is controlled uniformly in λ by Fourier decay, some other oscillatory phenomena, or by a multilinear smoothing inequality.

This decomposition was already implicit in Bourgain's Euclidean Szemerédi theorem [3] and has become a standard organizing principle in later multilinear work [8, 12, 10, 31].

Proposition 2.1 (Abstract all-large-scales criterion). *Let $\lambda_1 < \cdots < \lambda_J$ be lacunary, say $\lambda_{j+1} \geqslant 2\lambda_j$. Suppose that, for every $A_R \subseteq [0,R]^d$ with $|A_R| \geqslant \delta R^d$ and all R sufficiently large compared with λ_J, one has*

$$\mathcal{N}_\lambda^1(A_R) \geqslant c_0(\delta) R^d,$$

$$\frac{1}{J}\sum_{j=1}^{J} \left|\mathcal{N}_{\lambda_j}^\varepsilon(A_R) - \mathcal{N}_{\lambda_j}^1(A_R)\right| \leqslant \eta_\varepsilon(J) R^d, \qquad \eta_\varepsilon(J) \to 0 \text{ as } J \to \infty,$$

$$\left|\mathcal{N}_\lambda^0(A_R) - \mathcal{N}_\lambda^\varepsilon(A_R)\right| \leqslant u(\varepsilon) R^d, \qquad u(\varepsilon) \to 0 \text{ as } \varepsilon \to 0.$$

Then every measurable set of positive upper Banach density contains the corresponding configuration at all sufficiently large scales.

Proof. Fix a measurable set A with $\overline{\delta}(A) > 0$ and choose δ with $0 < \delta < \overline{\delta}(A)$. We prove that the set of "bad scales" is bounded. Choose $\varepsilon > 0$ so small that

$$u(\varepsilon) \leqslant \frac{1}{3} c_0(\delta),$$

and then choose J so large that

$$\eta_\varepsilon(J) \leqslant \frac{1}{3} c_0(\delta).$$

Assume, toward a contradiction, that there are arbitrarily large bad scales. We may then select bad scales

$$\lambda_1 < \lambda_2 < \cdots < \lambda_J, \qquad \lambda_{j+1} \geqslant 2\lambda_j,$$

with λ_1 as large as needed for the estimates. By the definition of upper Banach density there are arbitrarily large cubes on which A has density at least δ. Choose such a cube $Q = x_0 + [0,R]^d$ with R larger than the threshold in the hypotheses, i.e., $R \gg \lambda_J$. After translating, set

$$A_R = (A - x_0) \cap [0,R]^d, \qquad |A_R| \geqslant \delta R^d.$$

Because each λ_j is bad for A, the localized exact count is zero:

$$\mathcal{N}_{\lambda_j}^0(A_R) = 0, \qquad j = 1, \ldots, J.$$

Indeed, any configuration counted inside A_R would translate back to a configuration of scale λ_j inside A. The averaged error estimate implies, by pigenoholing, that at least one index $j \in \{1, \dots, J\}$ satisfies

$$\left|\mathcal{N}^{\varepsilon}_{\lambda_j}(A_R) - \mathcal{N}^{1}_{\lambda_j}(A_R)\right| \leqslant \eta_\varepsilon(J) R^d \leqslant \frac{1}{3} c_0(\delta) R^d.$$

For this index j, the decomposition (2.1) gives

$$\begin{aligned}\mathcal{N}^{0}_{\lambda_j}(A_R) &\geqslant \mathcal{N}^{1}_{\lambda_j}(A_R) - \left|\mathcal{N}^{\varepsilon}_{\lambda_j}(A_R) - \mathcal{N}^{1}_{\lambda_j}(A_R)\right| - \left|\mathcal{N}^{0}_{\lambda_j}(A_R) - \mathcal{N}^{\varepsilon}_{\lambda_j}(A_R)\right| \\ &\geqslant c_0(\delta) R^d - \frac{1}{3} c_0(\delta) R^d - \frac{1}{3} c_0(\delta) R^d > 0,\end{aligned}$$

contradicting the badness of λ_j. Thus the bad scales are bounded. The corresponding threshold for "sufficiently large scales" may depend on A, but it is finite. □

2.2 Basic analytic tools: heat flow, curvature, and lacunarity

The Fourier normalization is

$$\widehat{f}(\xi) = \int_{\mathbb{R}^d} f(x) e^{-2\pi i x \cdot \xi} \, \mathrm{d}x.$$

We use the Gaussian

$$g(x) = e^{-\pi |x|^2},$$

with derivatives $h^{(l)} = \partial_l g$ and Laplacian $k = \Delta g$. Thus

$$\widehat{g}(\xi) = e^{-\pi|\xi|^2}, \qquad \widehat{h^{(l)}}(\xi) = 2\pi i \xi_l e^{-\pi|\xi|^2}, \qquad \widehat{k}(\xi) = -4\pi^2 |\xi|^2 e^{-\pi|\xi|^2}.$$

For $\alpha > 0$ write $f_\alpha(x) = \alpha^{-d} f(x/\alpha)$, so $\widehat{f_\alpha}(\xi) = \widehat{f}(\alpha\xi)$. In this parametrization the heat equation reads

$$\frac{\partial}{\partial t} g_t(x) = \frac{1}{2\pi t} k_t(x). \tag{2.2}$$

Curvature enters through Fourier decay. For the normalized circle measure σ in $\mathbb{R}^2$,

$$|\widehat{\sigma}(\xi)| \lesssim (1 + |\xi|)^{-1/2}; \tag{2.3}$$

see Appendix A.3. Therefore,

$$|\widehat{\sigma}(\xi) \widehat{k}(t\xi)| \lesssim (1 + |\xi|)^{-1/2} (t|\xi|)^2 e^{-\pi t^2 |\xi|^2} \lesssim t^{1/2}, \qquad 0 < t \leqslant 1. \tag{2.4}$$

The factor $t^{1/2}$ in (2.4) will consequently be turned into the uniform-part gain $\varepsilon^{1/2}$, which will be typical for this whole chapter. Higher-dimensional analogues are straightforward and they can be found in Appendix A.3.

Finally, lacunarity converts square-function control into a scale average. If $\lambda_{j+1} \geqslant 2\lambda_j$, then, for every $C > 1$, the intervals

$$[\lambda_j, C\lambda_j], \quad j = 1, 2, \dots, J,$$

have uniformly bounded overlap. In Fourier variables this is turned into the elementary estimate

$$\sup_{\xi \in \mathbb{R}^d} \sum_{j=1}^{J} |\widehat{k}(\lambda_j \xi)|^2 \lesssim \sup_{\xi \in \mathbb{R}^d} \sum_{j=1}^{J} (\lambda_j |\xi|)^4 e^{-2\pi \lambda_j^2 |\xi|^2} \lesssim 1, \tag{2.5}$$

because the function $u^4 e^{-2\pi u^2}$ is concentrated where $u \sim 1$ and the numbers $\lambda_j |\xi|$ again form a lacunary sequence. This is the basic Littlewood–Paley ingredient behind the error in this chapter.

2.3 A pair of points: Székely's problem

The first benchmark is the two-point configuration.

Theorem 2.2 (Falconer–Marstrand; Furstenberg–Katznelson–Weiss [20, 22]). *Let $A \subseteq \mathbb{R}^2$ be measurable with $\overline{\delta}(A) > 0$. Then there is $\lambda_0 = \lambda_0(A) > 0$ such that for every $\lambda \geqslant \lambda_0$ there exist $x, x' \in A$ with*

$$|x - x'| = \lambda.$$

For a compactly supported bounded measurable function f define

$$\boxed{\mathcal{N}^0_\lambda(f) := \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} f(x)f(x+y)\,\mathrm{d}\sigma_\lambda(y)\,\mathrm{d}x = \int_{\mathbb{R}^2} f(x)(f * \sigma_\lambda)(x)\,\mathrm{d}x,}$$

$$\boxed{\mathcal{N}^\varepsilon_\lambda(f) := \int_{\mathbb{R}^2} f(x)\big(f * \sigma_\lambda * g_{\varepsilon\lambda}\big)(x)\,\mathrm{d}x.}$$

Here σ_λ is the pushforward of σ under dilation by λ; this is simply the normalized arclength measure on the circle of radius λ about the origin. Thus, $\mathcal{N}^0_\lambda(f) > 0$ means that the set contains a pair of points separated by exactly λ.

Proposition 2.3 (Estimates for a pair of points). *Let $A_R \subseteq [0, R]^2$, $|A_R| \geqslant \delta R^2$, and $R \gg \lambda_J$. For every lacunary list $\lambda_1 < \cdots < \lambda_J$ and every $0 < \varepsilon \leqslant 1$,*

$$\mathcal{N}^1_\lambda(\mathbb{1}_{A_R}) \gtrsim \delta^2 R^2, \tag{2.6}$$

$$\sum_{j=1}^{J} |\mathcal{N}^\varepsilon_{\lambda_j}(\mathbb{1}_{A_R}) - \mathcal{N}^1_{\lambda_j}(\mathbb{1}_{A_R})| \lesssim \log(1/\varepsilon) J^{1/2} R^2, \tag{2.7}$$

$$|\mathcal{N}^0_\lambda(\mathbb{1}_{A_R}) - \mathcal{N}^\varepsilon_\lambda(\mathbb{1}_{A_R})| \lesssim \varepsilon^{1/2} R^2. \tag{2.8}$$

Proof. We prove the three estimates in the same order in which they are used.

Structured part. The smoothed measure $\sigma * g$ is continuous and strictly positive on compact sets. Hence, after dilation,

$$\sigma_\lambda * g_\lambda \gtrsim \lambda^{-2}\mathbb{1}_{[-\lambda,\lambda]^2}.$$

Partition $[0, R]^2$ into $\sim (R/\lambda)^2$ many squares Q of side length $\leqslant \lambda$. If $x, y \in Q$, then $\|x-y\|_\infty \leqslant \lambda$. Therefore,

$$\mathcal{N}^1_\lambda(\mathbb{1}_{A_R}) \gtrsim \lambda^{-2}\sum_Q |A_R \cap Q|^2.$$

Since the number of squares is $\sim (R/\lambda)^2$, Cauchy's inequality gives

$$\sum_Q |A_R \cap Q|^2 \geqslant \frac{(\sum_Q |A_R \cap Q|)^2}{(R/\lambda)^2} \gtrsim \frac{\delta^2 R^4}{(R/\lambda)^2} = \delta^2 R^2 \lambda^2.$$

This proves (2.6).

Uniform part. By Plancherel,

$$\mathcal{N}_\lambda^0(f) - \mathcal{N}_\lambda^\varepsilon(f) = \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 \widehat{\sigma}(\lambda\xi)\big(1 - \widehat{g}(\varepsilon\lambda\xi)\big)\,\mathrm{d}\xi.$$

Recall (2.3), so that

$$|\widehat{\sigma}(\lambda\xi)| \lesssim (1 + \lambda|\xi|)^{-1/2},$$

while $|1 - e^{-\pi\varepsilon^2\lambda^2|\xi|^2}| \lesssim \min\{1, \varepsilon^2\lambda^2|\xi|^2\}$. Bounding the product

$$(1+u)^{-1/2}\min\{1, \varepsilon^2 u^2\} \lesssim \varepsilon^{1/2},$$

see Appendix A.3, we get

$$|\mathcal{N}_\lambda^0(\mathbb{1}_{A_R}) - \mathcal{N}_\lambda^\varepsilon(\mathbb{1}_{A_R})| \lesssim \varepsilon^{1/2}\,\|\mathbb{1}_{A_R}\|_2^2 \leqslant \varepsilon^{1/2}R^2,$$

so (2.8) follows.

Error part. The heat equation (2.2) gives

$$\mathcal{N}_\lambda^\varepsilon(f) - \mathcal{N}_\lambda^1(f) = -\frac{1}{2\pi}\int_\varepsilon^1 \langle f, f * \sigma_\lambda * k_{t\lambda}\rangle\,\frac{\mathrm{d}t}{t}.$$

Fix $t \in [\varepsilon, 1]$ and set

$$T_{j,t}f := f * \sigma_{\lambda_j} * k_{t\lambda_j}.$$

Cauchy–Schwarz in the scale index gives

$$\sum_{j=1}^J |\langle f, T_{j,t}f\rangle| \leqslant J^{1/2}\,\|f\|_2 \Big(\sum_{j=1}^J \|T_{j,t}f\|_2^2\Big)^{1/2}.$$

By Plancherel, $|\widehat{\sigma}| \leqslant 1$, and (2.5),

$$\begin{aligned}\sum_{j=1}^J \|T_{j,t}f\|_2^2 &= \int_{\mathbb{R}^2} |\widehat{f}(\xi)|^2 \sum_{j=1}^J |\widehat{\sigma}(\lambda_j\xi)|^2 |\widehat{k}(t\lambda_j\xi)|^2\,\mathrm{d}\xi \\ &\lesssim \int_{\mathbb{R}^2} |\widehat{f}(\xi)|^2\,\mathrm{d}\xi = \|f\|_2^2.\end{aligned}$$

Hence

$$\sum_{j=1}^J |\langle \mathbb{1}_{A_R}, T_{j,t}\mathbb{1}_{A_R}\rangle| \lesssim J^{1/2}\,\|\mathbb{1}_{A_R}\|_2^2 \leqslant J^{1/2}R^2.$$

Integrating $\mathrm{d}t/t$ from ε to 1 proves (2.7). $\square$

Proof of Theorem 2.2. Apply Proposition 2.1. The structured function is $c_0(\delta) := c\delta^2$, the error function is $\eta_\varepsilon(J) := C_1\log(1/\varepsilon)J^{-1/2}$, while the uniform function is $u(\varepsilon) := C_2\varepsilon^{c_d}$, where c, C_1, and C_2 are the implicit constants in (2.6), (2.7), and (2.8), respectively. The abstract criterion (Proposition 2.1) gives all sufficiently large distances. Positivity of the exact count produces actual points because the integrand is nonnegative and the measure $\mathrm{d}x\ \mathrm{d}\sigma_\lambda(y)$ is supported precisely on pairs $(x, x+y)$ with $|y| = \lambda$. $\square$

The presented proof is not the shortest possible proof of the density theorem for two points, but it fits well to our more general philosophy.

2.4 Bourgain's simplex theorem

Let $d \geqslant 2$, put $q = d-1$, and let

$$\Delta = \{0, u_1, \ldots, u_q\} \subseteq \mathbb{R}^q$$

be a non-degenerate q-dimensional simplex, meaning that $u_1, \ldots, u_q$ are linearly independent. We view Δ inside $\mathbb{R}^d$ by adding one zero coordinate. The increase from the affine dimension q to the ambient dimension $d = q+1$ supplies one extra degree of rotational freedom and allows one to prove the following.

Theorem 2.4 (Bourgain [3]). *Let $A \subseteq \mathbb{R}^d$ be measurable with $\overline{\delta}(A) > 0$. Then there is $\lambda_0 = \lambda_0(A, \Delta) > 0$ such that, for every $\lambda \geqslant \lambda_0$, the set A contains an isometric copy of $\lambda\Delta$.*

Bourgain established this in [3]. Later papers gave related direct or product-type proofs and extensions [40, 29, 41, 42], some of which we are about to discuss in the following sections.

Counting forms

Let μ be the Haar probability measure on the group $\mathrm{SO}(d)$ of d-dimensional rotations about the origin. For compactly supported measurable $0 \leqslant f \leqslant 1$ set

$$\boxed{\mathcal{N}_\lambda^0(f) := \int_{\mathbb{R}^d} \int_{\mathrm{SO}(d)} f(x) \prod_{k=1}^{q} f(x + \lambda U u_k) \,\mathrm{d}\mu(U)\,\mathrm{d}x,}$$

$$\boxed{\mathcal{N}_\lambda^\varepsilon(f) := \int_{\mathbb{R}^d} \int_{\mathrm{SO}(d)} f(x) \prod_{k=1}^{q} (f * g_{\varepsilon\lambda})(x + \lambda U u_k) \,\mathrm{d}\mu(U)\,\mathrm{d}x.}$$

The exact count is positive precisely when f sees a translated and rotated copy of the simplex at scale λ, i.e., a desired copy of $\lambda\Delta$.

The rotation integral may be written as an iterated integration over spheres. Choose $y_1 = \lambda U u_1$ first; it lies on a sphere. Once $y_1, \ldots, y_{k-1}$ have been chosen, the vector $y_k = \lambda U u_k$ is constrained to a sphere inside the affine plane determined by the prescribed inner products with the previous vectors. The dimension of this sphere decreases by one at each step. The last step is a circle, and it is precisely the decay of this remaining circle that drives the uniform estimate.

Proposition 2.5 (Simplex estimates). *For $A_R \subseteq [0, R]^d$ with $|A_R| \geqslant \delta R^d$ and $R \gg \lambda_J$, one has*

$$\mathcal{N}_\lambda^1(\mathbb{1}_{A_R}) \gtrsim_\Delta \delta^d R^d, \tag{2.9}$$

$$\sum_{j=1}^{J} |\mathcal{N}_{\lambda_j}^\varepsilon(\mathbb{1}_{A_R}) - \mathcal{N}_{\lambda_j}^1(\mathbb{1}_{A_R})| \lesssim_\Delta \varepsilon^{-C_q} J^{1/2} R^d, \tag{2.10}$$

$$|\mathcal{N}_\lambda^0(\mathbb{1}_{A_R}) - \mathcal{N}_\lambda^\varepsilon(\mathbb{1}_{A_R})| \lesssim_\Delta \varepsilon^{1/2} R^d. \tag{2.11}$$

Here C_q is a large constant depending only on q.

Proof. Recall that the ambient dimension is $d = q+1$.

Structured part. Since g is strictly positive, for each fixed vertex vector u_k there is a constant $a_k > 0$ depending only on Δ such that, uniformly in $U \in \mathrm{SO}(d)$,

$$(f * g_\lambda)(x + \lambda U u_k) \gtrsim (f * \varphi_{a_k\lambda})(x).$$

Indeed, if $|z - x| \leqslant a_k\lambda$ with a_k small, then

$$|x + \lambda U u_k - z| \leqslant (|u_k| + a_k)\lambda,$$

so the Gaussian kernel $g_\lambda(x + \lambda U u_k - z)$ is bounded below by a constant multiple of λ^{-d} on the ball $\mathrm{B}(x, a_k\lambda)$. Therefore,

$$\mathcal{N}^1_\lambda(\mathbb{1}_{A_R}) \gtrsim_\Delta \lambda^{-qd} \int \mathbb{1}_{A_R}(x) \prod_{k=1}^{q} (\mathbb{1}_{A_R} * \mathbb{1}_{B(0,a_k\lambda)})(x)\,\mathrm{d}x.$$

Partition $[0,R]^d$ into $N \sim (R/\lambda)^d$ axes-parallel cubes so that each cube has diameter at most

$$\min\{a_1, \dots, a_q\}\lambda/10.$$

Splitting the above integral we get

$$\mathcal{N}^1_\lambda(\mathbb{1}_{A_R}) \gtrsim_\Delta \lambda^{-qd} \sum_Q |A_R \cap Q|^{q+1},$$

which is

$$\sim R^d \lambda^{-d^2} \frac{\sum_Q |A_R \cap Q|^d}{N} \geqslant R^d \lambda^{-d^2} \left(\frac{\sum_Q |A_R \cap Q|}{N} \right)^d \sim R^{-d(d-1)} |A_R|^d \geqslant \delta^d R^d,$$

where we used Jensen's inequality for discrete averages. This proves (2.9).

Error part. By the heat equation, for $0 < \alpha < \beta \leqslant 1$,

$$\mathcal{N}^\alpha_\lambda(f) - \mathcal{N}^\beta_\lambda(f) = -\sum_{m=1}^{q} \int_\alpha^\beta \mathcal{L}^m_{\lambda,t}(f) \frac{\mathrm{d}t}{t},$$

where $\mathcal{L}^m_{\lambda,t}$ is obtained from $\mathcal{N}^t_\lambda$ by replacing the mth smoothed factor $(f * g_{t\lambda})(x + \lambda U u_m)$ by $(f * k_{t\lambda})(x + \lambda U u_m)$ and leaving the other smoothed factors unchanged.

We prove the required averaged bound for one value of m; summing over m only changes the constant. In the iterated-sphere parametrization, choose the order so that the mth vertex is selected last. The last integration is over a circle in a two-dimensional orthogonal plane. Insert the harmless identity

$$1 = \int_{\theta t\lambda_j}^{e\theta t\lambda_j} \frac{\mathrm{d}s}{s}$$

with a fixed small $\theta > 0$. For $s \sim t\lambda_j$, the Gaussian identities imply a decomposition of the cancellative kernel of the form

$$k_{t\lambda_j} = \sum_{l=1}^{d} h^{(l)}_s * h^{(l)}_s * g_{r(j,t,s)}, \qquad r(j,t,s) \sim s, \tag{2.12}$$

up to constants depending only on θ. This follows directly on the Fourier side from

$$|\xi|^2 e^{-\pi(t\lambda_j)^2|\xi|^2} = \sum_{l=1}^{d} \xi_l^2 e^{-2\pi s^2|\xi|^2} e^{-\pi r^2|\xi|^2}$$

with $r^2 = (t\lambda_j)^2 - 2s^2 > 0$.

Using (2.12), one derivative $h_s^{(l)}$ is placed on f and the other derivative is absorbed into the final circular average. The remaining spherical and Gaussian factors are positive probability kernels with rapidly decreasing tails. They are dominated by a finite superposition of ball or Gaussian averages at scale comparable to s, and hence by Hardy–Littlewood maximal functions. (However, since the function $f = \mathbb{1}_A$ we are interested in takes values in $[0,1]$ and the maximal operator is trivially bounded on L^∞, this domination is not really needed.) Applying Cauchy–Schwarz in the base variable and in the already chosen directions gives the estimate

$$|\mathcal{L}^m_{\lambda_j,t}(f)| \lesssim_\Delta \varepsilon^{-C_q} \sum_{l=1}^{d} \left(\int_{\theta t\lambda_j}^{e\theta t\lambda_j} \left\| f * h_s^{(l)} \right\|_2^2 \frac{\mathrm{d}s}{s} \right)^{1/2} R^{d/2}. \tag{2.13}$$

Here the factor ε^{-C_q} comes from comparing all smoothing parameters between $t\lambda_j$ and s while $t \geqslant \varepsilon$; the maximal-function factors are bounded in L^{2q} and use $0 \leqslant f \leqslant 1$ and $\operatorname{supp} f \subseteq [0,R]^d$.

Sum (2.13) over j and apply Cauchy–Schwarz in j. For fixed t, lacunarity gives bounded overlap of the intervals $[\theta t\lambda_j, e\theta t\lambda_j]$, and therefore

$$\sum_{j=1}^{J} |\mathcal{L}^m_{\lambda_j,t}(f)| \lesssim_\Delta \varepsilon^{-C_q} J^{1/2} R^{d/2} \left(\int_0^\infty \sum_{l=1}^{d} \left\| f * h_s^{(l)} \right\|_2^2 \frac{\mathrm{d}s}{s} \right)^{1/2}.$$

Plancherel's identity gives

$$\int_0^\infty \sum_{l=1}^{d} \left\| f * h_s^{(l)} \right\|_2^2 \frac{\mathrm{d}s}{s} \lesssim_d \|f\|_2^2 \leqslant R^d.$$

Thus $\sum_j |\mathcal{L}^m_{\lambda_j,t}(f)| \lesssim \varepsilon^{-C_q} J^{1/2} R^d$. Integrating $\mathrm{d}t/t$ over $[\varepsilon, 1]$ and summing over m proves (2.10), only with C_q slightly enlarged to swallow up the logarithmic loss.

