Finite Configurations in Sparse Sets

by

Vincent Chan

B. Mathematics, University of Waterloo, 2009

M. Mathematics, University of Waterloo, 2010

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Mathematics)

The University Of British Columbia

(Vancouver)

April 2014

Abstract

We prove a result which adds to the study of continuous analogues of Szemerédi-type problems. Let $E \subseteq \mathbb{R}^n$ be a Lebesgue-null set of Hausdorff dimension α, k, m be integers satisfying a suitable relationship, and $\{B_1, \ldots, B_k\}$ be $n \times (m-n)$ matrices. We prove that if the set of matrices B_i are non-degenerate in a particular sense, α is sufficiently close to n, and if E supports a probability measure satisfying certain dimensionality and Fourier decay conditions, then E contains a k-point configuration of the form $\{x + B_1 y, \ldots, x + B_k y\}$. In particular, geometric configurations such as collinear triples, triangles, and parallelograms are contained in sets satisfying the above conditions.

Preface

Part of Chapter 1 and Chapters 2 through 5 of this thesis were produced from a manuscript [7] accepted for publication in Journal d'Analyse Mathematique:

> V. Chan, I. Łaba, and M. Pramanik. *Finite configurations in sparse sets*, March 2014.

The identification of the research program was made by Izabella Łaba and Malabika Pramanik. Research activities were conducted equally by all three authors. Writing of the manuscript was conducted largely by Vincent Chan, with feedback and additions given by Izabella Łaba and Malabika Pramanik.

Table of Contents

List of Tables

List of Symbols

$\mathbb{Z}$	Integers
$\mathbb{N}$	Strictly positive integers
$\mathbb{Q}$	Rational numbers
$\mathbb{R}$	Real numbers
$\mathbb{C}$	Complex numbers
$\mathcal{P}$	Prime numbers
$\mathbb{S}^1$	Unit circle in $\mathbb{R}^2$
E^k	k-fold Cartesian product of E
$\#E$	Number of elements in set E
$\|E\|$	Lebesgue measure of set E
$\mathcal{M}(E)$	Set of probability measures supported on E
$C_c^\infty(\mathbb{R}^n)$	Infinitely differentiable functions on $\mathbb{R}^n$ with compact support
$\mathcal{S}(\mathbb{R}^n)$	Schwartz functions on $\mathbb{R}^n$
$\widehat{f}$	Fourier transform of the function f
$\widehat{\mu}$	Fourier transform of the measure μ
A^t	Transpose of matrix A
$I_{n\times n}$	$n \times n$ identity matrix
$0_{a\times b}$	$a \times b$ zero matrix
$\mathbb{1}_E(x)$	Indicator function on E
$f^{-1}(E)$	Preimage of E under f
$\mathbb{E}(b\|\mathcal{B})$	Conditional expectation of b with respect to sigma-algebra $\mathcal{B}$
$f \lesssim g$	$f \leq Cg$ for a constant C
$f \gtrsim g$	$f \geq Cg$ for a constant C
∇	Gradient

$\|f\|_{L^p_x}$	L^p norm of f in the x variable
$\|f\|_{L^p(E)}$	L^p norm of f on space E
$\|f\|_p$	L^p norm of f
$\|\|\|x\|\|\|$	Distance from x to closest integer
$\lceil x \rceil$	Smallest integer larger than x
$\lfloor x \rfloor$	Largest integer smaller than x
$\mathcal{N}(A)$	Null space of matrix A
$E_{[\delta]}$	δ-thickened E
$\mathbb{Q}_\varepsilon$	Squares of side-length ε with corners in $\varepsilon\mathbb{Z}^2$
Sym_k	Symmetric group on k elements
id	Identity permutation
$\varnothing$	Empty set

Acknowledgments

I would like to thank my advisors, Izabella Łaba and Malabika Pramanik, for their constant support and assistance throughout the past four years. Their knowledge, insight, and advice has made this work possible. I would also like to thank Akos Magyar for being a part of my supervisory committee.

I am thankful to the National Sciences and Engineering Research Council and to the University of British Columbia for their financial support and employment opportunities.

The UBC mathematics department, in particular the staff members, deserves credit for making my experience a pleasant one, and for all the support they have given me in my activities.

My life has been enriched by the friends by my side, both old and new, and for that they have my heartfelt gratitude. I am deeply grateful to my family for their affection and support. I am especially indebted to my parents, without whom I would not be.

Dedication

To my wife, Vicki, for her support, sacrifices, encouragement, and love.

Chapter 1

Introduction

1.1 History

Erdős [9] once made the following conjecture: for any infinite set of real numbers F, there exists a set E of positive measure which does not contain any non-trivial affine copies of F. This is an example of a larger class of problems, measuring a set from a geometrical point of view: one would like to determine conditions on a set E which will guarantee the existence of certain configurations in E, or in the negative direction, will ensure the absence of these configurations. The precise meaning of this will vary, as will the ambient space for E. For instance, we may have another fixed set F in consideration, and ask if there is a geometrically similar copy of F contained in E; or we may ask if E contains various geometric shapes or sets of points which satisfy some set of linear equations. The solution to Erdős' problem is still unknown in general, however there are many similar problems that have been solved.

In the special case when $F \subseteq \mathbb{R}$ consists of a decreasing sequence x_n converging to 0 such that $\liminf x_{i+1}/x_i = 1$ (that is, a slowly decaying sequence), Falconer [12] shows there exists a closed set $E \subseteq \mathbb{R}$ of positive measure which does not contain any affine copies of F, using a Cantor-like construction. In fact, his construction is much stronger; his set E misses all but finitely many entries of any affine copy of F. It is worth noting that if the sequence decays geometrically fast, such as $\{2^{-i}\}$, the result is unknown.

Bourgain [6] proves that for any infinite set $S \subseteq \mathbb{R}$, there exists a set $E \subseteq \mathbb{R}^3$ of positive measure which does not contain any affine copies of $F = S \times S \times S \subseteq \mathbb{R}^3$, using a probabilistic method. His proof of this result also yields that for any infinite sets $S_1, S_2, S_3 \subseteq \mathbb{R}$, there exists a set $E \subseteq \mathbb{R}^3$ of positive measure which does not contain any affine copies of $F = S_1 + S_2 + S_3$, where $+$ refers to the usual set addition.

For more results concerning the case when F is a sequence, see Arias de Reyna [1] and Miller [34], where they construct sets E of second Baire category; Borwein and Ditor [5], which examines the related problem when 'affine' is replaced with 'translated'; Komjáth [29], a generalization of Borwein and Ditor's result; Bourgain [6] , which considers the case when F is certain cross products or sum sets; and Kolountzakis [28], which gives a negative result concerning almost all affine copies (in Lebesgue measure sense) rather than all copies.

We will focus on the case when the configuration is finite. Although the continuous case is of primary interest, the discrete case offers some rich history and motivation for the analogous results.

1.1.1 Discrete case

In the case when the ambient space is the positive integers $\mathbb{N}$, a simple configuration to consider is an arithmetic progression $\{x, x+y, \ldots, x+(k-1)y\}$ where $y \neq 0$, called a *k-term arithmetic progression.* In 1936, Erdős and Turán [11] made the following conjecture related to sets containing arithmetic progressions.

Conjecture 1.1. *Let $k \geq 3$ be an integer and $0 < \delta \leq 1$. If N is sufficiently large depending on k and δ, then every subset A of $\{1, 2, \ldots, N\}$ of cardinality at least δN contains a k-term arithmetic progression.*

This notion of the set A containing a proportion of $\{1, 2, \ldots, N\}$ leads us to consider the density of a set relative to the natural numbers. To be precise, define the *upper asymptotic density* of a set $A \subseteq \mathbb{N}$ to be the quantity

$$\overline{\delta^*}(A) = \limsup_{N \to \infty} \frac{\#(A \cap \{1, \dots, N\})}{N},$$

where $\#S$ denotes the number of elements of $S \subseteq \mathbb{N}$. A seemingly weaker but equivalent statement to Conjecture 1.1 is as follows:

Conjecture 1.2. *Let $E \subseteq \mathbb{N}$ be an arbitrary set with*

$$\overline{\delta^*}(E) > 0. \tag{1.1}$$

Then E contains a k-term arithmetic progression for any integer $k \geq 3$.

The simplest case of $k = 3$ was proved by Roth [35] in 1953. Later, Szemerédi [38] proved the case $k = 4$, and then completed the conjecture by proving the result for any $k \geq 4$ in 1975 [39] in a combinatorial manner. Furstenberg [14] applied ergodic theory to provide a second proof of the case $k \geq 4$, and Gowers [16] used harmonic analysis and introduced higher order Fourier analysis to prove this case. Furstenberg and Katznelson [15] proved a multidimensional variant of Szemerédi's Theorem.