Uniform part. It suffices to control the heat-flow tail between smoothing parameters 0 and ε. Fix the derivative on the final vertex in the iterated-sphere description. After the first $q-1$ simplex directions have been chosen, the last direction ranges over a circle in a two-dimensional plane E. The relevant piece of the heat derivative has the schematic form

$$\int f_0(x)\,(f * k_{t\lambda} * \sigma_{\lambda,E})(x)\,\mathrm{d}x,$$

where $\sigma_{\lambda,E}$ is circle measure in the plane E and $0 \leqslant f_0 \leqslant 1$ is a product of the already selected vertex factors. Cauchy–Schwarz in all outer variables and Plancherel in x reduce the square of this expression to an integral bounded by

$$R^d \int |\widehat{f}(\xi)|^2 |\widehat{k}(t\lambda\xi)|^2 \mathcal{I}(\lambda\xi)\,\mathrm{d}\xi.$$

Here $\mathcal{I}(\zeta)$ is the average, over the possible first $q-1$ directions, of the reciprocal distance from ζ to their span. This factor is exactly what remains from the stationary-phase estimate for the last circle: the Fourier transform of a circle in the plane E decays like the inverse square root of the length of the projection of ζ onto E.

The averaging over rotations removes the singularity. By rotational invariance, fix the span of the first $q-1$ directions to be the coordinate $(q-1)$-plane and average the direction of ζ over S^{d-1}. In spherical coordinates the singular factor is

$$\frac{1}{|\zeta|\sin\phi_1\cdots\sin\phi_{q-1}},$$

whereas the surface element contributes

$$\sin^{q-1}\phi_1\sin^{q-2}\phi_2\cdots\sin\phi_{q-1}\,\mathrm{d}\phi_1\cdots\mathrm{d}\phi_q.$$

The quotient is integrable, hence

$$\mathcal{I}(\zeta)\lesssim_\Delta|\zeta|^{-1}. \tag{2.14}$$

Combining (2.14) with the formula for $\widehat{k}$ gives

$$|\widehat{k}(t\zeta)|^2\mathcal{I}(\zeta)\lesssim(t|\zeta|)^4e^{-2\pi t^2|\zeta|^2}|\zeta|^{-1}\lesssim t,$$

because $u^3e^{-2\pi u^2}$ is bounded. Thus the derivative at scale t is $O(t^{1/2}R^d)$. Integrating from 0 to ε gives

$$|\mathcal{N}^0_\lambda(f)-\mathcal{N}^\varepsilon_\lambda(f)|\lesssim R^d\int_0^\varepsilon t^{1/2}\frac{\mathrm{d}t}{t}\lesssim\varepsilon^{1/2}R^d,$$

which proves (2.11). □

Proof of Theorem 2.4. Use Proposition 2.1 and the estimates in Proposition 2.5. The structured function is $c_0(\delta):=c\delta^d$, the uniform function is $u(\varepsilon):=C\varepsilon^{c_q}$, and the averaged error has $\eta_\varepsilon(J):=C\varepsilon^{-C_q}J^{-1/2}$. After choosing ε and then J as in the abstract criterion, a bad lacunary list of scales would force a positive exact count at one of those scales, which is impossible. Positivity of $\mathcal{N}^0_\lambda(\mathbb{1}_A)$ gives an actual congruent copy of $\lambda\Delta$ because Haar measure is supported on rotations preserving all distances among the vertices of Δ. □

2.5 Rectangular boxes

Let $q\geqslant 1$ and

$$\square=\{0,b_1\}\times\cdots\times\{0,b_q\}\subseteq\mathbb{R}^q,\qquad b_1,\dots,b_q>0.$$

The ambient space is $(\mathbb{R}^2)^q$. The kth side of the box is allowed to rotate only in the kth distinguished copy of $\mathbb{R}^2$.

Theorem 2.6 (Lyall–Magyar, rectangular-box case [40, 42]). *Let $A\subseteq(\mathbb{R}^2)^q$ be measurable with $\overline{\delta}(A)>0$. Then there is $\lambda_0=\lambda_0(A,\square)>0$ such that, for every $\lambda\geqslant\lambda_0$, there exist $x_1,\dots,x_q,y_1,\dots,y_q\in\mathbb{R}^2$ with*

$$|y_k|=\lambda b_k,\qquad k=1,\dots,q,$$

and

$$\big\{(x_1+r_1y_1,\dots,x_q+r_qy_q):(r_1,\dots,r_q)\in\{0,1\}^q\big\}\subseteq A.$$

In particular, A contains an isometric copy of $\lambda\square$.

This is a particular case of Lyall–Magyar's product and hypergraph results [40, 42]. The following heat-flow proof that smoothens up the measures (rather than the functions) is due to Durcik and the author [10, 11], and its advantage is sufficient flexibility to also obtain a certain quantitative compact variant in Theorem 2.8 below.

Counting forms

The exact and smoothed counting forms are

$$\boxed{\mathcal{N}_\lambda^0(f) := \int_{(\mathbb{R}^2)^{2q}} \prod_{(r_1,\dots,r_q)\in\{0,1\}^q} f(x_1+r_1y_1,\dots,x_q+r_qy_q) \prod_{k=1}^q \mathrm{d}\sigma_{\lambda b_k}(y_k) \prod_{k=1}^q \mathrm{d}x_k,}$$

$$\boxed{\mathcal{N}_\lambda^\varepsilon(f) := \int_{(\mathbb{R}^2)^{2q}} \prod_{(r_1,\dots,r_q)\in\{0,1\}^q} f(x_1+r_1y_1,\dots,x_q+r_qy_q) \prod_{k=1}^q (\sigma_{\lambda b_k} * g_{\varepsilon\lambda b_k})(y_k) \prod_{k=1}^q \mathrm{d}x_k\,\mathrm{d}y_k.}$$

For $\mathbf{x} = (x_1^0, x_1^1, \dots, x_q^0, x_q^1) \in (\mathbb{R}^2)^{2q}$ we write

$$\mathcal{F}(\mathbf{x}) := \prod_{(r_1,\dots,r_q)\in\{0,1\}^q} f(x_1^{r_1},\dots,x_q^{r_q}).$$

If we substitute $x_k^0 = x_k$, $x_k^1 = x_k + y_k$, then $\mathcal{N}_\lambda^\varepsilon(f)$ takes a nicer and more symmetric expression:

$$\mathcal{N}_\lambda^\varepsilon(f) := \int_{(\mathbb{R}^2)^{2q}} \mathcal{F}(\mathbf{x}) \prod_{k=1}^q (\sigma * g_\varepsilon)_{\lambda b_k}(x_k^1 - x_k^0)\,\mathrm{d}\mathbf{x}.$$

The variables x_k^0 and x_k^1 represent the two endpoints of the kth side. Choosing all 2^q combinations of endpoints gives all vertices of the box.

Proposition 2.7 (Box estimates). *For $A_R \subseteq ([0,R]^2)^q$ with $|A_R| \geqslant \delta R^{2q}$ and $R \gg \lambda_J$,*

$$\mathcal{N}_\lambda^1(\mathbb{1}_{A_R}) \gtrsim_\square \delta^{2^q} R^{2q}, \tag{2.15}$$

$$\sum_{j=1}^J |\mathcal{N}_{\lambda_j}^\varepsilon(\mathbb{1}_{A_R}) - \mathcal{N}_{\lambda_j}^1(\mathbb{1}_{A_R})| \lesssim_\square \varepsilon^{-C_q} R^{2q}, \tag{2.16}$$

$$|\mathcal{N}_\lambda^0(\mathbb{1}_{A_R}) - \mathcal{N}_\lambda^\varepsilon(\mathbb{1}_{A_R})| \lesssim_\square \varepsilon^{1/2} R^{2q}. \tag{2.17}$$

Here C_q is a large constant depending only on q.

Proof. Structured part. Partition the kth square $[0,R]^2$ in a 2-dimensional coordinate plane $\mathbb{R}^2$ into $\sim (R/\lambda)^2$ squares Q_k of side length $\leqslant b_k\lambda$. In the smoothest form, restrict x_k^0 and x_k^1 to lie in the same square Q_k. Since $(\sigma * g)_{\lambda b_k}$ is bounded below by $\gtrsim \lambda^{-2}$ on a $\sim \lambda$ neighborhood of the origin, this restriction gives:

$$\mathcal{N}_\lambda^1(\mathbb{1}_{A_R}) \gtrsim_\square \sum_{Q_1,\dots,Q_q} |Q_1|\cdots|Q_q| \fint_{Q_1\times Q_1\times\cdots\times Q_q\times Q_q} \mathcal{F}(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

For each product cell and $f \geqslant 0$, we have:

$$\fint_{Q_1\times Q_1\times\cdots\times Q_q\times Q_q} \prod_{r\in\{0,1\}^q} f(x_1^{r_1},\dots,x_q^{r_q})\,\mathrm{d}\mathbf{x} \geqslant \Big(\fint_{Q_1\times\cdots\times Q_q} f\Big)^{2^q}. \tag{2.18}$$

For $q=1$ this is easy by Jensen's inequality. For the induction step, write the left side as an average over the last pair (x_q^0,x_q^1) of a product of two $(q-1)$-dimensional box forms. Cauchy–Schwarz in the variables indexed by the first $q-1$ coordinates bounds this average below by the square of the corresponding $(q-1)$-dimensional average, and the induction hypothesis applies. The reader is encouraged to write up the details and compare them with the proof of Proposition A.5 (the celebrated Gowers–Cauchy–Schwarz inequality) given in the appendix.

Summing (2.18) over all product cells gives

$$\mathcal{N}_\lambda^1(f) \gtrsim_\Box \sum_{Q_1,\dots,Q_q} |Q_1|\cdots|Q_q| \Big(\frac{|A_R\cap(Q_1\times\cdots\times Q_q)|}{|Q_1|\cdots|Q_q|}\Big)^{2^q}.$$

Jensen's inequality over the product partition bounds this as

$$\gtrsim R^{2q}\left(\frac{|A_R|}{R^{2q}}\right)^{2^q},$$

which is (2.15).

Error part. By differentiating the product of the q smoothed circular factors, the contribution of the derivative falling on the mth edge is

$$(\sigma*k_t)_{\lambda b_m}(x_m^0-x_m^1)\prod_{k\neq m}(\sigma*g_t)_{\lambda b_k}(x_k^0-x_k^1).$$

It suffices to treat $m=q$. Insert the average $1=\int_{\theta t\lambda}^{e\theta t\lambda}\mathrm{d}s/s$. For $s\sim t\lambda$, the Fourier identity for the Laplacian gives

$$\sigma_{\lambda b_q}*k_{t\lambda b_q} = \sum_{l=1}^{2}\sigma_{\lambda b_q}*g_r*h_s^{(l)}*h_s^{(l)}, \qquad r\sim s\sim t\lambda, \tag{2.19}$$

up to constants depending on $\Box$. The other smoothed circle kernels and the factor $\sigma_{\lambda b_q}*g_r$ are dominated, for $t\geqslant\varepsilon$, by superpositions of Gaussian kernels at scale $s\gamma$, $\gamma\geqslant 1$, with total mass $O_\Box(\varepsilon^{-C_q})$.

Expand the two derivatives in (2.19). The variables with $r_q=0$ form one face of the box and the variables with $r_q=1$ form the opposite face. Applying Cauchy–Schwarz between these two faces converts the absolute value of the derivative term into a positive cubical heat-flow form. More explicitly, one obtains the bound

$$\sum_{j=1}^{J}|\mathcal{L}_{\lambda_j}^q(f)| \lesssim_{\Box,q} \varepsilon^{-C_q}\log(1/\varepsilon)\int_{[1,\infty)^{q-1}}\Theta^{q,q}_{\gamma_1,\dots,\gamma_{q-1},\sqrt{2}}(f)\prod_{k=1}^{q-1}\frac{\mathrm{d}\gamma_k}{\gamma_k^2}, \tag{2.20}$$

where

$$\Theta^{q,m}_{\gamma_1,\dots,\gamma_q}(f) := -\int_0^\infty\int\prod_{r\in\{0,1\}^q} f(x_1^{r_1},\dots,x_q^{r_q})\,k_{s\gamma_m}(x_m^0-x_m^1)$$

$$\times \prod_{k\neq m} g_{s\gamma_k}(x_k^0 - x_k^1)\,\mathrm{d}\mathbf{x}\frac{\mathrm{d}s}{s}.$$

The minus sign is chosen because $k = \Delta g$ and the corresponding quadratic form is nonnegative. Indeed, writing the two m-faces as a function F_m of the common mth coordinate, the identity

$$k_{s\gamma_m} = 2\sum_{l=1}^{2} h^{(l)}_{s\gamma_m/\sqrt{2}} * h^{(l)}_{s\gamma_m/\sqrt{2}}$$

gives

$$\Theta^{q,m}_{\gamma_1,\ldots,\gamma_q}(f) = 2\sum_{l=1}^{2}\int_0^\infty \int \left\|F_m * h^{(l)}_{s\gamma_m/\sqrt{2}}\right\|_2^2 \prod_{k\neq m} g_{s\gamma_k}(x_k^0 - x_k^1)\frac{\mathrm{d}s}{s} \geqslant 0.$$

Summing these nonnegative forms over m telescopes by the heat equation:

$$\sum_{m=1}^{q} \Theta^{q,m}_{\gamma_1,\ldots,\gamma_q}(f) \lesssim \|f\|_{2^q}^{2^q}.$$

To see the telescoping, differentiate the fully Gaussian cubical average

$$\int \prod_{r\in\{0,1\}^q} f(x_1^{r_1},\ldots,x_q^{r_q}) \prod_{k=1}^{q} g_{s\gamma_k}(x_k^0 - x_k^1)\,\mathrm{d}\mathbf{x}$$

with respect to s and integrate from $s=0$ to $s=\infty$. At $s=\infty$ the normalized Gaussians vanish weakly on compactly supported data, while at $s=0$ all paired coordinates coalesce and the limit is $\int f^{2^q} \leqslant R^{2q}$. Since each Θ is nonnegative, (2.20) is bounded by $C\varepsilon^{-C_q}\log(1/\varepsilon)R^{2q}$. Summing over the possible derivative positions proves (2.16). This is the cubical singular-integral estimate in its soft heat-flow form; related estimates appear in [9, 14].

Uniform part. Let the heat derivative fall on the mth edge and keep all other endpoint variables fixed. The product of vertex factors splits into two faces,

$$F_0(u) = \prod_{r_m=0} f(\ldots,u,\ldots), \qquad F_1(v) = \prod_{r_m=1} f(\ldots,v,\ldots).$$

The relevant bilinear form in $u, v \in \mathbb{R}^2$ is

$$\int F_0(u)F_1(v)(\sigma_{\lambda b_m} * k_{t\lambda b_m})(u-v)\,\mathrm{d}u\,\mathrm{d}v.$$

By Plancherel it is equal to

$$\int \widehat{F_0}(\xi)\overline{\widehat{F_1}(\xi)}\widehat{\sigma}(\lambda b_m\xi)\widehat{k}(t\lambda b_m\xi)\,\mathrm{d}\xi.$$

The multiplier is $O(t^{1/2})$ by (2.4). Hence the bilinear form is bounded by $Ct^{1/2}\|F_0\|_2\|F_1\|_2$. After integrating the remaining variables against probability kernels and using $0 \leqslant f \leqslant 1$ and $\operatorname{supp} f \subseteq ([0,R]^2)^q$, this gives

$$|\mathcal{L}^m_{\lambda,t}(f)| \lesssim t^{1/2}R^{2q}.$$

Integrating $0 < t < \varepsilon$ and summing over m proves (2.17). $\square$

Proof of Theorem 2.6. Apply Proposition 2.1. The structured function is $c_0(\delta) := c\delta^{2^q}$, the uniform function is $u(\varepsilon) := C\varepsilon^{c_q}$, and the error estimate can take $\eta_\varepsilon(J) := C\varepsilon^{-C_q}J^{-1}$. The abstract contradiction argument therefore gives $\mathcal{N}^0_\lambda(\mathbb{1}_A) > 0$ at all sufficiently large scales. Positivity of $\mathcal{N}^0_\lambda(\mathbb{1}_A)$ produces the required box because each singular factor is supported on the circle $|x^0_k - x^1_k| = \lambda b_k$. □

With no additional effort one can prove the following compact counterpart to the last (cubical all-large-scale) theorem.

Theorem 2.8 (Quantified cube interval [10, 11]). *For every positive integer q there exists $C'_q \in (0,\infty)$ such that, whenever $A \subseteq ([0,1]^2)^q$ is measurable with $|A| \geqslant \delta$, $0 < \delta \leqslant 1/2$, there is an interval $I \subseteq (0,1]$ of length at least*

$$\exp(-\delta^{-C'_q})$$

such that for every $\lambda \in I$ one can find $x_i, y_i \in \mathbb{R}^2$, $1 \leqslant i \leqslant q$, satisfying

$$(x_1 + r_1 y_1, \ldots, x_q + r_q y_q) \in A$$

for all $(r_1, \ldots, r_q) \in \{0,1\}^q$, with

$$\|y_i\|_2 = \lambda \quad \textit{for all } i.$$

Namely, one considers the intervals $(2^{-j}, 2^{-j+1}]$, $1 \leqslant j \leqslant J$, for a sufficiently large number of scales J. If every one of them contained a scale λ_j at which no configuration is found, then we would obtain a contradiction just as in the previous proof. Thus, in one of these intervals, every scale is realized by a cube in A. Its length is at least $2^{-J} \geqslant \exp(-\delta^{-C})$ and we are done. We omit the details, since the whole next chapter on arithmetic progressions will be concerned with a much more involved compact density theorem.

2.6 The hypercube distance graph

The reader can find this section primarily as useful practice of the previously seen techniques. Let Γ_q be the distance graph isomorphic to the 1-skeleton of the q-dimensional unit hypercube. A planar set A contains an isometric copy of $\lambda\Gamma_q$ if there are $x \in \mathbb{R}^2$ and vectors $y_1, \ldots, y_q \in \mathbb{R}^2$ with $|y_i| = \lambda$ such that

$$x + r_1 y_1 + \cdots + r_q y_q \in A$$

for all $(r_1, \ldots, r_q) \in \{0,1\}^q$, and the 2^q points are all mutually distinct.

Theorem 2.9 (All large dilates of Γ_q [33]). *For every positive integer q and every measurable $A \subseteq \mathbb{R}^2$ of positive upper Banach density, there exists $\lambda_0(A,q) > 0$ such that for every $\lambda \geqslant \lambda_0(A,q)$ the set A contains an isometric copy of $\lambda\Gamma_q$.*

Theorem 2.10 (Compact interval version for Γ_q [33]). *For every q there is $C_q \in (0,\infty)$ such that if $A \subseteq [0,1]^2$ is measurable with $|A| \geqslant \delta$, $0 < \delta \leqslant 1/2$, then there is an interval $I \subseteq (0,1]$ of length at least*

$$\exp(-\delta^{-C_q})$$

such that A contains an isometric copy of $\lambda\Gamma_q$ for every $\lambda \in I$.

These are dimensionally sharp: the graph is embedded in the plane even though its "degeneracy" (as defined in [41]) is q.

Let $f = \mathbb{1}_{A_R}$ and define

$$\mathcal{F}_q(x; y_1, \ldots, y_q) = \prod_{(r_1,\ldots,r_q)\in\{0,1\}^q} f(x + r_1 y_1 + \cdots + r_q y_q).$$

With σ_λ being the circle measure for radius λ,

$$\boxed{\mathcal{N}^0_\lambda(f) := \int_{(\mathbb{R}^2)^{q+1}} \mathcal{F}_q(x; y_1, \ldots, y_q)\, \mathrm{d}\sigma_\lambda(y_1) \cdots \mathrm{d}\sigma_\lambda(y_q)\, \mathrm{d}x.}$$

The smoothed form replaces each σ_λ by $\sigma_\lambda * g_{\varepsilon\lambda}$:

$$\boxed{\mathcal{N}^\varepsilon_\lambda(f) := \int_{(\mathbb{R}^2)^{q+1}} \mathcal{F}_q(x; y_1, \ldots, y_q)\Big(\prod_{k=1}^{q} (\sigma_\lambda * g_{\varepsilon\lambda})(y_k)\Big)\, \mathrm{d}x.}$$

Positivity of $\mathcal{N}^0_\lambda(\mathbb{1}_A)$ implies the existence of a nondegenerate copy of $\lambda\Gamma_q$ because degenerate coincidences are described by finitely many linear relations among $y_1, \ldots, y_q$ and therefore have zero $\sigma_\lambda^{\otimes q}$-measure.

Proposition 2.11 (Hypercube graph estimates [33]). *For $R \geqslant \lambda > 0$ and $A_R \subseteq [0, R]^2$,*

$$\mathcal{N}^1_\lambda(\mathbb{1}_{A_R}) \gtrsim_q \left(\frac{|A_R|}{R^2}\right)^{2^q} R^2. \tag{2.21}$$

For $0 < \lambda_1 < \cdots < \lambda_J$ with $\lambda_{j+1} \geqslant 2\lambda_j$, $R \geqslant 2\lambda_J$, and $\varepsilon \in (0, 1]$,

$$\sum_{j=1}^{J} \Big|\mathcal{N}^\varepsilon_{\lambda_j}(\mathbb{1}_{A_R}) - \mathcal{N}^1_{\lambda_j}(\mathbb{1}_{A_R})\Big| \lesssim_q \varepsilon^{-3q-1} R^2. \tag{2.22}$$

For all $\lambda > 0$ and $\varepsilon \in (0, 1]$,

$$\big|\mathcal{N}^0_\lambda(\mathbb{1}_{A_R}) - \mathcal{N}^\varepsilon_\lambda(\mathbb{1}_{A_R})\big| \lesssim_q \varepsilon^{1/2} R^2. \tag{2.23}$$

The proof is left as an exercise, quite similar to the previous section. A notable difference is that the structured part will need the Gowers–Cauchy–Schwarz inequality (Proposition A.5), which is moved to Appendix A.4, not to spoil the exercise here. Likewise, the reader is invited to deduce Theorems 2.9 and 2.10 from Estimates (2.21), (2.22), and (2.23) following analogous deductions, which have been seen a few times already.

2.7 Open problem

Problem 2.12 (Equilateral triangles in the plane). A classical problem asks whether every planar set of positive upper density contains a congruent copy of every sufficiently large equilateral triangle; see Bourgain [3] and the discussion surrounding the Euclidean obstructions in [19]. In fact, all dimensionally sharp embeddings of simplices, i.e., $d = q \geqslant 2$, are still open.

Density theorems for arithmetic progressions

This chapter treats the compact quantitative density theorem for arithmetic progressions, a certain Euclidean strengthening of Szemerédi's theorem. Here one fixes a dense set inside a unit cube and asks for a whole interval of realized scales by the progression "gaps." The analytic proof follows a paper by Durcik and the author [10], who were highly influenced by the three-term predecessor due to Cook–Magyar–Pramanik [8]. The proof combines Szemerédi's theorem [50] and its quantitative improvements [27, 30, 2, 38, 45], Varnavides-type averaging [53], Gowers norms [23, 15], and multiscale cancellation related to the multilinear Hilbert transforms [36, 37, 52, 13, 55]. However, we begin the section by first explaining why the usual Euclidean norm (i.e., the ℓ^2-norm) is no longer suitable for this problem.

3.1 Bourgain's example and a generalization

The main issue is that the Euclidean norm cannot guarantee a density theorem for arithmetic progressions, either the "all large scales" variant or the variant with "an interval of scales."

An obstruction is already visible for three-term progressions,

$$x, \quad x+y, \quad x+2y; \tag{3.1}$$

it was first observed by Bourgain [3]. Namely, the parallelogram law in $\mathbb{R}^d$ immediately gives

$$|x+2y|^2 - 2|x+y|^2 + |x|^2 = 2|y|^2.$$

An immediate consequence of this identity can be phrased as follows: if all three numbers $|x|^2$, $|x+y|^2$, $|x+2y|^2$ are close to integers, the expression $2|y|^2$ also needs to be close to an integer. More quantitatively, if there exist integers m_0, m_1, m_2 such that

$$\big||x|^2 - m_0\big| < \frac{1}{10}, \quad \big||x+y|^2 - m_1\big| < \frac{1}{10}, \quad \big||x+2y|^2 - m_2\big| < \frac{1}{10},$$

then

$$\big|2|y|^2 - (m_0 - 2m_1 + m_2)\big| < \frac{2}{5},$$

so $|y|^2$ can never be of the form $k/2 + 1/4$ for $k \in \mathbb{Z}$. Thus, if we define

$$A := \bigcup_{m=1}^{\infty} \Big\{x \in \mathbb{R}^d : m - \frac{1}{10} \leqslant |x|^2 \leqslant m + \frac{1}{10}\Big\},$$

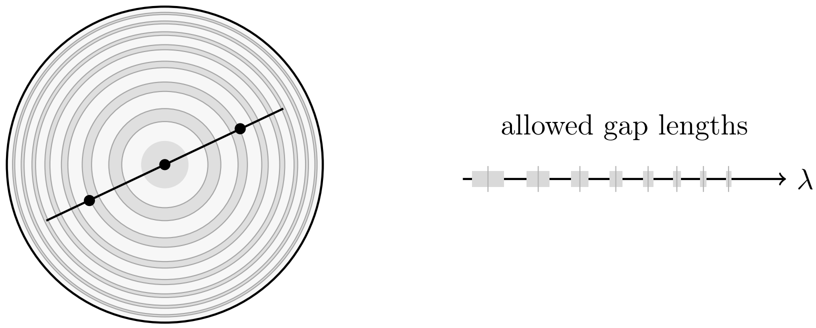

Figure 3.1: Bourgain's annular obstruction.

then its density $\mathrm{d}(A)$ is easily seen to be $1/5$, while the set

$$\big\{|y| : (\exists x, y \in \mathbb{R}^d)(x,\, x+y,\, x+2y \in A)\big\} \tag{3.2}$$

of gap lengths of three-term progressions (3.1) fully contained in A does not contain any interval of the form $[\lambda_0, \infty)$, simply because it avoids all numbers of the form $\lambda = \sqrt{(2k+1)/4}$, $k \in \mathbb{N}$. This rules out the possibility of a result analogous to Theorems 2.2, 2.4, 2.6, and 2.9. Figure 3.1 contains an illustration of the set A defined above. Geometrically it is a union of a sequence of annuli that get thinner as they move further away from the origin.

It is worth observing that a compact density result, like Theorems 2.8 and 2.10, is also not possible for three-term arithmetic progressions, again in any number of dimensions. Namely, for an arbitrary $\epsilon > 0$ define the compact set

$$A_\epsilon := \bigcup_{m=1}^{\infty} \Big\{ x \in [0,1]^d : m - \frac{1}{10} \leqslant \Big|\frac{x}{\epsilon}\Big|^2 \leqslant m + \frac{1}{10} \Big\}.$$

It is not difficult to show

$$\lim_{\epsilon \to 0^+} |A_\epsilon| = \frac{1}{5}.$$

On the other hand, just as before we see that its corresponding set (3.2) does not contain numbers $\epsilon\sqrt{(2k+1)/4}$, $k = 1, 2, 3, \ldots$. Consequently, its connected components all have length $O(\epsilon)$, so the set of progression-gaps cannot contain an interval of length that depends only on $|A_\epsilon|$.

A similar obstruction for longer progressions

$$x, \quad x+y, \quad x+2y, \ldots, \quad x+(n-1)y \tag{3.3}$$

was noted in [12]. For arithmetic progressions of length n we again do not have a density theorem that we would like, even when the progression gaps are measured in the ℓ^p-norm,

$$\|y\|_p = \|(y_1, y_2, \ldots, y_d)\|_p := \big(|y_1|^p + |y_2|^p + \cdots + |y_d|^p\big)^{1/p},$$

for $p \in \{1, 2, \ldots, n-1\}$. The obstruction is now based on the identity

$$\sum_{k=0}^{p} (-1)^{p-k} \binom{p}{k} \|x + ky\|_p^p = p! \|y\|_p^p \quad \text{for } x, y \in [0,\infty)^d,$$

and one can consider the sets

$$A := \bigcup_{m=1}^{\infty} \Big\{ x \in [0,\infty)^d : m - \frac{1}{2^{p+2}} \leqslant \|x\|_p^p \leqslant m + \frac{1}{2^{p+2}} \Big\},$$
$$A_\epsilon := \bigcup_{m=1}^{\infty} \Big\{ x \in [0,1]^d : m - \frac{1}{2^{p+2}} \leqslant \Big\|\frac{x}{\epsilon}\Big\|_p^p \leqslant m + \frac{1}{2^{p+2}} \Big\}.$$

We invite the reader to fill in the details.

3.2 The correct problem

Let $A \subseteq [0,1]^d$ be measurable. For an integer $n \geqslant 3$ define

$$\ell^p\text{-gaps}_n(A) = \{\lambda \geqslant 0 : \exists x, y \in \mathbb{R}^d,\ x, x+y, \ldots, x+(n-1)y \in A,\ \|y\|_p = \lambda\}.$$

We want to prove a compact density theorem, which will give an interval of gap lengths whose size is bounded from below using only the density of the set. The previous section explains that we must exclude the values $p = 1, 2, \ldots, n-1$. Quite surprisingly, these are the only values of $p \in [1,\infty)$ that we need to exclude.

Theorem 3.1 (Compact interval theorem for ℓ^p gaps). *Let $n \geqslant 3$,*

$$p \in [1,\infty) \setminus \{1, 2, \ldots, n-1\},$$

and let

$$D(n,p) = 2^{n+3}(n+p).$$

If $d \geqslant D(n,p)$, then every measurable set $A \subseteq [0,1]^d$ with $|A| \geqslant \delta > 0$ has ℓ^p-gaps$_n(A)$ containing an interval of length $\gtrsim_\delta 1$.

We can even be more concrete (but unlikely optimal) about this lower bound. Take $0 < \delta \leqslant 1/2$ and put $L_\delta := 1 + \log(1/\delta)$. Then the interval can have length

$$\begin{cases} \exp\big(-\exp\big(CL_\delta^6(1+\log L_\delta)\big)\big), & n = 3, \\ \exp\big(-\exp(\delta^{-C})\big), & n = 4, \\ \exp\big(-\exp\big(\exp(L_\delta^C)\big)\big), & n \geqslant 5 \end{cases}$$

for some constant $C = C(n,p,d) \in (0,\infty)$.

These strange quantities are dictated by the current state-of-the-art bounds in Szemerédi's theorem.

The "all-large-scales" variant of this result is known only for $n = 3$ and it is due to Cook, Magyar, and Pramanik [8]. The same paper [8], between the lines, also established Theorem 3.1 for $n = 3$, so an additional difficulty lies in handing longer progressions.

3.3 Szemerédi's theorem as the structured lower bound

Let $N(n,\eta)$ be the least integer such that every subset of $\{0,1,\dots,N-1\}$ with at least ηN elements contains a nontrivial n-term arithmetic progression. It is convenient to record this finitary input in the inverse form as follows. For each fixed $n \geqslant 3$ there is $C_n \in (0,\infty)$ such that

$$N(n,\eta) \leqslant \begin{cases} \exp\bigl(C_n L_\eta^6(1+\log L_\eta)\bigr), & n=3, \\ \exp(\eta^{-C_n}), & n=4, \\ \exp\bigl(\exp(L_\eta^{C_n})\bigr), & n \geqslant 5. \end{cases} \tag{3.4}$$

Indeed, if $r_n(M)$ denotes the largest cardinality of an n-term-progression-free subset of $\{1,\dots,M\}$, then Raghavan's improvement of Roth's theorem [45] (after the Kelley–Meka breakthrough [30]) gives

$$r_3(M) \leqslant M\exp\{-c(\log M)^{1/6}(\log\log M)^{-1/6}\}$$

for an absolute constant $c>0$. Taking $\log M = CL_\eta^6(1+\log L_\eta)$ with C large makes the exponential saving at least e^{-L_η}, and hence forces a nontrivial three-term progression in every subset of density at least η. The second line follows from Green and Tao's bound $r_4(M) \ll M(\log M)^{-c}$ [27]. For $n \geqslant 5$, Leng, Sah, and Sawhney prove $r_n(M) \ll_n M\exp[-(\log\log M)^{c_n}]$ [38], which gives the third line.

The continuous ingredient needed below is a Varnavides averaging version of Szemerédi's theorem. Notice that the lemma itself is qualitative in form and keeps the exact dependence on $N(n,\delta/4)$; the estimates in (3.4) are substituted only afterwards.

Lemma 3.2 (Continuous Varnavides lower bound). *For $n \geqslant 3$, $d \geqslant 1$, $0<\delta \leqslant 1/2$, $0<\lambda\leqslant 1$, and every measurable $A \subseteq [0,1]^d$ with $|A| \geqslant \delta$,*

$$\lambda^{-d}\int_{[0,\lambda]^d}\int_{[0,1]^d}\prod_{i=0}^{n-1}\mathbb{1}_A(x+iy)\,\mathrm{d}x\,\mathrm{d}y \gtrsim_d \delta^{d+1}N(n,\delta/4)^{-d-2}. \tag{3.5}$$

Proof. Put

$$N = N(n,\delta/4), \qquad \theta = \frac{\delta}{4d},$$

and consider the parameter set

$$T = \Bigl(\frac{\theta\lambda}{2N},\frac{\theta\lambda}{N}\Bigr]^d \times [0,1-\theta]^d,$$

whose elements are denoted by (t,u). Since $0<\lambda\leqslant 1$, for every $0 \leqslant k \leqslant N-1$ and $(t,u)\in T$ the point $u+kt$ lies in $[0,1]^d$; indeed $0 \leqslant kt \leqslant \theta\lambda \leqslant \theta$ coordinatewise.

Average the density of A along the N-point progression $u, u+t, \dots, u+(N-1)t$:

$$\fint_T \frac{1}{N}\sum_{k=0}^{N-1}\mathbb{1}_A(u+kt)\,\mathrm{d}t\,\mathrm{d}u.$$

For a fixed k and t, the change of variables $v=u+kt$ sends $u\in[0,1-\theta]^d$ to the box $[0,1-\theta]^d+kt$. This box contains $[\theta,1-\theta]^d$, because $0\leqslant kt\leqslant\theta$. Hence

$$\fint_T \frac{1}{N}\sum_{k=0}^{N-1}\mathbb{1}_A(u+kt)\,\mathrm{d}t\,\mathrm{d}u \geqslant |A\cap[\theta,1-\theta]^d|$$

$$
\begin{aligned}
&\geqslant |A| - \big|[0,1]^d \setminus [\theta, 1-\theta]^d\big| \\
&\geqslant \delta - 2d\theta = \delta/2.
\end{aligned}
$$

Let

$$T_{\text{large}} = \Big\{(t,u) \in T : \frac{1}{N}\sum_{k=0}^{N-1} \mathbb{1}_A(u+kt) \geqslant \delta/4\Big\}.$$

Since the averaged discrete density is bounded by 1 on T_{large} and by $\delta/4$ on its complement, the preceding inequality implies

$$\frac{\delta}{2} \leqslant \frac{|T_{\text{large}}|}{|T|} + \frac{\delta}{4},$$

and therefore

$$|T_{\text{large}}| \geqslant \frac{\delta}{4}|T| \gtrsim_d \delta^{d+1} N^{-d} \lambda^d. \tag{3.6}$$

For every $(t,u) \in T_{\text{large}}$ the index set

$$S_{t,u} = \{0 \leqslant k \leqslant N-1 : u + kt \in A\}$$

has cardinality at least $(\delta/4)N$. By the choice of N, it contains a nontrivial n-term arithmetic progression. Thus there exist integers k and $l \geqslant 1$ with $k + (n-1)l \leqslant N-1$ such that

$$\prod_{i=0}^{n-1} \mathbb{1}_A(u + (k+il)t) = 1.$$

Consequently,

$$|T_{\text{large}}| \leqslant \sum_{k=0}^{N-1} \sum_{1\leqslant l \leqslant (N-1-k)/(n-1)} \int_T \prod_{i=0}^{n-1} \mathbb{1}_A(u+(k+il)t)\,\mathrm{d}t\,\mathrm{d}u.$$

In a summand perform the change of variables

$$x = u + kt, \qquad y = lt.$$

The Jacobian is l^{-d}, and the new y lies in $(0,\theta\lambda]^d \subseteq [0,\lambda]^d$. Enlarging the x-domain to $[0,1]^d$, and using $l^{-d} \leqslant 1$, gives

$$|T_{\text{large}}| \leqslant N^2 \int_{[0,\lambda]^d} \int_{[0,1]^d} \prod_{i=0}^{n-1} \mathbb{1}_A(x+iy)\,\mathrm{d}x\,\mathrm{d}y.$$

Combining this with (3.6) yields (3.5). □

It is interesting to remark that the celebrated Szemerédi's theorem is actually equivalent to (for a fixed $n \geqslant 3$):

$$A \subseteq [0,1], \quad |A| \geqslant \delta > 0 \quad \Longrightarrow \quad \int_{[0,1]} \int_{[0,1]} \prod_{i=0}^{n-1} \mathbb{1}_A(x+iy)\,\mathrm{d}y\,\mathrm{d}x \gtrsim_\delta 1.$$

One direction follows from Lemma 3.2 with $d=1$ and $\lambda = 1$. In the other direction one applies this estimate to a set $S \subseteq \{0,1,\ldots,N-1\}$ blown up to a positive measure set

$$A := \sum_{k\in S} \Big[\frac{5k+2}{5N}, \frac{5k+3}{5N}\Big] \subseteq [0,1].$$

We will see that Lemma 3.2 gives a lower bound for the structured part that we are about to define.

3.4 Exact and smoothed counting forms

Let $S_p^{d-1} = \{y \in \mathbb{R}^d : \|y\|_p = 1\}$. Following Cook–Magyar–Pramanik and the work of Durcik and the author [8, 10], let σ denote the natural hypersurface measure on S_p^{d-1}, equivalently the vague limit of smooth densities

$$\sigma^\eta(y) = \psi_\eta(\|y\|_p^p - 1)$$

with ψ a fixed smooth approximate identity on the line. Let φ be a nonnegative even C^∞ bump on $\mathbb{R}^d$, supported in $[-3,3]^d$, positive on $[-2,2]^d$, and normalized to have integral one. For $\lambda > 0$ write σ_λ and $\varphi_{\varepsilon\lambda}$ for L^1-normalized dilates.

The exact counting form is

$$\boxed{\mathcal{N}_\lambda^0(A) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\prod_{i=0}^{n-1}\mathbb{1}_A(x+iy)\,\mathrm{d}\sigma_\lambda(y)\,\mathrm{d}x,}$$

so $\mathcal{N}_\lambda^0(A) > 0$ implies that A contains an n-term progression with ℓ^p gap λ. The smoothed form is

$$\boxed{\mathcal{N}_\lambda^\varepsilon(A) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\prod_{i=0}^{n-1}\mathbb{1}_A(x+iy)\,(\sigma_\lambda * \varphi_{\varepsilon\lambda})(y)\,\mathrm{d}y\,\mathrm{d}x.}$$

For fixed A and λ, $\mathcal{N}_\lambda^\varepsilon(A) \to \mathcal{N}_\lambda^0(A)$ as $\varepsilon \to 0^+$. Indeed, if

$$F(x,y) = \prod_{i=0}^{n-1}\mathbb{1}_A(x+iy), \qquad f(y) = \int F(x,y)\,\mathrm{d}x,$$

then f is continuous: translating y to y' changes each factor by an L^1-small translate of $\mathbb{1}_A$, and the product difference is bounded by a telescoping sum of these L^1 differences. By Fubini,

$$\mathcal{N}_\lambda^\varepsilon(A) = \int (f * \varphi_{\varepsilon\lambda})(y)\,\mathrm{d}\sigma_\lambda(y),$$

and dominated convergence with respect to the finite measure σ_λ gives the claim.

The proof again uses the decomposition (2.1).

3.5 The three estimates for arithmetic progressions

Proposition 3.3 (Structured part [10], with quantitative Szemerédi input). *There exists $E = E(n,p,d)$ such that, for every $\lambda \in (0,1]$ and every $A \subseteq [0,1]^d$ with $|A| \geqslant \delta$,*

$$\mathcal{N}_\lambda^1(A) \geqslant \Theta_{n,p,d}(\delta),$$

where

$$\Theta_{n,p,d}(\delta) = \begin{cases} \exp\bigl(-EL_\delta^6(1+\log L_\delta)\bigr), & n = 3, \\ \exp(-\delta^{-E}), & n = 4, \\ \exp\bigl(-\exp(L_\delta^E)\bigr), & n \geqslant 5. \end{cases}$$

Proof. The only point is to compare the coarse annular kernel with normalized Lebesgue measure on a cube of gap vectors. Since φ is positive on $[-2,2]^d$ and every point $z \in S_p^{d-1}$ has coordinates in $[-1,1]$, for every $u \in [0,1]^d$ one has

$$(\sigma * \varphi)(u) = \int_{S_p^{d-1}} \varphi(u-z)\,\mathrm{d}\sigma(z) \geqslant \sigma(S_p^{d-1}) \min_{w\in[-2,2]^d} \varphi(w) > 0.$$

After dilation this becomes

$$(\sigma_\lambda * \varphi_\lambda)(y) \gtrsim_{p,d} \lambda^{-d} \mathbb{1}_{[0,\lambda]^d}(y), \qquad 0 < \lambda \leqslant 1.$$

Therefore

$$\mathcal{N}_\lambda^1(A) \gtrsim_{p,d} \lambda^{-d} \int_{[0,\lambda]^d} \int_{[0,1]^d} \prod_{i=0}^{n-1} \mathbb{1}_A(x+iy)\,\mathrm{d}x\,\mathrm{d}y.$$

Lemma 3.2 bounds the right hand side below by

$$\gtrsim_{p,d} \delta^{d+1} N(n,\delta/4)^{-d-2}.$$

Substitution of (3.4), with polynomial factors in δ and constants depending on p,d absorbed by increasing E, gives precisely the three cases in the statement. $\square$

The other two estimates require harmonic analysis. Define

$$\rho(x) = d\varphi(x) + x\cdot\nabla\varphi(x).$$

Then $-t\partial_t\varphi_t = \rho_t$, so

$$\sigma_\lambda * \varphi_{a\lambda} - \sigma_\lambda * \varphi_{b\lambda} = \int_a^b \sigma_\lambda * \rho_{t\lambda} \frac{\mathrm{d}t}{t}. \tag{3.7}$$

Moreover $\int \rho = 0$, because $\rho(x) = \operatorname{div}(x\varphi(x))$.

Lemma 3.4 (The multiscale telescoping estimate). *Let $0 \leqslant f \leqslant 1$ be supported in $[0,1]^d$, and put*

$$F_{n-1}(x,y) = \prod_{i=0}^{n-1} f(x+iy).$$

For $a<b$, $\alpha,\gamma \geqslant 2^{-1/2}$, and $l \in \{1,\ldots,d\}$ define

$$\Lambda_{a,b}^{\alpha,\gamma,l} = \int_a^b \int_{(\mathbb{R}^d)^2} \left| \int_{\mathbb{R}^d} F_{n-1}(x,y) h_{s\gamma}^{(l)}(y-q)\,\mathrm{d}y \right| g_{s\alpha}(q)\,\mathrm{d}q\,\mathrm{d}x \frac{\mathrm{d}s}{s}.$$

Then

$$\Lambda_{a,b}^{\alpha,\gamma,l} \lesssim_{n,d} (\alpha\gamma)^2 \Big(\log\frac{b}{a}\Big)^{1-2^{-n+2}}. \tag{3.8}$$

Proof. The proof is the Cauchy–Schwarz "telescoping argument."

For $1 \leqslant k \leqslant n-1$ introduce the partially doubled products

$$F_k(x,y,u_{k+1},\ldots,u_{n-1}) = \prod_{i=0}^{k} \prod_{r\in\{0,1\}^{n-k-1}} f\Big(x+iy+\sum_{s=k+1}^{n-1} r_s(i+s-k)u_s\Big).$$

Thus F_{n-1} is the product in the statement. The inductive quantity at level k is obtained from the preceding display by inserting one derivative in the y variable and Gaussian weights in the already created difference variables. In the notation relevant for the induction,

$$\Lambda_{k,l,a,b}^{\alpha,\alpha_k,\dots,\alpha_{n-1}} = \int_a^b \int \left| \int F_k(x,y,u_{k+1},\dots,u_{n-1}) h_{t\alpha_k}^{(l)}(y-p_k)\,\mathrm{d}y \right| \\ \times g_{t\alpha}(p_k+\cdots+p_{n-1}) \prod_{j=k+1}^{n-1} g_{t\alpha_j}(p_j) g_{t\alpha_j}(u_j-p_j)\,\mathrm{d}\Omega \frac{\mathrm{d}t}{t},$$

where $\mathrm{d}\Omega$ denotes integration in the variables $x, p_k, \dots, p_{n-1}, u_{k+1}, \dots, u_{n-1}$. The lemma is the case $k = n-1$.

The claim is proved by induction on k. The basis $k = 1$ is the key positivity step. After changing variables so that the two factors $f(x)$ and $f(x+y)$ are separated, Cauchy–Schwarz writes the relevant expression as the geometric mean of two quantities of the form

$$\widetilde{\Theta}_{l,m,a,b} = \int_a^b \int G(x,u)\, \widetilde{K}_{l,m,t}(u)\,\mathrm{d}x\,\mathrm{d}u \frac{\mathrm{d}t}{t},$$

where $0 \leqslant G \leqslant \mathbb{1}_{[0,1]^d}$ and

$$\widetilde{K}_{l,m,t}(u) = -\int h_{t\alpha_m}^{(l)}(p_m) h_{t\alpha_m}^{(l)}(u_m-p_m) \prod_{j\neq m} g_{t\alpha_j}(p_j) g_{t\alpha_j}(u_j-p_j)\,\mathrm{d}p.$$

The convolution identities

$$g_a * g_b = g_{(a^2+b^2)^{1/2}}, \qquad h_a^{(l)} * h_b^{(l)} = \frac{ab}{a^2+b^2}\, k_{(a^2+b^2)^{1/2}}^{(l)}. \tag{3.9}$$

show that each $\widetilde{\Theta}_{l,m,a,b} \geqslant 0$: after integrating in all variables except one, it is a square. On the other hand, the heat equation gives the exact telescoping identity

$$\sum_l \sum_m \widetilde{\Theta}_{l,m,a,b} = \pi(\widetilde{\Xi}_a - \widetilde{\Xi}_b),$$

where $\widetilde{\Xi}_t$ is the same integral with every $\widetilde{K}$ replaced by the product of the corresponding Gaussians. Since $0 \leqslant \widetilde{\Xi}_t \leqslant 1$, every nonnegative summand $\widetilde{\Theta}_{l,m,a,b}$ is $O(1)$. This proves the induction basis.