A set can fail to have 3-term arithmetic progressions even if (1.1) fails just barely; Salem and Spencer [37] showed that for large N, there exists a subset $A \subseteq \{1, 2, \dots, N\}$ of cardinality about $N^{1-c/\log\log N}$ which contains no 3-term arithmetic progressions, and Behrend [2] improved on this with a power of $1 - c/\sqrt{N}$ in place of $1 - c/\log\log N$ on N. However, there are cases when (1.1) fails for the set E, yet E still contains k-term arithmetic progressions, so long as E is sufficiently random in some sense. The prime numbers $\mathcal{P}$ have an upper asymptotic density of 0, yet Roth/Szemerédi-type results still apply. Define the *relative upper asymptotic density* of a set $A \subseteq \mathbb{N}$ *relative to* a set $B \subseteq \mathbb{N}$ with $A \subseteq B$ to be the quantity

$$\overline{\delta^*_B}(A) = \limsup_{N \to \infty} \frac{\#(A \cap \{1, \dots, N\})}{\#(B \cap \{1, \dots, N\})}.$$

The analogous result is:

Theorem 1.3. *Let $E \subseteq \mathcal{P}$ be an arbitrary set with*

$$\overline{\delta^*_{\mathcal{P}}}(E) > 0. \tag{1.2}$$

Then E contains k-term arithmetic progression for any integer $k \geq 3$.

Green [18] proved the case $k = 3$, then with Tao [19] proved the general case $k \geq 3$. A multidimensional variant of this result was proven by Tao and Ziegler [41] using a weighted version of the Furstenberg correspondence principle, and also by Cook, Magyar, and Titichetrakun [8] based on a hypergraph approach.

In a similar vein, certain random sets probabilistically have 3-term arithmetic progressions even with 0 upper asymptotic density, as in the following theorem proved by Kohayakawa, Łuczak, and Rödl [27].

Theorem 1.4. *For integers $1 \leq M \leq n$, let $\mathcal{R}(n, M)$ be the probability space consisting of M-element subsets of $\{1, \ldots, n\}$ equipped with the uniform measure. Let $0 < \delta \leq 1$. There exists a constant $C = C(\delta)$ such that if $C\sqrt{n} \leq M \leq n$, then with probability tending to 1 as $n \to \infty$, the random set $R \in \mathcal{R}(n, M)$ has the property that any subset of R of cardinality at least $\delta(\#R)$ contains k-term arithmetic progression.*

We shall see that these results provide motivation for some of the results in the continuous case.

1.1.2 Continuous case

A first step is to see how the measure theoretic size of $E \subseteq \mathbb{R}^n$ determines if it contains particular configurations; we will use $|E|$ to denote the Lebesgue measure of E.

Definition 1.5. *(a) We say $\psi : \mathbb{R}^n \to \mathbb{R}^n$ is a* similarity map *if there exists $c > 0$ such that $|\psi(x) - \psi(y)| = c|x - y|$ for every $x, y \in \mathbb{R}^n$.*

(b) We say $E \subseteq \mathbb{R}^n$ contains a similar copy of *$F \subseteq \mathbb{R}^n$ if there exists a similarity map ψ such that $\psi(F) \subseteq E$.*

It is a consequence of the Lebesgue density theorem that a set of positive measure contains a similar copy of every finite set. We may ask if the converse is true, that is, if a set E contains a copy of every finite set, must we have $|E| > 0$? Erdős and Kakutani [10] showed the answer to this is negative, by constructing a compact (and perfect), measure 0 set which contains a similar copy of every finite set. Certainly not every set of measure 0 has this property, but what condition on E would ensure various configurations? To answer this, we require some way of distinguishing between sets of measure 0. In this case, we can explore how the dimensional size of a set will affect its capacity to contain or avoid configurations. There are several ways of defining dimension, of particular use is Hausdorff dimension.