Assume now that the estimate has been proved at level $k-1$ and consider level $k \geqslant 2$. Split the product defining F_k into the $i = 0$ factor and the remaining factors. Cauchy–Schwarz gives

$$\Lambda_k \leqslant S_k^{1/2} T_k^{1/2}.$$

The easy factor has no derivative and is bounded by

$$S_k \leqslant \int_a^b \frac{\mathrm{d}t}{t} = \log \frac{b}{a}.$$

For the second factor, expanding the square introduces a new difference variable u_k; after the harmless changes $x \mapsto x-y$ and $p_k \mapsto p_k + y$, the expression becomes a positive form $\Theta_{k,l,k,a,b}$ with two first derivatives joined into one second derivative. More generally, for $m = k, \dots, n-1$ one obtains nonnegative forms $\Theta_{k,l,m,a,b} \geqslant 0$. The same Gaussian heat identity as in the basis gives

$$\sum_l \Big(\Psi_{k,l,a,b} + \sum_{m=k}^{n-1} \Theta_{k,l,m,a,b} \Big) = \pi(\Xi_{k,a} - \Xi_{k,b}),$$

where $0 \leqslant \Xi_{k,t} \leqslant 1$. The term Ψ is the only one not visibly positive; it contains a second derivative in the outer Gaussian. Using (3.9), this second derivative is split into two first derivatives at scale $2^{-1/2}t$, and the induction hypothesis at level $k-1$ gives

$$|\Psi_{k,l,a,b}| \lesssim_{n,d} (\alpha\alpha_k \cdots \alpha_{n-1})^4 \Big(\log \frac{b}{a} \Big)^{1-2^{-k+2}}.$$

Since the Θ's are nonnegative and their sum is bounded by $1 + \sum_l |\Psi_{k,l,a,b}|$, each $\Theta_{k,l,m,a,b}$ satisfies the same bound. Returning to $\Lambda_k \leqslant S_k^{1/2} T_k^{1/2}$ gives

$$\Lambda_k \lesssim_{n,d} (\alpha\alpha_k \cdots \alpha_{n-1})^2 \Big(\log \frac{b}{a} \Big)^{1-2^{-k+1}}.$$

This closes the induction. Setting $k = n-1$ gives (3.8). $\square$

Proposition 3.5 (Error part [10]). *There exists $F = F(n,p,d)$ such that, whenever $\lambda_j \in (2^{-j}, 2^{-j+1}]$, $j = 1, \dots, J$,*

$$\sum_{j=1}^{J} \Big| \mathcal{N}^{\varepsilon}_{\lambda_j}(A) - \mathcal{N}^{1}_{\lambda_j}(A) \Big| \leqslant \varepsilon^{-F} J^{1-2^{-n+2}}$$

for all $0 < \varepsilon \leqslant 1/2$ and all measurable $A \subseteq [0,1]^d$.

Proof. Choose complex numbers κ_j of modulus at most 1 so that the left side is the absolute value of the signed sum. It suffices to prove

$$\Bigg| \sum_{j=1}^{J} \kappa_j (\mathcal{N}^{\varepsilon}_{\lambda_j}(A) - \mathcal{N}^{1}_{\lambda_j}(A)) \Bigg| \lesssim \varepsilon^{-F} J^{1-2^{-n+2}}. \tag{3.10}$$

By (3.7),

$$\sigma_{\lambda_j} * \varphi_{\varepsilon\lambda_j} - \sigma_{\lambda_j} * \varphi_{\lambda_j} = \int_{\varepsilon}^{1} \sigma_{\lambda_j} * \rho_{t\lambda_j} \frac{\mathrm{d}t}{t}.$$

The cancellation $\int \rho = 0$ allows us to insert a continuous Littlewood–Paley identity. Choose a Schwartz function θ satisfying

$$-\int_0^{\infty} \widehat{\theta}(u\xi)\widehat{k}(u\xi) \frac{\mathrm{d}u}{u} = 1, \qquad \xi \neq 0.$$

Applying this to the mean-zero kernel $\rho_{t\lambda_j}$ and using the triangle inequality reduces (3.10) to bounding

$$\sum_{j=1}^{J} \int_0^{\infty} \int_{\varepsilon}^{1} \Big| \int\int F_{n-1}(x,y)(\sigma_{\lambda_j} * \rho_{t\lambda_j} * \theta_{tu\lambda_j} * k_{tu\lambda_j})(y)\,\mathrm{d}y\,\mathrm{d}x \Big| \frac{\mathrm{d}t}{t}\frac{\mathrm{d}u}{u}. \tag{3.11}$$

We now split the second derivative at the scale of the jth gap. Since $\lambda_j \sim 2^{-j}$,

$$1 = \frac{1}{\log 2} \int_{2^{-j-5}t}^{2^{-j-4}t} \frac{\mathrm{d}s}{s}.$$

For such s set $r_j(s,t) = (t^2\lambda_j^2 - s^2)^{1/2}$; then $s \sim r_j(s,t) \sim t\lambda_j$. Formula (3.9) yields

$$k_{tu\lambda_j} = \frac{t^2\lambda_j^2}{s r_j(s,t)} \sum_{l=1}^{d} h^{(l)}_{r_j(s,t)u} * h^{(l)}_{su},$$

with a coefficient bounded above and below by constants. Expanding one of the two first derivatives and moving absolute values outward, (3.11) is bounded by a constant times

$$\sum_{j=1}^{J}\sum_{l=1}^{d} \int_0^\infty \int_\varepsilon^1 \int_{2^{-j-5}t}^{2^{-j-4}t} \iint \left| \int F_{n-1}(x,y) h^{(l)}_{su}(y-q)\,\mathrm{d}y \right| \left| L^{(l)}_{j,t,u,s}(q) \right| \mathrm{d}q\,\mathrm{d}x \frac{\mathrm{d}s}{s}\frac{\mathrm{d}t}{t}\frac{\mathrm{d}u}{u}, \qquad (3.12)$$

where

$$L^{(l)}_{j,t,u,s} = \sigma_{\lambda_j} * \rho_{t\lambda_j} * \theta_{tu\lambda_j} * h^{(l)}_{r_j(s,t)u}.$$

We next dominate this harmless kernel by Gaussians at the common scale su. The Fourier transform of ρ vanishes at the origin and is rapidly decreasing. Hence, for any large N,

$$\left| (\rho_{u^{-1}} * \theta * h^{(l)}_{r_j(s,t)/(t\lambda_j)})(z) \right| \lesssim_{N,d} \min\{u, u^{-1}\}(1 + \|z\|_2)^{-N}.$$

Convolving with σ and using $t \geqslant \varepsilon$ gives, after rescaling and choosing $N = d+3$,

$$\left| L^{(l)}_{j,t,u,s}(q) \right| \lesssim_{p,d} \varepsilon^{-d-3} \min\{u, u^{-1}\} \int_1^\infty g_{\beta su}(q) \frac{\mathrm{d}\beta}{\beta^4}.$$

Substitute this estimate into (3.12), interchange the integrals, and then replace s by su. The sum over j fills a single interval of s-scales, namely

$$2^{-J-5}tu \leqslant s \leqslant 2^{-5}tu.$$

Lemma 3.4 gives, for each fixed t, u, β, l,

$$\int_{2^{-J-5}tu}^{2^{-5}tu} \iint \left| \int F_{n-1}(x,y) h^{(l)}_{s}(y-q)\,\mathrm{d}y \right| g_{\beta s}(q)\,\mathrm{d}q\,\mathrm{d}x \frac{\mathrm{d}s}{s} \lesssim_{n,d} \beta^2 J^{1-2^{-n+2}}.$$

Therefore (3.11) is at most

$$C_{n,p,d}\varepsilon^{-d-3} J^{1-2^{-n+2}} \int_1^\infty \beta^2 \frac{\mathrm{d}\beta}{\beta^4} \int_0^\infty \min\{u, u^{-1}\} \frac{\mathrm{d}u}{u} \int_\varepsilon^1 \frac{\mathrm{d}t}{t}.$$

The β integral is finite. The u integral is finite after the sharper form $\min\{u, u^{-1}\}\,\mathrm{d}u/u = \min\{1, u^{-2}\}\,\mathrm{d}u$ that results from the preceding rescaling. Finally $\int_\varepsilon^1 \mathrm{d}t/t \leqslant \varepsilon^{-1}$. All fixed powers of ε^{-1} are absorbed into ε^{-F}, proving the proposition. □

Proposition 3.6 (Uniform part [8, 10]). *If $d \geqslant D(n,p)$, then*

$$\left|\mathcal{N}_\lambda^0(A) - \mathcal{N}_\lambda^\varepsilon(A)\right| \lesssim_{n,p,d} \varepsilon^{1/3}$$

for all $\lambda, \varepsilon \in (0,1]$ and all measurable $A \subseteq [0,1]^d$.

Proof. We use Euclidean Gowers norms. For a compactly supported measurable f on $\mathbb{R}^d$ put

$$\|f\|_{U^n(\mathbb{R}^d)}^{2^n} = \int_{(\mathbb{R}^d)^{n+1}} \prod_{\omega\in\{0,1\}^n} \mathcal{C}^{|\omega|} f(x+\omega\cdot h)\,\mathrm{d}x\,\mathrm{d}h,$$

where $\mathcal{C}$ denotes complex conjugation. We will use two elementary facts: the scaling law

$$\|f_\lambda\|_{U^n} = \lambda^{-d(1-(n+1)2^{-n})}\,\|f\|_{U^n}$$

and the generalized von Neumann inequality

$$\left|\int\!\!\int \prod_{i=0}^{n-1} \mathbb{1}_A(x+iy)g(y)\,\mathrm{d}y\,\mathrm{d}x\right| \lesssim_{n,d} \lambda^{d(1-(n+1)2^{-n})}\,\|g\|_{U^n(\mathbb{R}^d)} \tag{3.13}$$

whenever g is supported in a cube of sidelength $O(\lambda)$. To prove (3.13), define for $1 \leqslant k \leqslant n$ the iterated expressions

$$\Lambda_k = \int \prod_{i=0}^{k-1}\Big(\prod_{m=1}^{n-k} \Delta_{(i+m)h_m}\mathbb{1}_A(x+iy)\Big)\Big(\prod_{m=1}^{n-k}\Delta_{h_m} g(y)\Big)\,\mathrm{d}y\,\mathrm{d}x\,\mathrm{d}h_1\cdots\,\mathrm{d}h_{n-k}.$$

The case $k=n$ is the left side of (3.13). Splitting off the $i=0$ factor and applying Cauchy–Schwarz gives

$$|\Lambda_k| \leqslant L_k^{1/2} M_k^{1/2}.$$

The support of g forces all h_m to have size $O_n(\lambda)$, so $L_k \lesssim_{n,d} \lambda^{d(n-k)}$. Expanding M_k, changing variables, and relabelling the new difference variable turns it into Λ_{k-1}. Starting from $k=1$, where the second Cauchy–Schwarz factor is exactly the U^n norm of g, induction gives (3.13).

It remains to estimate the U^n norm of the difference between the sharp and smoothed ℓ^p spheres. We first record the one-dimensional oscillatory estimate

$$\left\|\mathbb{1}_{[-3,3]}(x)e^{2\pi i u|x|^p}\right\|_{U^n(\mathbb{R})} \lesssim_{n,p} (1+|u|)^{-2/D(n,p)}. \tag{3.14}$$

For $|u| \leqslant 1$ this follows from the trivial $L^{2^n/(n+1)}$ bound. For $|u|>1$ split the interval into $(-\eta,\eta)$, $[\eta,3]$, and $[-3,-\eta]$. The middle interval is representative. Expanding the U^n norm reduces the 2^nth power to integrals of

$$I_h(u) = \int_a^b e^{2\pi i u\Phi_h(x)}\,\mathrm{d}x,$$

where

$$\Phi_h(x) = \sum_{r\in\{0,1\}^{n-1}} (-1)^{|r|}\,|x+r\cdot h|^p\,.$$

If some $|h_i| \leqslant \eta$, we use $|I_h(u)| \leqslant 3$, and the set of such h has measure $O_n(\eta)$. Otherwise every $|h_i| > \eta$. Since $p \notin \{1, \ldots, n-1\}$, the repeated finite difference is not identically zero, and the fundamental theorem of calculus gives

$$\Phi_h'(x) = c_{n,p}\, h_1 \cdots h_{n-1} \int_{[0,1]^{n-1}} (x + t \cdot h)^{p-n}\, \mathrm{d}t,$$

with $c_{n,p} \neq 0$, and a similar formula for Φ_h''. On the interval $x \geqslant \eta$ this yields

$$|\Phi_h'(x)| \gtrsim_{n,p} \min\{\eta^{p-1}, \eta^{n-1}\}, \qquad |\Phi_h''(x)| \lesssim_{n,p} \max\{\eta^{p-n-1}, 1\}.$$

One integration by parts therefore gives

$$|I_h(u)| \lesssim_{n,p} \eta^{-4(n+p)+1} |u|^{-1}.$$

After squaring and integrating in h, one obtains

$$\left\| \mathbb{1}_{[\eta,3]} e^{2\pi i u |x|^p} \right\|_{U^n} \lesssim_{n,p} \left(\eta + \eta^{-8(n+p)+1} |u|^{-2} \right)^{2^{-n}}.$$

The small interval contributes $O(\eta^{(n+1)2^{-n}})$, and choosing $\eta = |u|^{-1/(4(n+p))}$ proves (3.14).

Now let $0 < \eta < t < 1$. We claim

$$\|\sigma^\eta * \rho_t\|_{U^n(\mathbb{R}^d)} \lesssim_{n,p,d} t^{1/3}. \tag{3.15}$$

Choose an intermediate thickness $t < \tau < 1$ and write

$$\sigma^\eta * \rho_t = \sigma^\tau * \rho_t + (\sigma^\eta - \sigma^\tau) * \rho_t.$$

For the first term, use $\rho = \operatorname{div}(x\varphi)$, hence $\rho_t = t \sum_m \partial_m v_t^{(m)}$ with $v^{(m)}(x) = x_m \varphi(x)$. The convolution inequality for U^n norms and the bound $\|f\|_{U^n} \leqslant \|f\|_{L^{2^n/(n+1)}}$ give

$$\|\sigma^\tau * \rho_t\|_{U^n} \lesssim t \sum_m \|\partial_m \sigma^\tau\|_{L^{2^n/(n+1)}} \lesssim_{p,d} t\tau^{-2}.$$

For the second term, Fourier inversion in the radial variable gives

$$\sigma^\eta(x) - \sigma^\tau(x) = \int_{\mathbb{R}} (\widehat{\psi}(\eta u) - \widehat{\psi}(\tau u)) e^{2\pi i u (\|x\|_p^p - 1)}\, \mathrm{d}u.$$

On the fixed support $[-3,3]^d$ the phase separates into a product of one-dimensional phases. By the tensor property of the U^n norm and (3.14),

$$\|\sigma^\eta - \sigma^\tau\|_{U^n} \lesssim_{n,p} \int_{\mathbb{R}} \left| \widehat{\psi}(\eta u) - \widehat{\psi}(\tau u) \right| (1 + |u|)^{-2d/D(n,p)}\, \mathrm{d}u.$$

Splitting the u-integral into $|u| < 1$, $1 \leqslant |u| < \tau^{-1}$, and $|u| \geqslant \tau^{-1}$, and using $d \geqslant D(n,p)$, gives

$$\|\sigma^\eta - \sigma^\tau\|_{U^n} \lesssim_{n,p,d} \tau.$$

Convolution with ρ_t costs only $\|\rho_t\|_1 = \|\rho\|_1$, so

$$\|\sigma^\eta * \rho_t\|_{U^n} \lesssim t\tau^{-2} + \tau.$$

Choosing $\tau = t^{1/3}$ proves (3.15).

We now estimate the uniform term. For $0 < \vartheta < \varepsilon$, (3.7), vague convergence of σ^η to σ, and Fubini give

$$\mathcal{N}_\lambda^\vartheta(A) - \mathcal{N}_\lambda^\varepsilon(A) = \int_\vartheta^\varepsilon \lim_{\eta\to 0^+} \int\int F(x,y)(\sigma_\lambda^\eta * \rho_{t\lambda})(y)\,\mathrm{d}y\,\mathrm{d}x \frac{\mathrm{d}t}{t}.$$

By scaling and (3.15),

$$\left\|\sigma_\lambda^\eta * \rho_{t\lambda}\right\|_{U^n} \lesssim_{n,p,d} \lambda^{-d(1-(n+1)2^{-n})} t^{1/3}.$$

The support of $\sigma_\lambda^\eta * \rho_{t\lambda}$ is contained in a cube of side $O(\lambda)$, so (3.13) cancels the power of λ and yields

$$\left|\int\int F(x,y)(\sigma_\lambda^\eta * \rho_{t\lambda})(y)\,\mathrm{d}y\,\mathrm{d}x\right| \lesssim_{n,p,d} t^{1/3},$$

uniformly in η. Letting $\eta \to 0^+$ and integrating,

$$\left|\mathcal{N}_\lambda^\vartheta(A) - \mathcal{N}_\lambda^\varepsilon(A)\right| \lesssim_{n,p,d} \int_\vartheta^\varepsilon t^{1/3} \frac{\mathrm{d}t}{t} \lesssim \varepsilon^{1/3}.$$

Finally let $\vartheta \to 0^+$, using $\mathcal{N}_\lambda^\vartheta(A) \to \mathcal{N}_\lambda^0(A)$. □

3.6 Obtaining the interval of scales

The three estimates above do not say that the error term is small at every scale. They say something weaker but sufficient: a long family of separated bad scales cannot exist. This is the compact analogue of the lacunary contradiction in Proposition 2.1.

Proof of Theorem 3.1 from the three estimates. Let $\Theta = \Theta_{n,p,d}(\delta)$ denote the structured lower bound from Proposition 3.3. Thus, one has

$$\Theta = \begin{cases} \exp\bigl(-EL_\delta^6(1+\log L_\delta)\bigr), & n = 3, \\ \exp(-\delta^{-E}), & n = 4, \\ \exp\bigl(-\exp(L_\delta^E)\bigr), & n \geqslant 5. \end{cases}$$

Let G be the implicit constant in Proposition 3.6. Choose

$$\varepsilon = \left(\frac{\Theta}{3G}\right)^3, \qquad J = \left\lfloor (3\Theta^{-1}\varepsilon^{-F})^{2^{n-2}} \right\rfloor + 1,$$

where F is the exponent in Proposition 3.5. Then

$$G\varepsilon^{1/3} \leqslant \Theta/3, \qquad \varepsilon^{-F} J^{-2^{-n+2}} \leqslant \Theta/3.$$

We claim that some dyadic interval $(2^{-j}, 2^{-j+1}]$, $1 \leqslant j \leqslant J$, has small error at every scale in it:

$$\left|\mathcal{N}_\lambda^\varepsilon(A) - \mathcal{N}_\lambda^1(A)\right| \leqslant \varepsilon^{-F} J^{-2^{-n+2}} \quad \text{for all } \lambda \in (2^{-j}, 2^{-j+1}]. \tag{3.16}$$

If no such interval existed, then for each j one could choose $\lambda_j \in (2^{-j}, 2^{-j+1}]$ violating (3.16). Summing the violations would contradict Proposition 3.5. Hence such a j exists.

For every λ in this good dyadic interval, decomposition (2.1) and the three estimates give

$$\begin{aligned}\mathcal{N}_\lambda^0(A) &\geqslant \mathcal{N}_\lambda^1(A) - \left|\mathcal{N}_\lambda^\varepsilon(A) - \mathcal{N}_\lambda^1(A)\right| - \left|\mathcal{N}_\lambda^0(A) - \mathcal{N}_\lambda^\varepsilon(A)\right| \\ &\geqslant \Theta - \Theta/3 - \Theta/3 = \Theta/3 > 0.\end{aligned}$$

Thus every λ in $(2^{-j}, 2^{-j+1}]$ is an ℓ^p gap length of an n-term progression in A.

It remains only to translate the size of J into the displayed density dependence. Since ε is a fixed power of Θ, the choice of J gives

$$J \lesssim_{n,p,d} \Theta^{-C_0}$$

for some $C_0 = C_0(n,p,d)$. Therefore the good dyadic interval has length at least

$$2^{-J} \geqslant \exp(-C\Theta^{-C}).$$

Evaluating this expression in the three regimes yields exactly the theorem. If $n = 3$, then

$$\Theta^{-C} \leqslant \exp\bigl(CL_\delta^6(1 + \log L_\delta)\bigr),$$

so the length is at least $\exp(-\exp(CL_\delta^6(1 + \log L_\delta)))$. If $n = 4$, then $\Theta^{-C} \leqslant \exp(\delta^{-C})$, giving $\exp[-\exp(\delta^{-C})]$. Finally, if $n \geqslant 5$, then

$$\Theta^{-C} \leqslant \exp\bigl(C\exp(L_\delta^E)\bigr) \leqslant \exp\bigl(\exp(L_\delta^C)\bigr)$$

after increasing C, and the interval length is at least $\exp(-\exp(\exp(L_\delta^C)))$. $\square$

3.7 Open problem

Problem 3.7 (All large ℓ^p arithmetic progressions). Let $n \geqslant 4$ and

$$p \in [1,\infty) \setminus \{1, 2, \ldots, n-1\}.$$

One expects that, in sufficiently high dimension, every set of positive upper Banach density contains n-term arithmetic progressions (3.3) with every sufficiently large ℓ^p gap $\|y\|_p$. The compact interval result proved in this chapter (Theorem 3.1) gives a density-dependent interval of gaps inside the unit cube, but the corresponding all-large-scales theorem remains open in this generality.

CHAPTER 4

Density theorems for configurations of prescribed volume

This chapter is concerned with configurations defined by a volume constraint:

- right simplices of a fixed volume,
- rectangular boxes of a fixed volume,
- box graph embeddings with a fixed product of edge-lengths,
- hyperbolic corners, and
- arbitrary unit-area triangles.

Only some of them will allow a density theorem that we would expect! Furthermore, even for those configurations, the parameter λ from our general strategy will not be the usual dilation scale, but rather a certain eccentricity coefficient.

The motivating questions go back to Graham and to Erdős, who contributed vastly to the Euclidean Ramsey theory [24, 18, 25]. The material for the first three sections is taken from [32], while the last two sections contain very recent density results in the plane due to Bulj and the author [5].

4.1 Right simplices

We begin with right-angled simplices and a compact density result.

Theorem 4.1 (Quantitative density theorem for right simplices [32]). *For every integer $m \geqslant 2$ there is $C_m \in (0, \infty)$ such that, for every $d \geqslant m+1$, if $A \subseteq [0, R]^d$ is measurable and*

$$\frac{|A|}{R^d} \geqslant \left(\frac{C_m}{\log R}\right)^{1/(9m^2)},$$

then A contains the $m+1$ vertices of a right-angled m-simplex of unit m-volume.

The simplex can be chosen with $m-1$ perpendicular edges parallel to $e_1, \dots, e_{m-1}$ and the last perpendicular edge lying in $\operatorname{span}(e_m, \dots, e_d)$.

Note that here we can be picky and require the volume of the simplex to be precisely 1. This is possible because the volume condition is "hyperbolic" as opposed to prescribing edge density (as we did earlier).

Proof. It is enough to prove the density statement in dimension $d = m+1$. Indeed, if $d > m+1$, then by Fubini at least one coordinate $(m+1)$-plane parallel to $\operatorname{span}(e_1, \dots, e_{m+1})$ has section-density at least $|A|/R^d$, and the lower-dimensional result applied to that section gives the required simplex in the original cube.

Write a point of $\mathbb{R}^{m+1}$ as (x,y) with $x \in \mathbb{R}^{m-1}$ and $y \in \mathbb{R}^2$. Let $A \subseteq [0,R]^{m+1}$ have density

$$\delta := \frac{|A|}{R^{m+1}} > 0, \qquad \theta := m^{-1} 2^{-m^2-m-1} \delta^{m+1}.$$

For $\lambda > 0$ define an exact counting form

$$\boxed{\begin{aligned}\mathcal{N}^0_\lambda(A) := \lambda^{-m+1} \int_{\mathbb{R}^{m-1}} \int_{\mathbb{R}^{m-1}} \int_{\mathbb{R}^2} \mathbb{1}_A(x,y) \prod_{k=1}^{m-1} \mathbb{1}_A(x+u_k e_k, y) \\ \times \mathbb{1}_A(x, y+v)\, \mathrm{d}\sigma_{m!|u_1\cdots u_{m-1}|^{-1}}(v) \prod_{k=1}^{m-1} \mathbb{1}_{\theta\lambda \leqslant |u_k| \leqslant \lambda}\, \mathrm{d}y\, \mathrm{d}x\, \mathrm{d}u.\end{aligned}}$$

If $\mathcal{N}^0_\lambda(A) > 0$, then for some x, y, u, v all points

$$(x,y), \quad (x+u_1e_1, y), \ldots, (x+u_{m-1}e_{m-1}, y), \quad (x, y+v)$$

belong to A. They form a right simplex, because the displayed edge vectors are mutually orthogonal, and its m-volume is

$$\frac{|u_1 \cdots u_{m-1}|\,|v|}{m!} = \frac{|u_1 \cdots u_{m-1}|}{m!}\, m!|u_1 \cdots u_{m-1}|^{-1} = 1.$$

Thus the task is to make one exact count positive. Define $\mathcal{N}^\varepsilon_\lambda(A)$ simply by replacing σ_r with $\sigma_r * g_{\varepsilon r}$.