Definition 1.6. *If $U \subseteq \mathbb{R}^n$ is a non-empty set, we define its* diameter *to be*

$$\operatorname{diam}(U_i) = \sup\{|x-y| : x, y \in U\}.$$

*We say that the countable collection $\{U_i\}$ is a δ-*cover *for $E \subseteq \mathbb{R}^n$ if $E \subseteq \bigcup_i U_i$ and $\operatorname{diam}(U_i) \leq \delta$.*

Suppose $E \subseteq \mathbb{R}^n$ and $s > 0$. For $\delta > 0$ we define

$$\mathcal{H}^s_\delta(E) = \inf\left\{\sum_{i=1}^{\infty} \operatorname{diam}(U_i)^s : \{U_i\} \text{ is a } \delta\text{-cover of } E\right\}.$$

We define the s-dimensional Hausdorff measure *of E to be*

$$\mathcal{H}^s(E) = \lim_{\delta \to 0} \mathcal{H}^s_\delta(E);$$

clearly the limit exists. Finally, we define the Hausdorff dimension *of E to be*

$$\dim_H(E) = \inf\{s : \mathcal{H}^s(E) = 0\} = \sup\{s : \mathcal{H}^s(E) = \infty\}.$$

Suppose F consists of three points, a triangle. Since any similar copy of a triangle preserves its angles, one question that arises is how large Hausdorff dimension can a set in $\mathbb{R}^n$ $(n \geq 2)$ have if it avoids three points forming a particular angle θ. For ease, we say $E \subseteq \mathbb{R}^n$ *contains the angle θ* if there exist distinct points $x, y, z \in E$ such that the angle between the vectors $y-x$ and $z-x$ is θ, and write $\angle\theta \in E$. Define

$$C(n, \theta) = \sup\{s : \exists E \subseteq \mathbb{R}^n \text{ compact with } \dim_H(E) = s, \angle\theta \notin E\}.$$

Harangi, Keleti, Kiss, Maga, Máthé, Mattila, and Strenner [20] give upper bounds on $C(n, \theta)$ (which they show is tight for $\theta = 0, \pi$), and Máthé [32] provides lower bounds. Their results are summarized below.

θ	lower bound on $C(n, \theta)$	upper bound on $C(n, \theta)$
$0, \pi$	$n-1$	$n-1$
$\pi/2$	$n/2$	$\lfloor (n+1)/2 \rfloor$
$\cos^2\theta \in \mathbb{Q}$	$n/4$	$n-1$
other θ	$n/8$	$n-1$

Table 1.1: Summary of results on bounds for $C(n, \theta)$

Also explored by Harangi et al [20] is the question of how large Hausdorff dimension a set can have if it avoids not just the angle α, but a δ-neighborhood of α.

If we instead fix the configuration, a simple case to consider is, like in the discrete case, a 3-term arithmetic progression. The following theorem gives a stronger result:

Theorem 1.7 (Keleti [25]). *There exists a compact set in $\mathbb{R}$ with Hausdorff dimension 1 which does not contain any "one-dimensional parallelogram" $\{x, x+y, x+z, x+y+z\}$, with $y, z \neq 0$.*

Keleti's proof involves a Cantor-like construction: Start with the interval $[0, 1]$ and on the mth step, each interval remaining is split into $6m$ intervals of length $1/(6^{m-1}m!)$, and every sixth interval is selected, but a shift is introduced on some of the intervals in order to avoid parallelograms. It can be seen to be of full dimension via the mass distribution principle (cf. [13] 4.6). One consequence of Keleti's result is that full dimension of a set E in $\mathbb{R}$ does not guarantee existence of 3-term arithemetic progressions in E, seen by taking $y = z$.

As in the discrete case, where even with 0 upper density certain types of sets can be guaranteed to contain 3-term arithmetic progressions, there are cases where measure 0 sets must contain a 3-term arithmetic progression. Łaba and Pramanik [30] show that essentially with sufficient Fourier dimension, E does contain 3-term arithemetic progressions. To make this precise:

Theorem 1.8 (Łaba, Pramanik [30])**.** *Suppose $E \subseteq [0,1]$ is a closed set which supports a probability measure μ with the following properties:*

(i) $\mu([x, x+\varepsilon]) \leq C_1\varepsilon^\alpha$ *for all* $0 < \varepsilon \leq 1$,

(ii) $|\widehat{\mu}(\xi)| \leq C_2|\xi|^{-\beta/2}$ *for all* $\xi \neq 0$,

$2/3 < \beta \leq 1$. *If α is close enough to 1 (depending on C_1, C_2, and β), then E contains a 3-term arithmetic progression.*