The three terms from (2.1) are controlled as follows.

Structured part. First, if $R^{-1/(m-1)} \leqslant \lambda \leqslant R$, then

$$\mathcal{N}^1_\lambda(A) \geqslant c_m \delta^{(m+1)(2m-1)} R^{m+1}. \tag{4.1}$$

Indeed, when $\theta\lambda \leqslant |u_k| \leqslant \lambda$, the smoothed circle kernel at radius $m!|u_1 \cdots u_{m-1}|^{-1}$ is bounded from below on a square of side comparable to λ^{-m+1} by a multiple of $(\theta\lambda)^{2(m-1)}$. Hence $\mathcal{N}^1_\lambda$ dominates, up to constants depending on m, the average number of parallelepiped-shaped boxes

$$I_1 \times \cdots \times I_{m-1} \times Q, \qquad |I_k| = \lambda, \qquad \ell(Q) = \lambda^{-m+1},$$

inside $[0,R]^{m+1}$ for which one point is chosen from A, each coordinate line through that point has another point in A, and the vertical two-dimensional fiber has another point in A. On each such box Hölder's inequality gives

$$\begin{aligned}\int \mathbb{1}_A(x,y) \prod_{k=1}^{m-1} \Big(\int_{I_k} \mathbb{1}_A(x_1, \ldots, x_{k-1}, x'_k, x_{k+1}, \ldots, y)\, \mathrm{d}x'_k \Big) \\ \times \Big(\int_Q \mathbb{1}_A(x, y')\, \mathrm{d}y' \Big)\, \mathrm{d}x\, \mathrm{d}y \gtrsim_m |A \cap (I_1 \times \cdots \times I_{m-1} \times Q)|^{2m-1}.\end{aligned}$$

Summing over the partition and using Jensen's inequality yields the lower bound in (4.1); the small exceptional region where some $|u_k| < \theta\lambda$ is swallowed by the choice of θ.

Uniform part. Second, for every $0 < \varepsilon \leqslant 1$,

$$|\mathcal{N}_\lambda^0(A) - \mathcal{N}_\lambda^\varepsilon(A)| \leqslant C_m \varepsilon^{1/2} R^{m+1}. \tag{4.2}$$

To see this, write $f(x, y; u) = \mathbb{1}_A(x, y) \prod_{k<m} \mathbb{1}_A(x + u_k e_k, y)$ and use the heat equation identity

$$\sigma_r - \sigma_r * g_{\varepsilon r} = \int_0^\varepsilon \sigma_r * k_{tr} \frac{\mathrm{d}t}{t},$$

where k_t is a cancellative Gaussian derivative. Taking the Fourier transform in the y variable and using the standard decay estimate for circular measure, one obtains

$$\left| \int f(x, y; u) \mathbb{1}_A(x, y + v)(\sigma_r * k_{tr})(v) \,\mathrm{d}v \,\mathrm{d}y \right| \lesssim t^{1/2} R^2.$$

After integrating in x, u and using the normalizing factor $R^{-m-1}\lambda^{-m+1}$, the integral in t is $\int_0^\varepsilon t^{-1/2}\,\mathrm{d}t$, which proves (4.2). This is the usual Fourier-decay estimate for the uniform part in Bourgain's method [3, 8].

Error part. Third, the error term satisfies the logarithmic square-function bound

$$\int_{\mathbb{R}} |\mathcal{N}_{e^\alpha}^\varepsilon(A) - \mathcal{N}_{e^\alpha}^1(A)|^2 \,\mathrm{d}\alpha \leqslant C_m \theta^{-4(m-1)} \left(\log \frac{1}{\varepsilon}\right)^2 R^{m+1}. \tag{4.3}$$

Here one writes $\mathcal{N}^\varepsilon - \mathcal{N}^1$ as $\int_\varepsilon^1$ of the same heat-flow derivative. The cancellative kernel k is split as a sum of products of two first derivatives of Gaussians. Cauchy–Schwarz places one derivative on the product f and the other on the final indicator $\mathbb{1}_A(x, \cdot)$. The first factor is estimated crudely by support size; the second factor is evaluated by Plancherel. After the change of variables $\lambda = e^\alpha$, the interval of admissible Gaussian scales has bounded overlap in the (α, s) plane, and the remaining integral is bounded by $|A| \leqslant R^{m+1}$. This proves (4.3); it is the same square-function mechanism as in the Euclidean Szemerédi proofs of Bourgain and Cook–Magyar–Pramanik [3, 8].

We now finish the proof. Choose

$$\varepsilon = c_m \delta^{2(m+1)(2m-1)}$$

with $c_m > 0$ small enough that (4.2) is at most one third of (4.1). By (4.3), if

$$J \geqslant C_m \theta^{-4(m-1)} \delta^{-2(m+1)(2m-1)} \left(\log \frac{1}{\varepsilon}\right)^2,$$

then for some $\beta \in [0, J]$ the error term at $\lambda = e^\beta$ is also at most one third of the structured term. Provided $R \geqslant e^J$, this scale lies in the range where the structured estimate applies. Therefore

$$\mathcal{N}_{e^\beta}^0(A) \geqslant \frac{1}{3} \mathcal{N}_{e^\beta}^1(A) > 0,$$

and hence A contains the desired unit-volume right simplex. Since the displayed value of J is bounded by $C_m \delta^{-9m^2}$ after increasing C_m, the condition $R \geqslant e^J$ follows from

$$\delta \geqslant \left(\frac{C_m}{\log R}\right)^{1/(9m^2)}$$

and we are done. $\square$

As an immediate consequence, we obtain a non-compact density result, in which a large set attains all volumes (with absolutely no exceptions).

Corollary 4.2 (Positive upper Banach density version [32]). *Let $m \geqslant 2$ and $d \geqslant m+1$. If $A \subseteq \mathbb{R}^d$ is measurable with $\overline{\delta}(A) > 0$, then for every $V > 0$ the set A contains a right-angled m-simplex of m-volume V. Moreover the ratio of any two perpendicular edge lengths may be bounded by*

$$\exp\left(C'_m \overline{\delta}(A)^{-9m^2}\right).$$

Proof of the corollary. It is enough to handle $V = 1$, since the dilation $V^{-1/m}A$ has the same upper Banach density as A, and scaling back by $V^{1/m}$ multiplies the simplex volume by V without changing ratios of perpendicular edge lengths. Put $\delta = \overline{\delta}(A)$ and choose

$$R_0 = \exp(2C_m \delta^{-9m^2}).$$

If some cube Q of side R_0 satisfies

$$\frac{|A \cap Q|}{|Q|} > \left(\frac{C_m}{\log R_0}\right)^{1/(9m^2)},$$

then the theorem applied to $A \cap Q$ gives a unit-volume right simplex in A. Its smallest perpendicular edge length is at least $m! R_0^{-m+1}$ when the product of the m edge lengths is $m!$, and every edge is at most R_0; hence the ratio of any two perpendicular edge lengths is at most $R_0^m/m! \leqslant \exp(C'_m \delta^{-9m^2})$.

If no such cube existed, then every large cube could be tiled by side-R_0 cubes plus a boundary remainder of relative measure $O(R_0/R)$. Taking suprema over translations and then $R \to \infty$ would give

$$\overline{\delta}(A) \leqslant \left(\frac{C_m}{\log R_0}\right)^{1/(9m^2)} < \delta,$$

contradicting the definition of δ. Thus the required cube exists, and the dilation argument gives the statement for every prescribed volume V. □

4.2 Rectangles and rectangular boxes

Surprisingly, rectangles of area 1 fail every desired density theorem.

Theorem 4.3 (No monochromatic unit-area rectangle [32]). *There exists a Jordan-measurable coloring of $\mathbb{R}^2$ in 25 colors such that no color class contains the vertices of a rectangle of area 1.*

In particular, there exists a set $A \subseteq \mathbb{R}^2$ of upper density $\overline{\mathrm{d}}(A) \geqslant 1/25$ that avoids the vertices of every unit-area rectangle.

Proof. Identify $\mathbb{R}^2$ with $\mathbb{C}$. For a parallelogram with vertices

$$z, \quad z+u, \quad z+u+v, \quad z+v,$$

consider the alternating complex invariant

$$I = z^2 - (z+u)^2 + (z+u+v)^2 - (z+v)^2 = 2uv.$$

If the parallelogram is a rectangle of area 1, then $u \perp v$ and $|u||v| = 1$, so $|I| = 2$. It is therefore enough to color the plane in such a way that a monochromatic parallelogram never has $|I| = 2$.

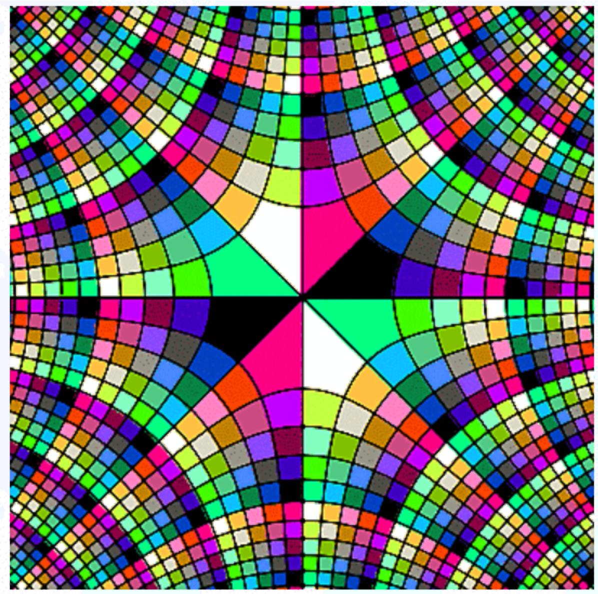

Figure 4.1: A coloring of the plane into 25 colors such that no rectangle of area 1 has all vertices of the same color.

For $0 \leqslant j, k \leqslant 4$, define

$$\mathscr{C}_{j,k} = \left\{ z \in \mathbb{C} : z^2 \in \frac{10}{3} \left(\mathbb{Z} + i\mathbb{Z} + \frac{j+ik}{5} + \left[0, \frac{1}{5}\right) + i\left[0, \frac{1}{5}\right) \right) \right\}.$$

These 25 sets partition $\mathbb{C}$ up to boundaries, and the boundaries are preimages of grid lines under the polynomial map $z \mapsto z^2$, hence have planar measure zero. Thus the coloring is Jordan-measurable in the usual sense. It is illustrated in Figure 4.1.

If all four vertices of a parallelogram are in the same class $\mathscr{C}_{j,k}$, then the four corresponding values of z^2 lie in translates of the same half-open square. Their alternating sum lies in

$$\frac{10}{3} \left(\mathbb{Z} + i\mathbb{Z} + \left(-\frac{2}{5}, \frac{2}{5} \right) + i\left(-\frac{2}{5}, \frac{2}{5} \right) \right).$$

The square around the origin has radius at most $(10/3)(2/5)\sqrt{2} = 4\sqrt{2}/3 < 2$, while every other such square is separated from the origin by distance greater than 2. Hence this set is disjoint from the circle $\{|w| = 2\}$. Consequently no monochromatic parallelogram has $|uv| = 1$, and in particular no monochromatic rectangle has area 1. $\square$

A proof of the same negative result in higher dimensions cannot use complex numbers, so we resort to the coloring that prevents only slightly tilted boxed and to the compactness of the group of rotations.

Theorem 4.4 (Higher-dimensional box avoidance [32]). *For every positive integer d there exists a finite Jordan-measurable coloring of $\mathbb{R}^d$ such that, for every $m \leqslant d$, no color class contains the 2^m vertices of an m-dimensional rectangular box of m-volume 1.*

Proof. Fix $m \leqslant d$ first; at the end we refine the finitely many colorings obtained for $m = 1, \ldots, d$. If an m-box is based at $p \in \mathbb{R}^d$ and spanned by mutually orthogonal vectors $v_1, \ldots, v_m$, write its vertices as $p + \sum_{j \in T} v_j$, $T \subseteq \{1, \ldots, m\}$. Define

$$J(\mathcal{R}) = \sum_{T \subseteq \{1,\ldots,m\}} (-1)^{m-|T|} \prod_{k=1}^{m} \left(p_k + \sum_{j \in T} v_{j,k} \right).$$

The elementary identity

$$\sum_{T \subseteq \{1,\ldots,m\}} (-1)^{m-|T|} \prod_{k=1}^{m} \left(p_k + \sum_{j \in T} v_{j,k} \right) = \sum_{\pi \in S_m} \prod_{j=1}^{m} v_{j,\pi(j)} \tag{4.4}$$

is proved by expanding the product and observing that every monomial missing some index j cancels after summing over T, while the surviving monomials choose exactly one entry from each row v_j and exactly one coordinate k. In particular, for an axis-parallel box with edge lengths $a_1, \ldots, a_m$, one has $J = a_1 \cdots a_m$.

The same remains approximately true for boxes whose orientation is close to the standard one. Let $\varepsilon_0 = (2^{m+2} m!)^{-1}$. Suppose $v_j = a_j U e_j$ with $U \in SO(d)$ and $\|U - I\|_{\mathrm{op}} < \varepsilon_0$. Then $|v_{j,j} - a_j| < \varepsilon_0 a_j$ and $|v_{j,k}| < \varepsilon_0 a_j$ for $k \neq j$. In (4.4), the identity permutation contributes within $m\varepsilon_0 (1+\varepsilon_0)^{m-1} a_1 \cdots a_m$ of $a_1 \cdots a_m$, while every non-identity permutation contains at least one off-diagonal entry and contributes at most $\varepsilon_0 (1+\varepsilon_0)^{m-1} a_1 \cdots a_m$. Hence

$$|J(\mathcal{R}) - a_1 \cdots a_m| \leqslant \frac{1}{4} a_1 \cdots a_m.$$

For a unit-volume box this implies

$$J(\mathcal{R}) \in (-5/4, -3/4) \cup (3/4, 5/4),$$

up to the harmless sign coming from the choice of base vertex and parity convention.

Now partition $\mathbb{R}^d$ into

$$\mathscr{S}_l = \left\{ x \in \mathbb{R}^d : x_1 \cdots x_m \in \frac{3}{2} \left(\mathbb{Z} + \left[\tfrac{l}{3 \cdot 2^m}, \tfrac{l+1}{3 \cdot 2^m} \right) \right) \right\}, \qquad 0 \leqslant l < 3 \cdot 2^m.$$

If all vertices of a near-standard unit-volume m-box belonged to the same $\mathscr{S}_l$, then the alternating sum defining J would lie in

$$\frac{3}{2}\mathbb{Z} + (-1/4, 1/4),$$

which is disjoint from $(-5/4, -3/4) \cup (3/4, 5/4)$. Therefore this coloring forbids all unit-volume m-boxes whose orientation lies in a fixed neighborhood of the standard orientation.

For arbitrary orientations, compactness of $SO(d)$ supplies finitely many rotations $U_1, \ldots, U_L$ such that every $U \in SO(d)$ has $U_i^{-1} U$ within operator distance ε_0 of I for some i. Refine the rotated colorings $U_i \mathscr{S}_l$ over all i. If a unit-volume m-box were monochromatic for the refinement, rotating it back by the appropriate U_i^{-1} would produce a forbidden near-standard monochromatic box for the basic coloring. Finally, refine the finitely many colorings corresponding to $m = 1, \ldots, d$. This gives one finite Jordan-measurable coloring excluding all unit-volume rectangular boxes of all dimensions $m \leqslant d$. $\square$

The negative result for fixed volume does not rule out sufficiently large volumes. In an increased ambient dimension one still has a valid density theorem.

Theorem 4.5 (Large rectangular boxes [32]). *Let m, d be positive integers with $d \geqslant m+1$. If $A \subseteq \mathbb{R}^d$ is measurable with $\overline{\delta}(A) > 0$, then there exists $V_0(A) > 0$ such that for every $V \geqslant V_0(A)$, A contains the vertices of an m-dimensional rectangular box of volume V. The boxes can be chosen with $m-1$ edges in the first coordinate directions and the remaining edge in the span of the remaining coordinates.*

The moral of formulating this theorem is that we can prove it already in $d \geqslant m+1$ dimensions, while Theorem 2.6, which realized a stronger type of embedding, used $d \geqslant 2m$.

Proof. Suppose, toward a contradiction, that A has positive upper Banach density and omits boxes of volumes λ_j^m for a sequence $\lambda_j \to \infty$. Passing to a subsequence, assume $\lambda_{j+1} \geqslant 2\lambda_j$. Choose a large cube $x_0 + [0,R]^d$ on which A has density at least $\delta > 0$; after translating, set $A_R = (A - x_0) \cap [0,R]^d$ and assume $R \geqslant \lambda_J$ for a large integer J to be chosen.

Write points as $(x, y) \in \mathbb{R}^{m-1} \times \mathbb{R}^{d-m+1}$. For $\lambda > 0$ define

$$\boxed{\begin{aligned}\mathcal{N}_\lambda^0(f) := \lambda^{-m+1} \int_{\mathbb{R}^{m-1}} \int_{\mathbb{R}^{d-m+1}} \int_{\mathbb{R}^{m-1}} &\prod_{r \in \{0,1\}^m} f(x_1 + r_1 u_1, \ldots, x_{m-1} + r_{m-1} u_{m-1}, y + r_m v) \\ &\mathrm{d}\sigma_{\lambda^m |u_1 \cdots u_{m-1}|^{-1}}(v) \prod_{k=1}^{m-1} \mathbb{1}_{\theta\lambda \leqslant |u_k| \leqslant \lambda} \,\mathrm{d}u \,\mathrm{d}y \,\mathrm{d}x,\end{aligned}}$$

where $\theta = m^{-1} 2^{-2^m d} \delta^{2^m}$. A positive value of $\mathcal{N}_\lambda^0$ gives a rectangular box whose first $m-1$ edge lengths are $|u_k|$ and whose last edge length is $\lambda^m |u_1 \cdots u_{m-1}|^{-1}$, hence whose m-volume is exactly λ^m.

Smoothen the spherical measure as before and decompose as in (2.1).

Structured part. First, for $0 < \lambda \leqslant R$,

$$\mathcal{N}_\lambda^1(\mathbb{1}_{A_R}) \gtrsim_{m,d,\delta} R^d. \tag{4.5}$$

Indeed, for $\theta\lambda \leqslant |u_k| \leqslant \lambda$ the smoothed spherical kernel dominates a constant multiple of $\lambda^{-d+m-1} \mathbb{1}_{[-\lambda,\lambda]^{d-m+1}}$. Partition $[0,R]$ in the first $m-1$ coordinates into intervals of length λ and the remaining coordinate block into cubes of side λ. On each product cell Q the Gowers-box Cauchy–Schwarz inequality, the product-form analogue of Proposition A.5, gives

$$\int_{Q^{\{0,1\}^m}} \prod_{r \in \{0,1\}^m} \mathbb{1}_{\mathbb{1}_{A_R}}(z_r) \,\mathrm{d}z \geqslant |Q|^{2-2^m} |\mathbb{1}_{A_R} \cap Q|^{2^m}.$$

Summing in Q and applying Jensen's inequality yields a lower bound of size $(R\lambda)^d \delta^{2^m}$. The exceptional region where one $|u_k| < \theta\lambda$ contributes at most $m\theta(R\lambda)^d$ and is smaller by the definition of θ, giving (4.5).

Uniform part. Second,

$$|\mathcal{N}_\lambda^0(\mathbb{1}_{A_R}) - \mathcal{N}_\lambda^\varepsilon(\mathbb{1}_{A_R})| \lesssim_{m,d} \varepsilon^{1/2} R^d.$$

This is the same Fourier-decay estimate as for right simplices: after fixing x, u, Fourier transform in the y variable and use the spherical estimate for $\sigma * k_t$ in dimension $d-m+1 \geqslant 2$; the normalizations cancel the sizes of the x and u regions, leaving $\int_0^\varepsilon t^{-1/2}\,\mathrm{d}t$.

Error part. Third, for separated scales,

$$\sum_{j=1}^{J} |\mathcal{N}^{\varepsilon}_{\lambda_j}(\mathbb{1}_{A_R}) - \mathcal{N}^{1}_{\lambda_j}(\mathbb{1}_{A_R})| \lesssim_{m,d,\delta,\varepsilon} R^d. \tag{4.6}$$

The proof is a soft singular Brascamp–Lieb argument. The heat-flow representation of $\mathcal{N}^{\varepsilon} - \mathcal{N}^{1}$ replaces one Gaussian by a cancellative derivative. Splitting the derivative into two first derivatives and applying Cauchy–Schwarz reduces the scale sum to a positive cubical form

$$\Theta^{(k)}_{\gamma}(F) = -\int_0^\infty \int \prod_{r\in\{0,1\}^m} F(z_r) k_{s\gamma_k}(z_k^0 - z_k^1) \prod_{i\neq k} g_{s\gamma_i}(z_i^0 - z_i^1)\,\mathrm{d}z\,\frac{\mathrm{d}s}{s},$$

with $F = \mathbb{1}_{A_R}$. Positivity follows from $-k_s = 2\sum_l h^{(l)}_{s/\sqrt{2}} * h^{(l)}_{s/\sqrt{2}}$. Also, differentiating the fully Gaussian cubical average in s gives

$$\sum_{k=1}^{m} \Theta^{(k)}_{\gamma}(F) = 2\pi\, \|F\|^{2^m}_{L^{2^m}}.$$

Thus every $\Theta^{(k)}_{\gamma}(F)$ is at most $2\pi|A_R| \leqslant 2\pi R^d$. The intervals of derivative scales associated with the separated λ_j have bounded overlap depending on ε; after the prefactor R^{-d} this proves (4.6). This is the cubical singular Brascamp–Lieb estimate in its heat-flow form, related to [9, 14].

From the general principle in Proposition 2.1 we conclude that a positive-density set contains vertices of boxes of every sufficiently large volume. □

4.3 Hypercube graph obstruction

Here we are studying a distance graph which is isomorphic to the 1-skeleton of the n-dimensional rectangular box. Just as in Section 2.5, its edge lengths will be denoted $b_1, b_2, \dots, b_n$.

Theorem 4.6 (Avoiding hypercube graph embeddings with product-one edge lengths [32]). *For every positive integer n there exists a Jordan-measurable finite coloring of $\mathbb{R}^2$ such that no color class contains an embedding*

$$z + r_1u_1 + \cdots + r_nu_n, \qquad (r_1,\dots,r_n) \in \{0,1\}^n,$$

of the 1-skeleton of an n-box with edge lengths $b_i = |u_i|$ satisfying

$$b_1b_2\cdots b_n = 1.$$

In particular, there exists a set $A \subseteq \mathbb{R}^2$ of positive upper density that avoids this point configuration.