In the prequel, $\widehat{\mu}(\xi) = \int e^{-2\pi i x\cdot\xi} d\mu(x)$. The first condition is related to Hausdorff dimension. Indeed, Frostman's Lemma (see [33]) says that if $E \subseteq \mathbb{R}^n$ is compact, then

$$\dim_H(E) = \sup\{\alpha \geq 0 : \exists \mu \in \mathcal{M}(E) \text{ with } \mu(B(x,\varepsilon)) \leq C\varepsilon^\alpha\},$$

where $\mathcal{M}(E)$ is the set of probability measures supported on E. Then condition (i) implies that E has Hausdorff dimension of at least α. The second condition is related to the Fourier dimension, defined as

$$\begin{aligned}\dim_F(E) = \sup\{\beta \in [0,n] :\ &\exists \mu \in \mathcal{M}(E) \\ &\text{with } |\widehat{\mu}(\xi)| \leq C(1+|\xi|)^{-\beta/2}\ \forall \xi \in \mathbb{R}^n\}.\end{aligned} \tag{1.3}$$

This means (ii) implies that E has Fourier dimension of at least β. It is thus useful to discuss the relationship between the Hausdorff dimension and Fourier dimension.

It can be shown (see [33]) that

$$\dim_H(E) = \sup\left\{s \geq 0 : \exists \mu \in \mathcal{M}(E) \text{ with } \int |x|^{s-n}|\widehat{\mu}(x)|^2\, dx < \infty\right\}.$$

Notice $\int |x|^{s-n}|\widehat{\mu}(x)|^2\, dx < \infty$ implies $|\widehat{\mu}(x)| < |x|^{-s/2}$ for "most" x with large norm, but this does not necessarily hold for all x. Comparing with (1.3), we see that $\dim_F(E) \leq \dim_H(E)$ for all $E \subseteq \mathbb{R}^n$. Strict inequality is possible, as the following example illustrates.

Example 1.9. It is well-known that the middle thirds Cantor set has Hausdorff dimension of $\log 2/\log 3$. Let μ be the standard Cantor measure, the restriction of $\mathcal{H}^s$ to the Cantor set where $s = \log 2/\log 3$. Then the Fourier transform is given by

$$\widehat{\mu}(x) = \prod_{j=1}^{\infty} \cos(3^{-j}x),$$

and so $\widehat{\mu}(3^\ell \pi) \not\to 0$ as $\ell \to \infty$. In general, it is also true that the Cantor set supports no non-zero Radon measure whose Fourier transform would tend to 0 at infinity [23], so has Fourier dimension 0.

When equality occurs for a set E, E is said to be a *Salem set*; the spheres S^{n-1} in $\mathbb{R}^n$ are simple examples of Salem sets with dimension $n-1$. Deterministic Salem sets of non-integral dimension are uncommon, some constructions are due to Kahane [21] and Kaufman [24]. Salem [36] constructed random Salem sets, and Kahane [22] has shown that Salem sets are abound as random sets. In particular, if $E \subseteq \mathbb{R}^n$ is compact with $\dim_H(E) = \alpha < n/2$, then the image of E under n-dimensional brownian motion is almost surely a Salem set of dimension 2α. Further probabbilistic constructions of Salem sets can be found in [3], [4], and [30].

In this sense, condition (ii) above relates to a "randomness" condition, which is an analogue of Theorems 1.3 and 1.4. It should be mentioned that although Salem sets are related to the two conditions, the concepts are not equivalent—conditions (i) and (ii) do not necessitate equality in Hausdorff and Fourier dimension, although we need to have some control over the constants C_1, C_2. Łaba and Pramanik modify Salem's construction to produce explicit constants in the estimates. Keleti improved upon the result that full dimension does not guarantee 3 term arithmetic progressions in the following.

Theorem 1.10 (Keleti [26])**.** *For a given distinct triple of points* $\{x, y, z\}$*, there exists a compact set in* $\mathbb{R}$ *with Hausdorff dimension 1 which does not contain any similar copy of* $\{x, y, z\}$*.*

The method of proof is similar to the ideas from [25]. Another way to extend Theorem 1.7 is in the ambient dimension:

Theorem 1.11 (Maga [31])**.** *There exists a compact set in* $\mathbb{R}^n$ *with Hausdorff dimension* n *which does not contain any parallelogram* $\{x, x+y, x+z, x+y+z\}$*, with* $y, z \neq 0$*.*

This construction is the exact analogue of the construction from Theorem 1.7, using cubes instead of intervals. In the same paper, Maga also extends Theorem 1.10.