Proof. The one-variable specialization of (4.4) is

$$\sum_{T\subseteq\{1,\dots,n\}} (-1)^{n-|T|} \left(z + \sum_{j\in T} u_j\right)^n = n!\,u_1u_2\cdots u_n.$$

Color $\mathbb{C}$ by

$$\mathscr{C}_{j,k} = \left\{ z \in \mathbb{C} : z^n \in 2n! \left(\mathbb{Z} + i\mathbb{Z} + \frac{j+ik}{2^{n+1}} + [0, 2^{-n-1}) + i[0, 2^{-n-1}) \right) \right\},$$

where $0 \leqslant j, k < 2^{n+1}$. If all 2^n vertices $z + \sum_{j\in T} u_j$ have the same color, then the alternating sum of their nth powers belongs to

$$2n! \left(\mathbb{Z} + i\mathbb{Z} + (-1/2, 1/2) + i(-1/2, 1/2) \right).$$

After division by $n!$, this says

$$u_1 \cdots u_n \in 2(\mathbb{Z} + i\mathbb{Z}) + (-1/2, 1/2) + i(-1/2, 1/2).$$

The latter set is disjoint from the unit circle: its central square lies strictly inside the unit disk, while every other square has distance greater than 1 from the origin. Hence $|u_1 \cdots u_n| \neq 1$, so the product of edge lengths cannot equal 1. $\square$

4.4 Hyperbolic corners

The following sections are based on the joint work by Bulj and the author [5]. The proof keeps Bourgain's scale-pigeonholing philosophy [3, 4], but the uniform estimate is supplied by a hyperbolic variant of the trilinear smoothing theory of Christ–Durcik–Roos [6].

A *hyperbolic corner* is a triple

$$(x, y), \quad (x + t, y), \quad (x, y + t^{-1}), \qquad t > 0. \tag{4.7}$$

It is an upward axes-aligned right triangle of area 1/2. A coloring theorem for this configuration has been shown by Graham [24]: every coloring of the Euclidean plane into finitely many colors necessarily contains a monochromatic triple of points of the form (4.7). However, no density theorem of this form was previously known.

Moreover, studying configuration (4.7) continues a line of research interested in *linear corners*,

$$(x, y), \quad (x + t, y), \quad (x, y + t),$$

see [12], and certain non-linear corners, such as

$$(x, y), \quad (x + t, y), \quad (x, y + t^2),$$

see [6]. However, the existing literature was mostly concerned with polynomial nonlinearities only.

Theorem 4.7 (Sets without hyperbolic corners [5]). *Let $M(R)$ be the supremum of $|A|$ over measurable $A \subseteq [0, R]^2$ containing no hyperbolic corner. For $R \geqslant 10$,*

$$R \log R \lesssim M(R) \lesssim R^2 \left(\frac{\log \log R}{\log R} \right)^{1/4}.$$

Proof. We first prove the lower bound. Let $m = \lfloor R/4 \rfloor$ and let

$$S_j = \left\{ (x, y) \in [0, R]^2 : R - 4j \leqslant x + y \leqslant R - 4j + \frac{1}{8j} \right\}, \qquad 1 \leqslant j \leqslant m.$$

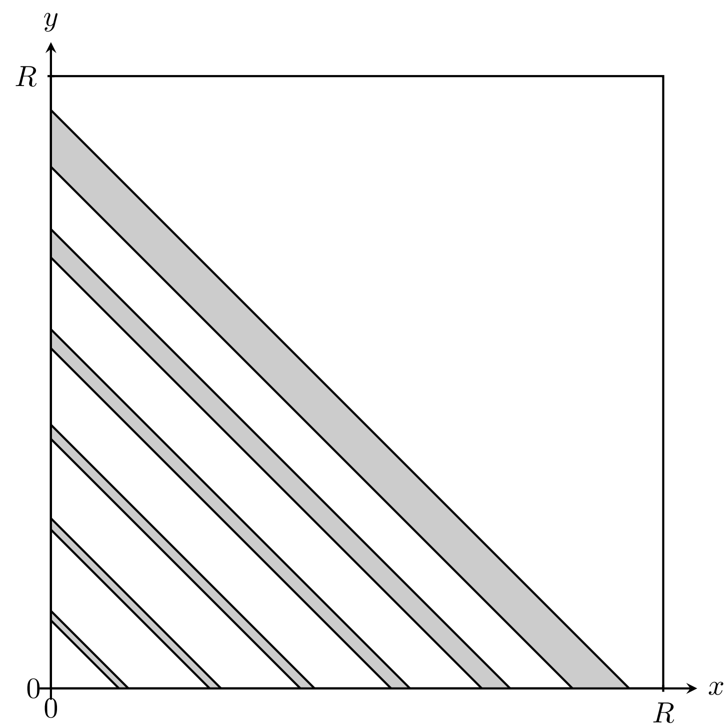

Figure 4.2: A set realizing $\Omega(R \log R)$.

Set $A_R = \bigcup_{j=1}^{m} S_j$; see Figure 4.2. The bands are disjoint, and the area of a strip $a \leqslant x + y \leqslant b$ in the relevant triangle is $(b^2 - a^2)/2$. Hence

$$|A_R| = \frac{1}{2} \sum_{j=1}^{m} \left(\left(R - 4j + \frac{1}{8j} \right)^2 - (R - 4j)^2 \right) = \frac{1}{8} R \log R + O(R).$$

If $(x, y) \in S_j$ and both $(x + t, y)$ and $(x, y + t^{-1})$ were in the same band S_j, then $t \leqslant 1/(8j)$ and $t^{-1} \leqslant 1/(8j)$, impossible. If both moved to earlier bands S_k, $k < j$, the gaps of size almost 4 between bands would force $t > 3$ and $t^{-1} > 3$, also impossible. If one point stays in S_j and the other moves to an earlier band, then, for instance, $t \leqslant 1/(8j)$ while $t^{-1} < 4j$, again impossible. The remaining mixed case is symmetric. Thus A_R contains no hyperbolic corner and has measure $\gtrsim R \log R$.

For the upper bound, let $A \subseteq [0, R]^2$ be corner-free and put $\delta = |A|/R^2$. Fix smooth nonnegative functions ζ and ϕ with ζ supported in $[1/2, 2]$, $\int \phi = 1$, and ϕ bounded below on a fixed neighborhood of the origin. For $\lambda > 0$ define

$$\boxed{\mathcal{N}_\lambda^0(A) = \int_0^\infty \int_{\mathbb{R}^2} \mathbb{1}_A(x, y) \mathbb{1}_A(x + \lambda u, y) \mathbb{1}_A \left(x, y + \frac{1}{\lambda u} \right) \zeta(u) \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}u.}$$

If A is corner-free, then $\mathcal{N}_\lambda^0(A) = 0$ for every $\lambda > 0$. Define the smoothed form

$$\boxed{\mathcal{N}_\lambda^\varepsilon(A) = \mathcal{N}_\lambda^0\big(\mathbb{1}_A, \mathbb{1}_A *_1 \phi_{\lambda\varepsilon}, \mathbb{1}_A *_2 \phi_{\lambda^{-1}\varepsilon}\big),}$$

where $*_1$ and $*_2$ denote convolution in the first and second coordinate. The same splitting (2.1) is useful once again.

Structured part. The structured term obeys, for $\lambda \in [1/R, R]$,

$$\mathcal{N}^1_\lambda(A) \geqslant c\frac{|A|^3}{R^4} = cR^2\delta^3. \tag{4.8}$$

Indeed, $\mathcal{N}^1_\lambda$ has a positive kernel $K^+_\lambda(x-x', y-y')$ satisfying

$$K^+_\lambda \gtrsim \mathbb{1}_{[-\lambda,\lambda]\times[-\lambda^{-1},\lambda^{-1}]}.$$

Partition $[0,R]^2$ into $N = O(R^2)$ rectangles $Q = I \times J$ with $|I| \leqslant \lambda$ and $|J| \leqslant \lambda^{-1}$. On each Q, Hölder gives

$$\int_{I\times I\times J\times J} \mathbb{1}_A(x,y)\mathbb{1}_A(x',y)\mathbb{1}_A(x,y')\,\mathrm{d}x\,\mathrm{d}x'\,\mathrm{d}y\,\mathrm{d}y' \geqslant |A\cap Q|^3.$$

Since $N \lesssim R^2$ and $\sum_Q |A \cap Q| = |A|$, Jensen's inequality gives (4.8).

Error part. The error term satisfies

$$\left(\int_0^\infty |\mathcal{N}^\varepsilon_\lambda(A) - \mathcal{N}^1_\lambda(A)|^2 \frac{\mathrm{d}\lambda}{\lambda}\right)^{1/2} \lesssim |A| \left(\log\frac{1}{\varepsilon}\right)^{1/2}. \tag{4.9}$$

To prove it, expand the difference between the two smoothing scales as a sum of a "horizontal" and a "vertical" term:

$$(\phi_{\lambda\varepsilon} - \phi_\lambda)(x)\phi_{\lambda^{-1}\varepsilon}(y) + \phi_\lambda(x)(\phi_{\lambda^{-1}\varepsilon} - \phi_{\lambda^{-1}})(y).$$

For the horizontal part, Cauchy–Schwarz in (x,y) gives

$$|E_1(\lambda)| \lesssim |A|^{1/2}\,\|\mathbb{1}_A *_1 (\phi_{\lambda\varepsilon} - \phi_\lambda)\|_2\,.$$

Integrating $|E_1(\lambda)|^2$ in $\mathrm{d}\lambda/\lambda$ and using Plancherel reduces the estimate to

$$\sup_{\xi\neq 0}\int_0^\infty |\widehat{\phi}(\varepsilon\lambda\xi) - \widehat{\phi}(\lambda\xi)|^2\frac{\mathrm{d}\lambda}{\lambda} \lesssim \log\frac{1}{\varepsilon}.$$

This follows by the change of variables $s = \lambda|\xi|$, smoothness at the origin, and rapid decay at infinity. The vertical part is identical after replacing λ by λ^{-1}. This proves (4.9).

Uniform part. For the uniform term, a hyperbolic trilinear smoothing estimate gives some $\sigma > 0$ such that

$$|\mathcal{N}^0_\lambda(A) - \mathcal{N}^\varepsilon_\lambda(A)| \lesssim |A|\varepsilon^\sigma = R^2\delta\varepsilon^\sigma. \tag{4.10}$$

Namely, the hyperbolic analogue of Christ–Durcik–Roos [6] reads

$$|\mathcal{N}^0_1(f_0, f_1, f_2)| \lesssim \|f_0\|_\infty\,\|f_1\|_{H^{-\sigma,0}}\,\|f_2\|_{H^{0,-\sigma}}\,,$$

for functions supported on a fixed square. To derive (4.10), anisotropically rescale by $(x,y) = (\lambda x', \lambda^{-1}y')$, so all λ become 1. Decompose $f - f *_1 \phi_\varepsilon$ and $f - f *_2 \phi_\varepsilon$, localize by a bounded-overlap partition of unity, and apply the smoothing estimate on each bounded window. Since

$$\|f - f *_1 \phi_\varepsilon\|_{H^{-\sigma,0}} \lesssim \varepsilon^\sigma\,\|f\|_2\,, \qquad \|f - f *_2 \phi_\varepsilon\|_{H^{0,-\sigma}} \lesssim \varepsilon^\sigma\,\|f\|_2\,,$$

the localized square sum is bounded by $\varepsilon^\sigma\,\|f\|_2^2 = \varepsilon^\sigma|A|$. This proves (4.10).

Combining bounds. Now choose $\varepsilon = c_0\delta^{2/\sigma}$ with c_0 small. Since $\mathcal{N}^0_\lambda(A) = 0$, Estimate (4.10) implies

$$\mathcal{N}^\varepsilon_\lambda(A) \leqslant \frac{1}{2}cR^2\delta^3 \leqslant \frac{1}{2}\mathcal{N}^1_\lambda(A) \qquad (1/R \leqslant \lambda \leqslant R).$$

Therefore $|\mathcal{N}^\varepsilon_\lambda(A) - \mathcal{N}^1_\lambda(A)| \gtrsim R^2\delta^3$ throughout this interval. Squaring, integrating in $\mathrm{d}\lambda/\lambda$, and applying (4.9) gives

$$R^4\delta^6 \log R \lesssim R^4\delta^2 \log\frac{1}{\varepsilon} \lesssim R^4\delta^2\Big(1 + \log\frac{1}{\delta}\Big).$$

Hence $\delta^4 \log R \lesssim 1 + \log(1/\delta)$, and solving this (nonlinear) inequality gives

$$\delta \lesssim \left(\frac{\log\log R}{\log R}\right)^{1/4}.$$

Thus $|A| = \delta R^2 \lesssim R^2(\log\log R/\log R)^{1/4}$, as claimed. □

4.5 Unit-area triangles

Here is a very nice geometric consequence of the proof from the previous section. In the popular Hungarian magazine *Matematikai Lapok*, Erdős gave a short proof that a set $A \subseteq \mathbb{R}^2$ of infinite measure contains vertices of a triangle of area precisely 1. He naturally wondered if the same property is retained by sets A that are only known to have sufficiently large Lebesgue measure, say greater than C for some absolute constant $C \in (0,\infty)$. This question kept reoccurring in Erdős's papers [17, 16, 18] and it is still open at the time of writing.

Our density techniques are not sufficient to handle sets that are scattered though space, but already Theorem gives a partial after obvious stretching of the plane by the factor $\sqrt{2}$ to switch from triangles of area $1/2$ to those of area 1. We can still do slightly better.

Theorem 4.8 (Sets without unit-area triangles [5]). *Let $M_\Delta(R)$ be the supremum of $|A|$ over measurable $A \subseteq [0,R]^2$ containing no triple of points spanning a triangle of area 1. For $R \geqslant 10$,*

$$M_\Delta(R) \lesssim R^2\left(\frac{\log\log R}{\log R}\right)^{1/2}.$$

Note that this theorem is a strengthening of Theorem 4.1 in the special case of right triangles. The problem becomes much more difficult in the ambient space of the minimal (i.e., "critical") dimension for a configuration, which is here the two-dimensional plane.

Proof. After a fixed dilation it is equivalent to exclude triangles of area $1/2$. Define the horizontal form

$$\boxed{\overrightarrow{\mathcal{M}}^0_\lambda(A) = \int_0^\infty \int_{\mathbb{R}^3} \mathbb{1}_A(x,y)\mathbb{1}_A(x+\lambda u, y)\mathbb{1}_A\left(x', y + \frac{1}{\lambda u}\right)\zeta(u)\,\mathrm{d}x\,\mathrm{d}x'\,\mathrm{d}y\,\mathrm{d}u,}$$

and average it over rotations:

$$\boxed{\mathcal{M}^0_\lambda(A) = \frac{1}{2\pi}\int_0^{2\pi} \overrightarrow{\mathcal{M}}^0_\lambda(\mathcal{R}_\theta A)\,\mathrm{d}\theta.}$$

A positive value detects three points spanning area $1/2$, because the two non-horizontal vertices have vertical separation $(\lambda u)^{-1}$ while the horizontal base has length λu. Thus, if A has no such triangle, $\mathcal{M}^0_\lambda(A) = 0$ for all $\lambda > 0$. Define $\mathcal{M}^\varepsilon_\lambda$ by the same one-dimensional smoothings as in the hyperbolic-corner proof:

$$\boxed{\overrightarrow{\mathcal{M}}^\varepsilon_\lambda(A) := \int_0^\infty \int_{\mathbb{R}^3} \mathbb{1}_A(x,y)(\mathbb{1}_A *_1 \phi_{\lambda\varepsilon})(x+\lambda u, y)(\mathbb{1}_A *_2 \phi_{\lambda^{-1}\varepsilon})\Big(x', y + \frac{1}{\lambda u}\Big)\zeta(u)\,\mathrm{d}x\,\mathrm{d}x'\,\mathrm{d}y\,\mathrm{d}u,}$$

$$\boxed{\mathcal{M}^\varepsilon_\lambda(A) := \frac{1}{2\pi}\int_0^{2\pi} \overrightarrow{\mathcal{M}}^\varepsilon_\lambda(\mathcal{R}_\theta A)\,\mathrm{d}\theta.}$$

Structured part. The structured estimate is stronger than for corners:

$$\mathcal{M}^1_\lambda(A) \geqslant c\frac{|A|^3}{R^3} = cR^3\delta^3, \qquad \delta = |A|/R^2, \tag{4.11}$$

whenever $1/R \leqslant \lambda \leqslant R$. It suffices to prove the bound for the horizontal form and then average rotations. The positive kernel again dominates $\mathbb{1}_{[-\lambda,\lambda]\times[-\lambda^{-1},\lambda^{-1}]}$. Partition the x-axis into intervals I of length λ and the y-axis into intervals J of length λ^{-1}. Since the third point has an independent horizontal coordinate x', the contribution of I, J contains

$$|A \cap ([0,R]\times J)| \int_J \sum_I \left(\int_I \mathbb{1}_A(x,y)\,\mathrm{d}x\right)^2 \mathrm{d}y.$$

Cauchy–Schwarz in x, then Jensen in J, gives the lower bound $|A|^3/R^3$, proving (4.11).

Error part. The error term is controlled by the Riesz energy of the set:

$$\mathcal{E}(A) = \iint \frac{\mathbb{1}_A(z)\mathbb{1}_A(z')}{|z-z'|}\,\mathrm{d}z\,\mathrm{d}z'.$$

More precisely,

$$\left(\int_0^\infty |\mathcal{M}^\varepsilon_\lambda(A) - \mathcal{M}^1_\lambda(A)|^2 \frac{\mathrm{d}\lambda}{\lambda}\right)^{1/2} \lesssim \mathcal{E}(A)\left(\log\frac{1}{\varepsilon}\right)^{1/2}. \tag{4.12}$$

For the horizontal form, let $G(y) = \int \mathbb{1}_A(x,y)\,\mathrm{d}x$. The same Plancherel calculation as in (4.9) gives

$$\left(\int_0^\infty |\overrightarrow{\mathcal{M}}^\varepsilon_\lambda(A) - \overrightarrow{\mathcal{M}}^1_\lambda(A)|^2 \frac{\mathrm{d}\lambda}{\lambda}\right)^{1/2} \lesssim \left(\log\frac{1}{\varepsilon}\right)^{1/2} \int_{\mathbb{R}} G(y)^2\,\mathrm{d}y.$$

Apply this to every rotation of A and average in the angle. Expanding the square of the X-ray transform and changing variables from (θ, x, x', y) to the two planar points z, z' gives the Jacobian $|z - z'|$. Consequently

$$\frac{1}{2\pi}\int_0^{2\pi}\int_{\mathbb{R}} G_\theta(y)^2\,\mathrm{d}y\,\mathrm{d}\theta \lesssim \iint \frac{\mathbb{1}_A(z)\mathbb{1}_A(z')}{|z-z'|}\,\mathrm{d}z\,\mathrm{d}z',$$

which proves (4.12).

Uniform part. The uniform term satisfies, for the same smoothing exponent $\sigma > 0$ as before,

$$|\mathcal{M}^0_\lambda(A) - \mathcal{M}^\varepsilon_\lambda(A)| \lesssim \varepsilon^\sigma R|A| = \varepsilon^\sigma R^3\delta. \tag{4.13}$$

For the horizontal form, replace the third indicator by the horizontal-section average

$$g(x,y) = \mathbb{1}_{[0,R]^2}(x,y)\frac{1}{R}\int_0^R \mathbb{1}_A(x',y)\,\mathrm{d}x'.$$

Then $\overrightarrow{\mathcal{M}}$ is R times a corner-type form with inputs $\mathbb{1}_A, \mathbb{1}_A, g$. The proof of (4.10) applies to these inputs and gives $R\varepsilon^\sigma \|\mathbb{1}_A\|_2 \|g\|_2 \lesssim R\varepsilon^\sigma |A|$. Averaging rotations gives (4.13).

Combining bounds. Choose $\varepsilon = c_0\delta^{2/\sigma}$ so that the uniform term is at most one half of the structured term. Since $\mathcal{M}^0_\lambda(A) = 0$, estimates (4.11)–(4.12) imply

$$R^6\delta^6 \log R \lesssim \mathcal{E}(A)^2\left(1 + \log\frac{1}{\delta}\right). \tag{4.14}$$

Hardy–Littlewood–Sobolev alone would give $\mathcal{E}(A) \lesssim |A|^{3/2} = R^3\delta^{3/2}$ and only the weaker exponent $1/3$. To obtain the stated exponent, we improve the energy estimate by induction on the side length of the containing square.[1]

Let

$$\eta(R) = \left(\frac{\log\log R}{\log R}\right)^{1/2}.$$

We prove, for sufficiently large absolute constants C and R_0, that every triangle-free $A \subseteq [0,R]^2$, $R \geqslant R_0$, satisfies $|A| \leqslant CR^2\eta(R)$. Assume this is already known for all side lengths between R_0 and $R/2$. For a fixed $z \in A$, use the layer-cake identity

$$\int \frac{\mathbb{1}_A(z')}{|z-z'|}\,\mathrm{d}z' \leqslant \int_0^{2R} \frac{|A \cap \mathrm{B}(z,r)|}{r^2}\,\mathrm{d}r + \frac{|A|}{2R}.$$

For $r < R_0$ we use the trivial area bound, for $R_0 \leqslant r \leqslant R/4$ we apply the induction hypothesis to $A \cap \mathrm{B}(z,r)$ inside a square of side comparable to r, and for $r \geqslant R/4$ we use $|A| = \delta R^2$. This gives

$$\int \frac{\mathbb{1}_A(z')}{|z-z'|}\,\mathrm{d}z' \leqslant C'R\big(\delta + C\eta(R)\big).$$

Integrating in $z \in A$ yields

$$\mathcal{E}(A) \leqslant C'R^3\delta\big(\delta + C\eta(R)\big). \tag{4.15}$$

Combining (4.14) and (4.15) gives

$$\delta^2 \leqslant C''\frac{(1+\log(1/\delta))^{1/2}}{(\log R)^{1/2}}\big(\delta + C\eta(R)\big).$$

If C is chosen sufficiently large, this inequality forces $\delta \leqslant C\eta(R)$; otherwise the left side dominates both terms on the right for large R. This closes the induction. Therefore $|A| \lesssim R^2(\log\log R/\log R)^{1/2}$. □

[1] The following improvement via the induction on scales was suggested to the authors of [5] by OpenAI's ChatGPT 5.5 Pro.

4.6 Open problems

Problem 4.9 (Rectangular boxes in the critical dimension). For m-dimensional rectangular boxes of sufficiently large volume, our positive result (Theorem 4.5) assumed ambient dimension $d \geqslant m+1$. The critical case $d = m \geqslant 2$ remains open and the problem was also posed in [32].

Problem 4.10 (Parallelograms of prescribed area). It is open whether every finite coloring of $\mathbb{R}^2$ has a color class containing the vertices of a parallelogram of every prescribed area [32]. Partial results from that paper show that any positive proof must deal with nearly degenerate parallelograms and with infinitely many directions.

Problem 4.11 (Unit-area triangles in sets of large finite measure). As we have already mentioned in Section 4.5, Erdős asked whether there is an absolute constant C such that every planar measurable set of area greater than C contains the vertices of a triangle of unit area; see [18]. Mauldin [44, 43] provided some partial results but in a different direction from ours. Erdős even suspected that the smallest such C could be the area of the disk circumscribed around an equilateral triangle of area 1.

Chapter 5

Density theorems for large point configurations

After the prescribed-volume examples, we return to arbitrary finite point patterns. The guiding question is how dense a set must be in order to guarantee every sufficiently large copy of every pattern of a fixed size. Here we are mainly interested in negative results. We refine Bourgain's annular obstruction [3], but also add a nontrivial input from equidistribution theory and Diophantine approximations.

5.1 The threshold problem and polynomial equidistribution

For fixed positive integer d and n, let $\rho(d,n)$ be the smallest threshold such that every measurable $A \subseteq \mathbb{R}^d$ with $\overline{\delta}(A) > \rho(d,n)$ contains, for every n-point pattern $P \subseteq \mathbb{R}^d$, all sufficiently large similar copies of P. The elementary translated-copy argument of Corollary 1.6 gives

$$\rho(d,n) \leqslant 1 - \frac{1}{n},$$

and Falconer, Yavicoli, and the author [19] improved this (slighly but with nontrivial effort) to

$$\rho(d,n) \leqslant 1 - \frac{1}{n-1}$$

for $d, n \geqslant 2$. We recommend [19] to the interested reader, but do not discuss this improvement here. It is natural to wonder about the asymptotic behavior of these numbers as $n \to \infty$. The near-optimal lower bound known available at present is the following.