Theorem 1.12 (Maga [31])**.** *For distinct points* $x, y, z \in \mathbb{R}^2$*, there exists a compact set in* $\mathbb{R}^2$ *with Hausdorff dimension 2 which does not contain any similar copy of* $\{x, y, z\}$*.*

Maga [31] posed the following question:

Question 1.13. *If* $E \subseteq \mathbb{R}^2$ *is compact with* $\dim_H(E) = 2$*, must* E *contain the vertices of an isosceles triangle?*

The previous results and question will motivate our present result, which requires some technical definitions to set up.

1.2 Notation and definitions

Throughout this document, norms of functions are used which may require special attention to details. We shall use $\|f\|_{L^p_x}$ to denote the L^p norm of f in the x variable, in case f depends on multiple variables, $\|f\|_{L^p(E)}$ to denote the L^p norm of f on the space E, in case specifying the space is necessary, and $\|f\|_p$ for the L^p norm when the context is clear.

Many estimates are made via inequalities where we are unconcerned about constants as a multiplier. We thus use $f \lesssim g$ to mean $f \leq Cg$ where C is a constant, and $f \gtrsim g$ to mean $f \geq Cg$ where C is a constant.

We now develop some notation specific to the main result.

Definition 1.14. *Fix integers $n \geq 2$, $k \geq 3$ and $m \geq n$. Suppose $B_1, \dots, B_k$ are $n \times (m-n)$ matrices.*

(a) We say E contains a k-point $\mathbb{B}$-configuration *if there exists $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^{m-n} \setminus \{0\}$ such that $\{x + B_j y\}_{j=1}^k \subseteq E$.*

(b) Given any finite collection of subspaces $V_1, \dots, V_q \subseteq \mathbb{R}^{m-n}$ such that $\dim(V_i) < m-n$, we say that E contains a non-trivial k-point $\mathbb{B}$-configuration with respect to $(V_1, \dots, V_q)$ *if there exists $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^{m-n} \setminus \bigcup_{i=1}^q V_i$ such that $\{x + B_j y\}_{j=1}^k \subseteq E$.*

For both of these definitions, we will drop the k from the notation if there is no confusion.

Let $A_1, \dots, A_k$ be $n \times m$ matrices. For any set of distinct indices $J = \{j_1, \dots, j_s\} \subseteq \{1, \dots, k\}$, define the $ns \times m$ matrix $\mathbb{A}_J$ by

$$\mathbb{A}_J^t = (A_{j_1}^t \ \cdots \ A_{j_s}^t),$$

where A^t denotes the transpose of A. We shall use $\mathbb{A} = \mathbb{A}_{\{1,\dots,k\}}$.

Let r be the unique positive integer such that

$$n(r-1) < nk - m \leq nr. \tag{1.4}$$

If we have $nk - m$ components and account for $r-1$ groups of size n, we are left with $nk - m - n(r-1)$. This quantity is useful in the main theorem, and bulky to repeatedly use. We shall denote

$$n' = nk - m - n(r-1). \tag{1.5}$$

Notice if $nk - m$ is a multiple of n, then $n' = n$, and in general, $0 < n' \leq n$.

Definition 1.15. *We say that $\{A_1, \dots, A_k\}$ is* non-degenerate *if for any $J \subseteq \{1, \dots, k\}$ with $\#(J) = k - r$ and any $j \in \{1, \dots, k\} \setminus J$, the $m \times m$ matrix*

$$(\mathbb{A}_J^t \ \widetilde{A_j}^t)$$

is non-singular for any choice of $(n-n') \times m$ submatrix $\widetilde{A_j}$ of A_j.

1.3 The main result

We use $\lceil x \rceil$ to denote the smallest integer larger than x.