Theorem 5.1 (Near-optimal Euclidean lower bound [34]). *There exists an absolute constant $C > 0$ such that, for every $d \geqslant 1$ and every sufficiently large n, there exist a measurable set $E \subseteq \mathbb{R}^d$, an n-point configuration $P \subseteq \mathbb{R}^d$ contained in a line, and a sequence $\lambda_j \to \infty$ such that*

$$\mathrm{d}(E) \geqslant 1 - C\frac{\log n}{n},$$

but E contains no Euclidean isometric copy of $\lambda_j P$ for any j.

Consequently,

$$\rho(d,n) \geqslant 1 - C\frac{\log n}{n}$$

for all sufficiently large n. Together with the upper bound from Corollary 1.6, this leaves only a logarithmic gap in the density threshold.

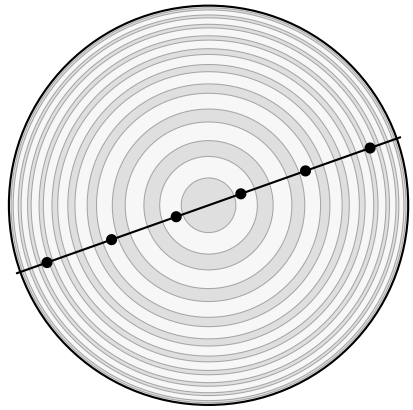

Figure 5.1: A typical annular construction underlying negative density results, together with a collinear pattern that is forced to interact with the annular gaps.

We will prove a slightly more general ℓ^p result; see Theorem 5.3 below. Its number-theoretic core is a polynomial density property on the circle. Write $\|x\|_{\mathbb{T}} = \operatorname{dist}(x, \mathbb{Z})$ for $x \in \mathbb{T} = \mathbb{R}/\mathbb{Z}$ and $e(t) = e^{2\pi i t}$.

Proposition 5.2 (Uniform polynomial density modulo 1). *Fix an integer $p \geqslant 2$. There is a constant $K_p < \infty$ such that, for every sufficiently large n, one can find a set $P \subseteq \mathbb{Z}$ with $|P| = n$ and a real number α with the following property. For every $\beta_1, \ldots, \beta_{p-1} \in \mathbb{R}$, the set*

$$\{(\alpha k^p + \beta_{p-1} k^{p-1} + \cdots + \beta_1 k) \bmod 1 : k \in P\}$$

meets every interval in $\mathbb{T}$ of length

$$\varepsilon_n = K_p \frac{\log n}{n}.$$

Proof. Choose a prime Q satisfying

$$n^{2^p} < Q < 2n^{2^p},$$

which is possible for all large n by Bertrand's postulate, and put $\alpha = 1/Q$. For $\beta = (\beta_1, \ldots, \beta_{p-1}) \in \mathbb{T}^{p-1}$ define

$$x_k(\beta) = \alpha k^p + \beta_{p-1} k^{p-1} + \cdots + \beta_1 k \pmod 1, \qquad 0 \leqslant k < Q.$$

We first show that the full collection $\{x_k(\beta) : 0 \leqslant k < Q\}$ is well distributed, uniformly in β.

Let $D_Q(\beta)$ be the discrepancy of this collection. The Erdős–Turán inequality [35] gives, for any $M \geqslant 1$,

$$D_Q(\beta) \lesssim \frac{1}{M} + \sum_{m=1}^{M} \frac{1}{m} \left| \frac{1}{Q} \sum_{k=0}^{Q-1} e(m x_k(\beta)) \right|.$$

Take $M = Q - 1$. For $1 \leqslant m \leqslant Q - 1$, the leading coefficient of the polynomial $m x_k(\beta)$ is m/Q, already in lowest terms because Q is prime. Weyl's inequality [54], applied with denominator Q and with a small exponent loss $\eta = 2^{-p}$, yields

$$\left| \sum_{k=0}^{Q-1} e(m x_k(\beta)) \right| \lesssim_p Q^{1-2^{-p}},$$

uniformly in m and in all lower-order coefficients. Hence

$$D_Q(\beta) \lesssim_p \frac{1}{Q} + Q^{-2^{-p}} \sum_{m=1}^{Q-1} \frac{1}{m} \lesssim_p \frac{\log Q}{Q^{2^{-p}}} \lesssim_p \frac{\log n}{n}.$$

Choose K_p so large that $D_Q(\beta) \leqslant \varepsilon_n/10$ for every β. Then every interval $I \subseteq \mathbb{T}$ of length $9\varepsilon_n/10$ contains at least

$$\left(\frac{9\varepsilon_n}{10} - \frac{\varepsilon_n}{10}\right) Q = \frac{4}{5}\varepsilon_n Q$$

of the points $x_k(\beta)$.

We now discretize the coefficient torus. For $i = 1, \ldots, p-1$, choose a grid $\mathcal{N}_i \subseteq \mathbb{T}$ of mesh

$$\Delta_i = \frac{\varepsilon_n}{100pQ^i},$$

so $|\mathcal{N}_i| \lesssim Q^i/\varepsilon_n$. Put $\mathcal{N} = \mathcal{N}_1 \times \cdots \times \mathcal{N}_{p-1}$; then

$$|\mathcal{N}| \lesssim_p \frac{Q^{p(p-1)/2}}{\varepsilon_n^{p-1}}.$$

Let $\mathcal{I}$ be the family of intervals of length $9\varepsilon_n/10$ whose left endpoints are integer multiples of $\varepsilon_n/100$. Then $|\mathcal{I}| \lesssim 1/\varepsilon_n$.

For $b \in \mathcal{N}$ and $I \in \mathcal{I}$, set

$$R_{b,I} = \{0 \leqslant k < Q : x_k(b) \in I\}.$$

The discrepancy estimate gives $|R_{b,I}| \geqslant (4/5)\varepsilon_n Q$. Choose an n-element subset

$$P \subseteq \{0, 1, 2, \ldots, Q-1\}$$

uniformly at random.[1] The probability that P misses this particular $R_{b,I}$ is at most

$$\frac{\binom{Q-|R_{b,I}|}{n}}{\binom{Q}{n}} \leqslant \left(1 - \frac{|R_{b,I}|}{Q}\right)^n \leqslant \exp\left(-\frac{4}{5}\varepsilon_n n\right).$$

By the union bound, the probability that some pair (b, I) is missed is at most

$$|\mathcal{N}|\,|\mathcal{I}| \exp\left(-\frac{4}{5}\varepsilon_n n\right) \lesssim_p \frac{Q^{p(p-1)/2}}{\varepsilon_n^p} \exp\left(-\frac{4}{5}K_p \log n\right).$$

Since $Q \asymp n^{2^p}$ and $\varepsilon_n = K_p(\log n)/n$, the last expression is < 1 for all large n if K_p is chosen sufficiently large. Therefore there exists a deterministic n-element set P such that

$$\{x_k(b) : k \in P\} \cap I \neq \emptyset \qquad \text{for every } b \in \mathcal{N},\ I \in \mathcal{I}. \tag{5.1}$$

It remains to pass from grid coefficients to arbitrary coefficients. Let $\beta \in \mathbb{T}^{p-1}$ and let $J \subseteq \mathbb{T}$ be an interval of length ε_n. Choose $b \in \mathcal{N}$ with $\|\beta_i - b_i\|_{\mathbb{T}} \leqslant \Delta_i$. Since $P \subseteq \{0, \ldots, Q-1\}$,

$$\|x_k(\beta) - x_k(b)\|_{\mathbb{T}} \leqslant \sum_{i=1}^{p-1} \Delta_i Q^i \leqslant \frac{\varepsilon_n}{100} \qquad (k \in P).$$

Let $J' \subseteq J$ be the concentric subinterval of length $96\varepsilon_n/100$. It contains some $I \in \mathcal{I}$. By (5.1), choose $k \in P$ with $x_k(b) \in I \subseteq J'$. The preceding estimate then gives $x_k(\beta) \in J$. Thus the desired polynomial set meets every interval of length ε_n. $\square$

[1] This randomization argument and the remainder of the proof of the proposition were suggested to the authors of [34] by OpenAI's ChatGPT 5.4 Pro.

5.2 Collinear ℓ^p obstructions

Theorem 5.3 (Collinear ℓ^p lower bound [34]). *Fix integers $d \geqslant 1$ and $p \geqslant 2$. There exists $C_{d,p} \in (0,\infty)$ such that, for all sufficiently large n, there exist a measurable set $A \subseteq \mathbb{R}^d$, an n-point configuration $P \subseteq \mathbb{R}^d$ contained in a line, and scales $\lambda_j \to \infty$ such that*

$$\underline{\mathrm{d}}(A) \geqslant 1 - C_{d,p}\frac{\log n}{n},$$

and A contains no ℓ^p-isometric copy of $\lambda_j P$ for any j. If p is even, $C_{d,p}$ may be chosen independently of d, and A may be chosen to have an actual density.

Proof. Let $P \subseteq \mathbb{Z}$ and α be supplied by Proposition 5.2, and put

$$\varepsilon_n = K_p \frac{\log n}{n}, \qquad \lambda_j = (\alpha + j)^{1/p}$$

for all sufficiently large positive integers j.

First assume that p is even. Define the annular set

$$A = \left\{ x \in \mathbb{R}^d : \operatorname{dist}(\|x\|_p^p, \mathbb{Z}) < \frac{1-\varepsilon_n}{2} \right\}.$$

It is not difficult to show that $\mathrm{d}(A) = 1 - \varepsilon_n$. Suppose that A contained an ℓ^p-isometric copy of $\lambda_j P$. Since P is collinear, its affine copy is collinear as well, so one can write it as

$$\{x + \lambda_j k v : k \in P\} \tag{5.2}$$

with $\|v\|_p = 1$. (Interestingly, here we used the fact that the ℓ^p space is strictly convex for $1 < p < \infty$.) For $k \in P$, the binomial theorem gives

$$\begin{aligned} \|x + \lambda_j k v\|_p^p &= \sum_{i=1}^{d} \sum_{l=0}^{p} \binom{p}{l} x_i^{p-l} \lambda_j^l k^l v_i^l \\ &= (\alpha + j) k^p + \beta_{p-1} k^{p-1} + \cdots + \beta_1 k + \beta_0 \end{aligned}$$

for suitable real coefficients $\beta_0, \ldots, \beta_{p-1}$; the leading coefficient is $(\alpha + j)\|v\|_p^p = \alpha + j$. Since all copied points lie in A, the values of this polynomial modulo 1 all lie in the interval

$$\left(-\frac{1-\varepsilon_n}{2}, \frac{1-\varepsilon_n}{2} \right) \subseteq \mathbb{T}$$

of length $1 - \varepsilon_n$. Subtracting the constant β_0 and the integer jk^p shows that

$$\{\alpha k^p + \beta_{p-1} k^{p-1} + \cdots + \beta_1 k \bmod 1 : k \in P\}$$

is contained in an interval of length $1 - \varepsilon_n$. The complementary interval has length ε_n, contradicting Proposition 5.2. Thus no such copy exists.

Now assume that p is odd. For each sign vector $\sigma = (\sigma_1, \ldots, \sigma_d) \in \{-1, 1\}^d$ set

$$F_\sigma(x) = \sum_{i=1}^{d} \sigma_i x_i^p, \qquad A_\sigma = \left\{ x : \operatorname{dist}(F_\sigma(x), \mathbb{Z}) < \frac{1-\varepsilon_n}{2} \right\},$$

and define

$$A = \bigcap_{\sigma \in \{-1,1\}^d} A_\sigma.$$

Each A_σ has density $1 - \varepsilon_n$ in large cubes $[-R/2, R/2]^d$ up to an $O(R^{d-1})$ error. Therefore, by the union bound on complements,

$$\underline{\mathrm{d}}(A) \geqslant 1 - 2^d \varepsilon_n.$$

If an ℓ^p-isometric copy of $\lambda_j P$ lay in A, we would again write it as (5.2) with $\|v\|_p = 1$. Choose signs so that $\sigma_i v_i^p = |v_i|^p$ for every i. Then

$$F_\sigma(x + \lambda_j k v) = (\alpha + j)k^p + \beta_{p-1}k^{p-1} + \cdots + \beta_1 k + \beta_0,$$

again because the leading coefficient is

$$\lambda_j^p \sum_i \sigma_i v_i^p = (\alpha + j) \sum_i |v_i|^p = \alpha + j.$$

Since the copied points lie in $A \subseteq A_\sigma$, the same interval-complement argument contradicts Proposition 5.2. This proves the theorem, with $C_{d,p} = K_p$ for even p and $C_{d,p} = 2^d K_p$ for odd p. □

Proof of Theorem 5.1. Take $p = 2$ in Theorem 5.3. The norm ℓ^2 is the Euclidean norm, the constructed pattern is collinear, the constructed set has actual density at least $1 - K_2(\log n)/n$, and the missing ℓ^2-isometric copies are exactly missing Euclidean isometric copies. □

5.3 Non-collinear ℓ^p patterns

For $p \neq 2$ the geometry of ℓ^p spaces supplies a simpler obstruction if non-collinear patterns are allowed.

A well-known result that will be useful are Clarkson's inequalities [7]. Let $1 < p < \infty$, $p \neq 2$, and let $u, v \in \mathbb{R}^d$. Then

$$\|u+v\|_p^p + \|u-v\|_p^p \begin{cases} \geqslant 2(\|u\|_p^p + \|v\|_p^p), & p > 2, \\ \leqslant 2(\|u\|_p^p + \|v\|_p^p), & 1 < p < 2. \end{cases}$$

In either case equality holds if and only if u and v have disjoint coordinate supports.

Theorem 5.4 (Asymptotically sharp ℓ^p obstruction [34]). *Let $d \geqslant 1$ and $p \in (1, \infty)$, $p \neq 2$. For every $n \geqslant 2d+1$ there exist a measurable set $A \subseteq \mathbb{R}^d$, an n-point configuration $P \subseteq \mathbb{R}^d$, and scales $\lambda_j \to \infty$ such that A contains no ℓ^p-isometric copy of $\lambda_j P$, while*

$$\mathrm{d}(A) = 1 - \frac{1}{n - 2d + 2}.$$

Combining this with the trivial upper bounds from Corollary 1.6, we get sharp asymptotics for the obviously defined ℓ^p analogues of the numbers $\rho(d, n)$:

$$\rho(d, n, p) = 1 - \frac{1}{n} + O_d\left(\frac{1}{n^2}\right).$$

Proof. Let $e_1, \dots, e_d$ be the standard basis and set

$$P = \{ke_1 : k = -1, 0, 1, \dots, n-2d\} \cup \{\pm e_2, \dots, \pm e_d\}.$$

The first set has $n - 2d + 2$ points and the second has $2d - 2$ points, so $|P| = n$. Put

$$\varepsilon = \frac{1}{n-2d+2}, \qquad A = \{x \in \mathbb{R}^d : (x_1 + \cdots + x_d) \bmod 1 \in [0, 1-\varepsilon)\}.$$

Fubini in the x_1 variable gives

$$|A \cap [-R/2, R/2]^d| = (1-\varepsilon)R^d + O_d(R^{d-1}),$$

so $\mathrm{d}(A) = 1 - \varepsilon$. Let

$$\lambda_j = j + \varepsilon, \qquad j = 1, 2, \dots.$$

We prove that no ℓ^p-isometric copy of $\lambda_j P$ lies in A.

Assume, for contradiction, that such a copy exists. The collinear part

$$\{\lambda_j k e_1 : k = -1, 0, 1, \dots, n-2d\}$$

must be mapped to points

$$y_k = x + \lambda_j k u, \qquad k = -1, 0, 1, \dots, n-2d,$$

where $\|u\|_p = 1$. If $d = 1$, then $u = \pm e_1$. Assume $d \geqslant 2$. Let z_i^+ and z_i^- be the images of $\lambda_j e_i$ and $-\lambda_j e_i$ for $i = 2, \dots, d$, and write

$$z_i^\pm = x + \lambda_j v_i^\pm.$$

Then $\|v_i^\pm\|_p = 1$.

Because the distance between e_i and $-e_i$ is 2, we have

$$\|v_i^+ - v_i^-\|_p = 2 = \|v_i^+\|_p + \|-v_i^-\|_p.$$

Strict convexity of ℓ^p implies equality in the triangle inequality only for positively collinear vectors; hence

$$v_i^- = -v_i^+. \tag{5.3}$$

Next, the model distances

$$\|e_i - e_1\|_p^p = \|e_i - (-e_1)\|_p^p = 2$$

become

$$\|v_i^+ - u\|_p^p = \|v_i^+ + u\|_p^p = 2.$$

Adding the two equalities gives

$$\|v_i^+ - u\|_p^p + \|v_i^+ + u\|_p^p = 4 = 2(\|v_i^+\|_p^p + \|u\|_p^p).$$

By the equality characterization in Clarkson's inequalities, v_i^+ and u have disjoint coordinate supports.

Similarly, for $2 \leqslant i < m \leqslant d$, the model distances

$$\|e_i - e_m\|_p^p = \|e_i + e_m\|_p^p = 2$$

and (5.3) give

$$\left\|v_i^+ - v_m^+\right\|_p^p = \left\|v_i^+ + v_m^+\right\|_p^p = 2.$$

Clarkson's inequality is again saturated with an equality, which shows that v_i^+ and v_m^+ also have disjoint coordinate supports. Thus the d nonzero unit vectors

$$u, v_2^+, \ldots, v_d^+$$

have pairwise disjoint nonempty supports inside the d coordinate set $\{1, \ldots, d\}$. Each support must therefore be a singleton. In particular,

$$u = \sigma e_l$$

for some $l \in \{1, \ldots, d\}$ and $\sigma \in \{-1, 1\}$.

All points $y_k = x + \sigma\lambda_j k e_l$ lie in A. If $c = x_1 + \cdots + x_d$, then

$$c + \sigma\lambda_j k \bmod 1 \in [0, 1-\varepsilon) \qquad (k = -1, 0, 1, \ldots, n-2d).$$

Since $\lambda_j = j + \varepsilon$, this says that a translate of the set

$$\{\sigma k\varepsilon \bmod 1 : k = -1, 0, 1, \ldots, n-2d\}$$

is contained in an interval of length $1-\varepsilon$. But this is exactly the set of $N = n - 2d + 2$ equally spaced points on the circle, where $\varepsilon = 1/N$. No half-open interval of length $1 - 1/N$ contains all N equally spaced points. This contradiction completes the proof. $\square$

5.4 Open problem

Problem 5.5 (Sharp density threshold for arbitrary Euclidean patterns). Just as before (see Section 5.1), let $\rho(d, n)$ denote the least density threshold forcing all sufficiently large similar copies of every n-point Euclidean pattern in $\mathbb{R}^d$. The best known bounds, due respectively to Santos Sepčić and the author [34] (Theorem 5.1) and to Falconer, Yavicoli, and the author [19], are

$$1 - C\frac{\log n}{n} \leqslant \rho(d, n) \leqslant 1 - \frac{1}{n-1}$$

for $d \geqslant 2$ and sufficiently large n. The natural problem is to remove the logarithm in the lower bound or to improve the upper bound. Closing the gap likely requires a new idea on one side of this dichotomy.

Preliminaries

This appendix records some technical background used throughout the notes.

A.1 Comparison notation

We use standard asymptotic notation throughout the notes. If X and Y are nonnegative quantities, then

$$X \lesssim Y$$

means that $X \leqslant CY$ for a finite constant C independent of the main variables under discussion. A subscript records allowed dependence: for example $X \lesssim_{d,n,\varepsilon} Y$ means that C may depend on d, n, ε, but not on scale parameters such as R or λ unless this is explicitly stated. We write

$$X \gtrsim Y$$

for $Y \lesssim X$, and

$$X \sim Y$$

when both inequalities hold. The notation

$$X = O(Y)$$

has the same meaning as $|X| \lesssim Y$; similarly $X = O_{d,n}(Y)$ permits the implicit constant to depend on d and n. For nonnegative quantities,

$$X = \Omega(Y)$$

means $X \gtrsim Y$, again with optional subscripts to indicate permitted parameter dependence. Thus a statement such as $|A| = \Omega_d(R^d)$ asserts that $|A| \geqslant c_d R^d$ for some positive constant depending only on d.

A.2 Upper Banach density and averaging bodies

We first justify two properties that are used throughout the notes. The definition of upper Banach density contains a genuine limit, not merely a limsup; and the value of that limit does not depend on choosing cubes rather than balls or any other fixed compact convex body. This is the elementary Følner comparison argument [21], specialized to dilates of bounded sets in Euclidean space.

For a bounded measurable set $F \subseteq \mathbb{R}^d$ with $0 < |F| < \infty$, define

$$D_F(E) := \sup_{x \in \mathbb{R}^d} \frac{|E \cap (x + F)|}{|F|}.$$

Thus the cube definition in Chapter 1 is the assertion that $D_{[0,R]^d}(E)$ has a limit as $R \to \infty$.

Lemma A.1 (Comparison of two averaging windows). *Let $F, G \subseteq \mathbb{R}^d$ be bounded measurable sets with positive finite measure. For every measurable $E \subseteq \mathbb{R}^d$,*

$$D_F(E) \leqslant D_G(E) + \eta(F, G), \qquad \eta(F,G) := \sup_{z \in G} \frac{|(F+z) \triangle F|}{|F|}.$$

Proof. Let

$$\mathcal{A}_F f(x) := \frac{1}{|F|} \int_F f(x+y)\ \mathrm{d}y.$$

Then $D_F(E) = \|\mathcal{A}_F \mathbb{1}_E\|_\infty$. For $f = \mathbb{1}_E$ and for every $z \in G$,

$$\left|\mathcal{A}_F f(x+z) - \mathcal{A}_F f(x)\right| \leqslant \frac{|(F+z) \triangle F|}{|F|} \leqslant \eta(F,G).$$

Averaging this inequality over $z \in G$ gives

$$\mathcal{A}_F f(x) \leqslant \frac{1}{|G|} \int_G \mathcal{A}_F f(x+z)\ \mathrm{d}z + \eta(F,G) = \mathcal{A}_F(\mathcal{A}_G f)(x) + \eta(F,G).$$

The operator $\mathcal{A}_F$ is a positive average, so

$$\mathcal{A}_F(\mathcal{A}_G f)(x) \leqslant \|\mathcal{A}_G f\|_\infty = D_G(E).$$

Taking the supremum over x proves the claim. □

Lemma A.2 (Mutual dependence of convex dilates). *Let $K, L \subseteq \mathbb{R}^d$ be compact convex sets with nonempty interior. For every fixed $S > 0$,*

$$\eta(RK, SL) \longrightarrow 0 \qquad \text{as } R \to \infty.$$

Proof. Since L is bounded, all $z \in SL$ satisfy $|z| \leqslant C_S$ for some constant C_S. Also

$$\frac{|(RK+z) \triangle RK|}{|RK|} = \frac{|(K+z/R) \triangle K|}{|K|}.$$

The map $u \mapsto \mathbb{1}_{K+u}$ is continuous at $u = 0$ in $L^1(\mathbb{R}^d)$, because translations are continuous in L^1. Hence, for every $\varepsilon > 0$, there is $\delta > 0$ such that

$$|u| < \delta \quad \Longrightarrow \quad |(K+u) \triangle K| < \varepsilon |K|.$$

If $R > C_S/\delta$, then $|z/R| < \delta$ for every $z \in SL$, and the displayed supremum is at most ε. □

Proposition A.3 (Existence of the upper-density limit). *Let $K \subseteq \mathbb{R}^d$ be a compact convex set with nonempty interior. Then, for every measurable $E \subseteq \mathbb{R}^d$, the limit*

$$\lim_{R \to \infty} D_{RK}(E)$$

exists. More precisely,

$$\lim_{R \to \infty} D_{RK}(E) = \inf_{S > 0} D_{SK}(E).$$

Proof. Fix $S > 0$ and apply Lemma A.1 with $F = RK$ and $G = SK$:

$$D_{RK}(E) \leqslant D_{SK}(E) + \eta(RK, SK).$$

By Lemma A.2, the error tends to 0 as $R \to \infty$. Therefore

$$\limsup_{R\to\infty} D_{RK}(E) \leqslant D_{SK}(E).$$

Taking the infimum over S gives

$$\limsup_{R\to\infty} D_{RK}(E) \leqslant \inf_{S>0} D_{SK}(E).$$

The quantity on the right is a lower bound for every value $D_{RK}(E)$, and hence it is at most $\liminf_{R\to\infty} D_{RK}(E)$. Thus the limsup and the liminf are equal, and both equal the stated infimum. $\square$

Theorem A.4 (Independence of the averaging body). *Let $K, L \subseteq \mathbb{R}^d$ be compact convex sets with nonempty interior. Then, for every measurable $E \subseteq \mathbb{R}^d$,*

$$\lim_{R\to\infty} D_{RK}(E) = \lim_{R\to\infty} D_{RL}(E).$$

Consequently the upper Banach density may be computed from translates and dilates of any such K:

$$\overline{\delta}(E) = \lim_{R\to\infty} \sup_{x\in\mathbb{R}^d} \frac{|E \cap (x + RK)|}{R^d|K|}.$$

In particular, cubes, balls, ellipsoids, simplices, and all compact convex bodies with nonempty interior give the same number.