Theorem 1.16. *Suppose*

$$n\left\lceil \frac{k+1}{2} \right\rceil \leq m < nk \tag{1.6}$$

and

$$\frac{2(nk-m)}{k} < \beta < n. \tag{1.7}$$

Let $\{B_1, \ldots, B_k\}$ *be a collection of* $n \times (m-n)$ *matrices such that* $A_j = (I_{n\times n}\ B_j)$ *is non-degenerate in the sense of Definition 1.15, where* $I_{n\times n}$ *is the* $n \times n$ *identity matrix. Then for any constant* C*, there exists a positive number* $\epsilon_0 = \epsilon_0(C, n, k, m, \mathbb{B}) \ll 1$ *with the following property. Suppose the set* $E \subseteq \mathbb{R}^n$ *with* $|E| = 0$ *supports a positive, finite, Radon measure* μ *with the two conditions:*

(a) (ball condition) $\sup_{\substack{x\in E\\ 0<r<1}} \frac{\mu(B(x;r))}{r^\alpha} \leq C$,

(b) (Fourier decay) $\sup_{\xi\in\mathbb{R}^n} |\widehat{\mu}(\xi)|(1+|\xi|)^{\beta/2} \leq C$.

If $n - \varepsilon_0 < \alpha < n$*, then:*

(i) E *contains a* k*-point* $\mathbb{B}$*-configuration in the sense of Definition 1.14 (a).*

(ii) Moreover, for any finite collection of subspaces $V_1, \ldots, V_q \subseteq \mathbb{R}^{m-n}$ *with* $\dim(V_i) < m-n$, E *contains a non-trivial* k*-point* $\mathbb{B}$*-configuration with respect to* $(V_1, \ldots, V_q)$ *in the sense of Definition 1.14 (b).*

This can be viewed as a multidimensional analogue of [30]. The existence and constructions of measures on $\mathbb{R}$ that satisfy $(a), (b)$ are discussed in detail in [30]. In higher dimensions, it should be possible to generalize the construction in [30, Section 6] to produce examples in $\mathbb{R}^n$; alternatively, it is easy to check that if $\mu = \tilde{\mu}(dr) \times \sigma(d\omega)$ is a product measure in radial coordinates (r, ω), where $\tilde{\mu}$ is a Salem measure on $[0,1]$ as in [30, Section 6] and σ is the Lebesgue measure on S^{n-1}, then μ satisfies the conditions $(a), (b)$ of Theorem 1.16.

1.4 Examples

Motivated by the discussion in Section 1.1.2, we provide some examples of geometric configurations which are covered by Theorem 1.16; proofs and further discussion will be contained in Chapter 5.

Corollary 1.17. *Let a, b, c be three distinct collinear points in $\mathbb{R}^n$. Suppose that $E \subset \mathbb{R}^n$ satisfies the assumptions of Theorem 1.16 with ε_0 sufficiently small depending on a, b, c, C. Then E must contain three distinct points x, y, z that form a similar image of the triple a, b, c.*

This result includes 3-term arithmetic progressions, so can be viewed as an extension of Theorem 1.8 to higher dimensions. It should be noted that their proof would also work for any fixed 3-point configuration, the 1-dimensional version of Corollary 1.17. For nonlinear triples, Theorem 1.16 does not seem to cover general triangles in dimensions $n \geq 3$, but there is a positive result for the plane.

Corollary 1.18. *Let a, b, c be three distinct points in $\mathbb{R}^2$. Suppose that $E \subset \mathbb{R}^2$ satisfies the assumptions of Theorem 1.16 with ε_0 sufficiently small depending on a, b, c, C. Then E must contain three distinct points x, y, z such that the triangle $\triangle xyz$ is a similar copy of the triangle $\triangle abc$.*

Compare this to the results of Harangi, Keleti, Kiss, Maga, Máthé, Mattila, and Strenner as summarized in Table 1.1. Corollary 1.18 shows that if θ is given, then any $E \subseteq \mathbb{R}^2$ as in Theorem 1.16 must not only contain the angle θ, but in fact θ can be realized as the angle of an isosceles triangle with vertices in E (or any other pre-determined shape); this answers, in part, Question 1.13. Note that Maga's result described in Theorem 1.12 shows the result is false without the assumption of Fourier decay.

For larger configurations, we first consider parallelograms.

Corollary 1.19. *Suppose that $E \subset \mathbb{R}^n$ satisfies the assumptions of Theorem 1.16, with ε_0 sufficiently small depending on C. Then E contains a parallelogram $\{x, x+y, x+z, x+y+z\}$, where the four points are all distinct and $x, y, z \in \mathbb{R}^n$.*

Recall Theorem 1.11 by Maga, which shows the result is false without the assumption of Fourier decay.