Proof. Write

$$\overline{\delta}_K(E) := \lim_{R\to\infty} D_{RK}(E), \qquad \overline{\delta}_L(E) := \lim_{R\to\infty} D_{RL}(E),$$

which exist by Proposition A.3. Fix $S > 0$ and compare the large window RK with the fixed window SL:

$$D_{RK}(E) \leqslant D_{SL}(E) + \eta(RK, SL).$$

Letting $R \to \infty$ and using Lemma A.2, we obtain

$$\overline{\delta}_K(E) \leqslant D_{SL}(E).$$

Taking the infimum in S and using Proposition A.3 for L gives $\overline{\delta}_K(E) \leqslant \overline{\delta}_L(E)$. Interchanging K and L gives the opposite inequality.

Finally, $|RK| = R^d|K|$, so the displayed formula is exactly the definition of $D_{RK}(E)$. The location of K is irrelevant: replacing K by $K + a$ changes $x + RK$ to $(x + Ra) + RK$, and the translation parameter x already ranges over all of $\mathbb{R}^d$. $\square$

A.3 Surface measures and decay

This section records the Fourier facts about curved surface measure that are used in the uniform estimates in the main text. We use the Fourier normalization

$$\widehat{f}(\xi) = \int_{\mathbb{R}^d} f(x) e^{-2\pi i x\cdot\xi}\,\mathrm{d}x.$$

Let $\Sigma \subseteq \mathbb{R}^d$ be a compact C^∞ hypersurface and let σ be a smooth compactly supported surface measure on Σ; in the most important example $\Sigma = S^{d-1}$ and σ is normalized spherical measure. If the Gaussian curvature of Σ is everywhere nonzero on the support of σ, then stationary phase gives

$$|\widehat{\sigma}(\xi)| \lesssim_\Sigma (1+|\xi|)^{-(d-1)/2}. \tag{A.1}$$

For the sphere this estimate can also be read from the Bessel-function formula for $\widehat{\sigma}$ and the standard asymptotics of Bessel functions; see Abramowitz–Stegun [1]. In this generality it is a classical theorem of Littman and Herz, and it is a standard consequence of the stationary-phase theorem as presented, for instance, in Stein [48], Littman [39], Herz [28], and Sogge [47].

Here is the sketch of the stationary-phase property in the form needed in these notes. Write $\xi = \rho\theta$, where $\rho = |\xi|$ and $\theta \in S^{d-1}$. On a coordinate patch $u \mapsto \Phi(u) \in \Sigma$,

$$\widehat{\sigma}(\rho\theta) = \int e^{-2\pi i\rho\theta\cdot\Phi(u)} a(u)\,\mathrm{d}u.$$

The critical points are precisely the points at which θ is normal to Σ. Non-vanishing Gaussian curvature says that the Hessian of $u \mapsto \theta\cdot\Phi(u)$ is non-singular at such a point. A partition of unity and the stationary-phase theorem therefore give, uniformly in θ,

$$\widehat{\sigma}(\rho\theta) = \rho^{-(d-1)/2} \sum_{u_c} a_c(\theta) e^{-2\pi i\rho\theta\cdot\Phi(u_c)} + O_\Sigma(\rho^{-(d+1)/2}), \qquad \rho \geqslant 1,$$

where the sum is over the finitely many critical points in the coordinate patches. Away from the critical set, repeated integration by parts gives rapid decay. Together with the trivial bound $|\widehat{\sigma}(\xi)| \leqslant \|\sigma\|$ for $|\xi| \lesssim 1$, this proves (A.1).

The same estimate applies to dilates. If σ_λ denotes the image of σ under $x \mapsto \lambda x$, then

$$\widehat{\sigma_\lambda}(\xi) = \widehat{\sigma}(\lambda\xi), \qquad |\widehat{\sigma_\lambda}(\xi)| \lesssim (1+\lambda|\xi|)^{-(d-1)/2}.$$

This is the estimate used whenever an exact counting form contains a spherical constraint such as $|y| = \lambda$.

The other recurring ingredient is a smooth approximate identity. Let

$$g(x) = e^{-\pi|x|^2}, \qquad g_t(x) = t^{-d} g(x/t),$$

and let $k = \Delta g$. Then

$$\widehat{g}(\xi) = e^{-\pi|\xi|^2}, \qquad \widehat{k_t}(\xi) = \widehat{k}(t\xi) = -4\pi^2 t^2|\xi|^2 e^{-\pi t^2|\xi|^2}.$$

Consequently, when σ satisfies (A.1),

$$\sup_{\xi\in\mathbb{R}^d} |\widehat{\sigma}(\lambda\xi)\widehat{k}(t\lambda\xi)| \lesssim t^{\min\{(d-1)/2,2\}}, \qquad 0 < t \leqslant 1.$$

Indeed, with $r = \lambda|\xi|$ the left side is bounded by

$$(1+r)^{-(d-1)/2}(tr)^2 e^{-\pi t^2 r^2}.$$

The maximum occurs either for $r \lesssim 1$ or for $r \sim t^{-1}$; the displayed bound follows in both ranges, and the intermediate range is monotone up to harmless constants. In the planar circular estimates used repeatedly in Chapters 2 and 3 this gives the familiar bound

$$\sup_{\xi\in\mathbb{R}^2} |\widehat{\sigma}(\lambda\xi)\widehat{k}(t\lambda\xi)| \lesssim t^{1/2}.$$

A closely related smoothing estimate is also used. Since

$$1-\widehat{g}(\varepsilon\lambda\xi) = 1-e^{-\pi\varepsilon^2\lambda^2|\xi|^2},$$

one has, for every $0<\gamma\leqslant \min\{(d-1)/2, 2\}$,

$$\sup_{\xi\in\mathbb{R}^d} |\widehat{\sigma}(\lambda\xi)(1-\widehat{g}(\varepsilon\lambda\xi))| \lesssim_\gamma \varepsilon^\gamma.$$

This follows by writing $u=\lambda|\xi|$ and bounding

$$(1+u)^{-(d-1)/2}\min\{1,\varepsilon^2u^2\}.$$

Plancherel then gives the model L^2 estimate

$$\left|\int F(x)\big(F*\sigma_\lambda - F*\sigma_\lambda * g_{\varepsilon\lambda}\big)(x)\,\mathrm{d}x\right| \leqslant C_\gamma \varepsilon^\gamma \|F\|^2_{L^2(\mathbb{R}^d)},$$

which is the prototype for the uniform part of Bourgain's method. More complicated configurations require the same idea after freezing all but one variable, or after applying a multilinear smoothing inequality, but the local oscillatory input is still usually just (A.1).

A.4 Gowers norms

The cube notation used in the notes is the continuum analogue of the Gowers uniformity norms introduced in quantitative proofs of Szemerédi's theorem [23]; see also the expositions and related estimates in Green–Tao [26], Tao [51], Shkredov [46], and Eisner–Tao [15]. We record the precise Cauchy–Schwarz inequality invoked in the main text.

Take $k\geqslant 1$. If $h=(h_1,\dots,h_k)\in(\mathbb{R}^d)^k$ and $\omega=(\omega_1,\dots,\omega_k)\in\{0,1\}^k$, write

$$\omega\cdot h = \omega_1 h_1+\cdots+\omega_k h_k, \qquad |\omega| = \omega_1+\cdots+\omega_k.$$

Let $\mathcal{C}z=\overline{z}$. For a family of bounded compactly supported functions $(f_\omega)_{\omega\in\{0,1\}^k}$ define the *Gowers inner product*

$$\langle f_\omega : \omega\in\{0,1\}^k\rangle_{U^k} := \int_{\mathbb{R}^d}\int_{(\mathbb{R}^d)^k} \prod_{\omega\in\{0,1\}^k} \mathcal{C}^{|\omega|} f_\omega(x+\omega\cdot h)\,\mathrm{d}h_1\cdots\mathrm{d}h_k\,\mathrm{d}x.$$

The *Gowers (semi)norm* or the *uniformity (semi)norm* or just the U^k *(semi)norm* is defined by putting all functions equal:

$$\|f\|^{2^k}_{U^k(\mathbb{R}^d)} = \langle f : \omega\in\{0,1\}^k\rangle_{U^k}.$$

For the nonnegative indicator functions used in the density arguments, the conjugations do not change the value and the formula becomes the positive cube integral

$$\int_{\mathbb{R}^d} \int_{(\mathbb{R}^d)^k} \prod_{\omega \in \{0,1\}^k} f(x + \omega \cdot h) \, \mathrm{d}h_1 \cdots \mathrm{d}h_k \, \mathrm{d}x.$$

Proposition A.5 (Gowers–Cauchy–Schwarz inequality). *For every $k \geqslant 1$ and every bounded compactly supported family $(f_\omega)_{\omega \in \{0,1\}^k}$,*

$$\left|\langle f_\omega : \omega \in \{0,1\}^k \rangle_{U^k}\right| \leqslant \prod_{\omega \in \{0,1\}^k} \|f_\omega\|_{U^k(\mathbb{R}^d)} .$$

Proof. The proof is the usual repeated Cauchy–Schwarz argument, written here in a form that is directly compatible with the Euclidean integrals in the notes. The case $k = 1$ is immediate: by Fubini,

$$\langle f_0, f_1 \rangle_{U^1} = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f_0(x) \overline{f_1(x+h)} \, \mathrm{d}h \, \mathrm{d}x = \left(\int f_0 \right) \overline{\left(\int f_1 \right)},$$

and this is bounded by $\|f_0\|_{U^1} \, \|f_1\|_{U^1}$.

We first prove a two-function estimate that will be used in the induction. For $k \geqslant 2$ and $s \in \mathbb{R}^d$, write $T_s g(x) = g(x+s)$. We claim that

$$\int_{\mathbb{R}^d} \left\| f \, \overline{T_s g} \right\|_{U^{k-1}}^{2^{k-1}} \, \mathrm{d}s \leqslant \|f\|_{U^k}^{2^{k-1}} \, \|g\|_{U^k}^{2^{k-1}} . \tag{A.2}$$

Expanding the left side gives

$$\int_{s,x,h'} \prod_{\eta \in \{0,1\}^{k-1}} \mathcal{C}^{|\eta|} \big(f(x + \eta \cdot h') \overline{g(x + s + \eta \cdot h')} \big) \, \mathrm{d}h' \, \mathrm{d}x \, \mathrm{d}s,$$

where $h' = (h_1, \ldots, h_{k-1})$. Put $y = x + s$. The expression becomes

$$\int_{h'} F(h') \overline{G(h')} \, \mathrm{d}h',$$

where

$$F(h') = \int_{\mathbb{R}^d} \prod_{\eta \in \{0,1\}^{k-1}} \mathcal{C}^{|\eta|} f(x + \eta \cdot h') \, \mathrm{d}x, \qquad G(h') = \int_{\mathbb{R}^d} \prod_{\eta \in \{0,1\}^{k-1}} \mathcal{C}^{|\eta|} g(y + \eta \cdot h') \, \mathrm{d}y.$$

By Cauchy–Schwarz in h' this is at most

$$\left(\int |F(h')|^2 \, \mathrm{d}h' \right)^{1/2} \left(\int |G(h')|^2 \, \mathrm{d}h' \right)^{1/2} .$$

Finally, after writing the second copy of the base variable as $x + h_k$, the first factor is exactly $\|f\|_{U^k}^{2^{k-1}}$, and similarly the second is $\|g\|_{U^k}^{2^{k-1}}$. This proves (A.2).

We now prove the proposition by induction on k. Assume that it has been proved for $k-1$. Write a vertex of $\{0,1\}^k$ as (η, ϵ) with $\eta \in \{0,1\}^{k-1}$ and $\epsilon \in \{0,1\}$. Separating the last difference variable $s = h_k$, the k-dimensional inner product can be written as

$$\int_{\mathbb{R}^d} \left\langle f_{\eta,0} \, \overline{T_s f_{\eta,1}} : \eta \in \{0,1\}^{k-1} \right\rangle_{U^{k-1}} \, \mathrm{d}s.$$

The induction hypothesis bounds the absolute value of the inner product by

$$\prod_{\eta \in \{0,1\}^{k-1}} \left\| f_{\eta,0} \, \overline{T_s f_{\eta,1}} \right\|_{U^{k-1}} .$$

Integrating in s and applying Hölder with the 2^{k-1} equal exponents gives

$$\left| \langle f_\omega : \omega \in \{0,1\}^k \rangle_{U^k} \right| \leqslant \prod_{\eta \in \{0,1\}^{k-1}} \left(\int_{\mathbb{R}^d} \left\| f_{\eta,0} \, \overline{T_s f_{\eta,1}} \right\|_{U^{k-1}}^{2^{k-1}} \, \mathrm{d}s \right)^{1/2^{k-1}} .$$

Applying the two-function estimate (A.2) to each pair $(f_{\eta,0}, f_{\eta,1})$ yields

$$\left| \langle f_\omega : \omega \in \{0,1\}^k \rangle_{U^k} \right| \leqslant \prod_{\eta \in \{0,1\}^{k-1}} \| f_{\eta,0} \|_{U^k} \, \| f_{\eta,1} \|_{U^k} = \prod_{\omega \in \{0,1\}^k} \| f_\omega \|_{U^k} .$$

This completes the induction. □

In the main text the last proposition is usually applied to indicator functions or to nonnegative functions bounded by indicators, when one may remove all conjugation signs from both sides of the inequality.

Bibliography

[1] M. Abramowitz and I. A. Stegun (eds.), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York, 1992.

[2] T. F. Bloom and O. Sisask, *An improvement to the Kelley–Meka bounds on three-term arithmetic progressions*, arXiv:2309.02353, 2023.

[3] J. Bourgain, *A Szemerédi type theorem for sets of positive density in* $\mathbf{R}^k$, Israel J. Math. **54** (1986), no. 3, 307–316.

[4] J. Bourgain, *A nonlinear version of Roth's theorem for sets of positive density in the real line*, J. Analyse Math. **50** (1988), 169–181.

[5] A. Bulj and V. Kovač, *On hyperbolic corners and unit-area triangles in planar sets of large measure*, preprint, https://arxiv.org/abs/2605.30033.

[6] M. Christ, P. Durcik, and J. Roos, *Trilinear smoothing inequalities and a variant of the triangular Hilbert transform*, Adv. Math. **390** (2021), Paper No. 107863.

[7] J. A. Clarkson, *Uniformly convex spaces*, Trans. Amer. Math. Soc. **40** (1936), no. 3, 396–414.

[8] B. Cook, Á. Magyar, and M. Pramanik, *A Roth-type theorem for dense subsets of* $\mathbb{R}^d$, Bull. London Math. Soc. **49** (2017), no. 4, 676–689.

[9] P. Durcik, *An* L^4 *estimate for a singular entangled quadrilinear form*, Math. Res. Lett. **22** (2015), no. 5, 1317–1332.

[10] P. Durcik and V. Kovač, *A Szemerédi-type theorem for subsets of the unit cube*, Anal. PDE **15** (2022), no. 2, 507–549.

[11] P. Durcik and V. Kovač, *Boxes, extended boxes and sets of positive upper density in the Euclidean space*, Math. Proc. Cambridge Philos. Soc. **171** (2021), no. 3, 481–501.

[12] P. Durcik, V. Kovač, and L. Rimanić, *On side lengths of corners in positive density subsets of the Euclidean space*, Int. Math. Res. Not. IMRN **2018**, no. 22, 6844–6869.

[13] P. Durcik, V. Kovač, and C. Thiele, *Power-type cancellation for the simplex Hilbert transform*, J. Anal. Math. **139** (2019), 67–82.

[14] P. Durcik and C. Thiele, *Singular Brascamp–Lieb inequalities with cubical structure*, Bull. London Math. Soc. **52** (2020), no. 2, 283–298.

[15] T. Eisner and T. Tao, *Large values of the Gowers–Host–Kra seminorms*, J. Anal. Math. **117** (2012), 133–186.

[16] P. Erdős, *My Scottish Book 'Problems'*, in *The Scottish Book*, Birkhäuser, Boston, 1981, pp. 27–33.

[17] P. Erdős, *Set-theoretic, measure-theoretic, combinatorial, and number-theoretic problems concerning point sets in Euclidean space*, Real Anal. Exchange **4** (1978/79), no. 2, 113–138.

[18] P. Erdős, *Some combinatorial, geometric and set theoretic problems in measure theory*, in *Measure Theory, Oberwolfach 1983*, Lecture Notes in Math. **1089**, Springer, 1984, pp. 321–327.

[19] K. J. Falconer, V. Kovač, and A. Yavicoli, *The density of sets containing large similar copies of finite sets*, J. Anal. Math. **148** (2022), no. 1, 339–359.

[20] K. J. Falconer and J. M. Marstrand, *Plane sets with positive density at infinity contain all large distances*, Bull. London Math. Soc. **18** (1986), no. 5, 471–474.

[21] E. Følner, *On groups with full Banach mean value*, Math. Scand. **3** (1955), 243–254.

[22] H. Furstenberg, Y. Katznelson, and B. Weiss, *Ergodic theory and configurations in sets of positive density*, in *Mathematics of Ramsey Theory*, Algorithms Combin. **5**, Springer, Berlin, 1990, pp. 184–198.

[23] W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), no. 3, 465–588.

[24] R. L. Graham, *On partitions of* $\mathbb{E}^n$, J. Combin. Theory Ser. A **28** (1980), no. 1, 89–97.

[25] R. L. Graham, *Recent trends in Euclidean Ramsey theory*, Discrete Math. **136** (1994), no. 1–3, 119–127.

[26] B. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. (2) **167** (2008), no. 2, 481–547.

[27] B. Green and T. Tao, *New bounds for Szemerédi's theorem, III: a polylogarithmic bound for* $r_4(N)$, Mathematika **63** (2017), no. 3, 944–1040.

[28] C. S. Herz, *Fourier transforms related to convex sets*, Ann. of Math. (2) **75** (1962), 81–92.

[29] L. Huckaba, N. Lyall, and Á. Magyar, *Simplices and sets of positive upper density in* $\mathbb{R}^d$, Proc. Amer. Math. Soc. **145** (2017), no. 6, 2335–2347.

[30] Z. Kelley and R. Meka, *Strong bounds for 3-progressions*, in *2023 IEEE 64th Annual Symposium on Foundations of Computer Science—FOCS 2023*, IEEE Computer Society, 2023, pp. 933–973.

[31] V. Kovač, *Density theorems for anisotropic point configurations*, Canad. J. Math. **74** (2022), no. 5, 1244–1284.

[32] V. Kovač, *Coloring and density theorems for configurations of a given volume*, Proc. Lond. Math. Soc. (3) **132** (2026), no. 3, Paper No. e70143.

[33] V. Kovač and B. Predojević, *Large dilates of hypercube graphs in the plane*, Anal. Math. **50** (2024), no. 3, 893–915.

[34] V. Kovač and A. A. Santos Sepčić, *Near-optimal density theorems for large dilates of large point configurations*, preprint, https://arxiv.org/abs/2604.18544.

[35] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, Wiley-Interscience, New York, 1974.

[36] M. Lacey and C. Thiele, *L^p estimates on the bilinear Hilbert transform for* $2 < p < \infty$, Ann. of Math. (2) **146** (1997), no. 3, 693–724.

[37] M. Lacey and C. Thiele, *On Calderón's conjecture*, Ann. of Math. (2) **149** (1999), no. 2, 475–496.

[38] J. Leng, A. Sah, and M. Sawhney, *Improved bounds for Szemerédi's theorem*, arXiv:2402.17995, 2024.

[39] W. Littman, *Fourier transforms of surface-carried measures and differentiability of surface averages*, Bull. Amer. Math. Soc. **69** (1963), 766–770.

[40] N. Lyall and Á. Magyar, *Product of simplices and sets of positive upper density in* $\mathbb{R}^d$, Math. Proc. Cambridge Philos. Soc. **165** (2018), no. 1, 25–51.

[41] N. Lyall and Á. Magyar, *Distance graphs and sets of positive upper density in* $\mathbb{R}^d$, Anal. PDE **13** (2020), no. 3, 685–700.

[42] N. Lyall and Á. Magyar, *Weak hypergraph regularity and applications to geometric Ramsey theory*, Trans. Amer. Math. Soc. Ser. B **9** (2022), 160–207.

[43] R. D. Mauldin, *Some problems and ideas of Erdős in analysis and geometry*, in *Erdős Centennial*, Bolyai Soc. Math. Stud. **25**, János Bolyai Math. Soc., Budapest, 2013, pp. 365–376.

[44] R. D. Mauldin, *Some problems in set theory, analysis and geometry*, in *Paul Erdős and his mathematics, I (Budapest, 1999)*, Bolyai Soc. Math. Stud. **11**, János Bolyai Math. Soc., Budapest, 2002, pp. 493–506.

[45] R. Raghavan, *Improved bounds for 3-progressions*, arXiv:2603.27045, 2026.

[46] I. D. Shkredov, *On a problem of Gowers*, Izv. Math. **70** (2006), no. 2, 385–425.

[47] C. D. Sogge, *Fourier Integrals in Classical Analysis*, Cambridge Tracts in Mathematics **105**, Cambridge University Press, Cambridge, 1993.

[48] E. M. Stein, *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals*, Princeton Mathematical Series **43**, Princeton University Press, Princeton, 1993.

[49] L. A. Székely, *Remarks on the chromatic number of geometric graphs*, in *Graphs and Other Combinatorial Topics*, Teubner-Texte Math. **59**, Teubner, 1983, pp. 312–315.

[50] E. Szemerédi, *On sets of integers containing no* k *elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245.

[51] T. Tao, *The ergodic and combinatorial approaches to Szemerédi's theorem*, in *Additive Combinatorics*, CRM Proc. Lecture Notes **43**, Amer. Math. Soc., Providence, 2007, pp. 145–193.

[52] T. Tao, *Cancellation for the multilinear Hilbert transform*, Collect. Math. **67** (2016), no. 2, 191–206.

[53] P. Varnavides, *On certain sets of positive density*, J. London Math. Soc. **34** (1959), 358–360.

[54] R. C. Vaughan, *The Hardy–Littlewood Method*, second edition, Cambridge University Press, 1997.

[55] P. Zorin-Kranich, *Cancellation for the simplex Hilbert transform*, Math. Res. Lett. **24** (2017), no. 2, 581–592.