We end with a polynomial example.

Corollary 1.20. *Let $a_1, \ldots, a_6$ be distinct numbers, all greater than 1. Suppose that $E \subset \mathbb{R}^3$ satisfies the assumptions of Theorem 1.16, with ε_0 small enough depending on C and a_i. Then E contains a configuration of the form*

$$x, \quad x + B_2 y, \quad x + B_3 y, \quad x + B_4 y, \tag{1.8}$$

for some $x \in \mathbb{R}^3$ and $y \in \mathbb{R}^6$ with $B_i y \neq 0$ for $i = 2, 3, 4$, where

$$B_2 = \begin{pmatrix} 1 & \ldots & 1 \\ a_1 & \ldots & a_6 \\ a_1^2 & \ldots & a_6^2 \end{pmatrix}, \quad B_3 = \begin{pmatrix} a_1^3 & \ldots & a_6^3 \\ a_1^4 & \ldots & a_6^4 \\ a_1^5 & \ldots & a_6^5 \end{pmatrix}, \quad B_4 = \begin{pmatrix} a_1^6 & \ldots & a_6^6 \\ a_1^7 & \ldots & a_6^7 \\ a_1^8 & \ldots & a_6^8 \end{pmatrix}.$$

This is a non-trivial result in the following sense. Since Vandermonde matrices are non-singular, the set of 6 vectors

$$\begin{pmatrix} 1 \\ a_1 \\ \vdots \\ a_1^5 \end{pmatrix}, \ \ldots, \ \begin{pmatrix} 1 \\ a_6 \\ \vdots \\ a_6^5 \end{pmatrix}$$

forms a basis for $\mathbb{R}^6$. It follows that if $a, b, c \in E$, there is a unique $y \in \mathbb{R}^6$ such that $b = a + B_2 y$ and $c = a + B_3 y$. This also determines uniquely the point $a + B_4 y$, which might or might not be in E. Our result asserts that, under the conditions of Corollary 1.20, we may choose x and y so that in fact all 4 points in (1.8) lie in E.

Chapter 2

A multilinear form for counting configurations

We will consider the following multilinear form, initially defined for $f_j \in C_c^\infty(\mathbb{R}^n)$, the smooth functions on $\mathbb{R}^n$ with compact support:

$$\Lambda(f_1, \ldots, f_k) = \int_{\mathbb{R}^m} \prod_{j=1}^{k} f_j(A_j\vec{x})\ d\vec{x}. \tag{2.1}$$

If $f_j = f$ for all j, we write $\Lambda(f)$ instead of $\Lambda(f, \ldots, f)$. Clearly if $\Lambda(f) \neq 0$, then the support of f contains configurations of the form $\{A_j\vec{x} : 1 \leq j \leq k\}$ for a set of $\vec{x}$ of positive measure. We will later rewrite $\Lambda(f_1, \ldots, f_k)$ in a form that will allow us to extend it to measures, not just functions, then show that it is suitable for counting configurations in sparse sets.

2.1 Fourier-analytic representation of the multilinear form

It turns out that working with the Fourier transforms of f_j allows for good estimates on Λ. Recall the Fourier transform of a function $f \in C_c^\infty(\mathbb{R}^n)$ is defined by $\widehat{f}(\xi) = \int_{\mathbb{R}^n} e^{-2\pi i x\cdot\xi} f(x)\ dx$. The following result provides a means to make use of the Fourier transform in Λ.

Proposition 2.1. *For $f_j \in C_c^\infty(\mathbb{R}^n)$, $\Lambda(f_1, \ldots, f_k)$ defined in (2.1) admits the representation*

$$\Lambda(f_1, \ldots, f_k) = C \int_S \prod_{j=1}^{k} \widehat{f}_j(\xi_j) \; d\sigma(\xi_1, \cdots, \xi_k),$$

where σ is the Lebesgue measure on the subspace

$$S = \left\{ \xi = (\xi_1, \ldots, \xi_k) \in (\mathbb{R}^n)^k : \; \sum_{j=1}^{k} A_j^t \xi_j = \vec{0} \right\} \tag{2.2}$$

and $C = C(\mathbb{A})$ is a constant only depending on the matrices A_j.

We will require a sort of approximate identity result to prove this proposition. For a $p \times d$ $(p \leq d)$ matrix P of full rank p, define

$$V = \{\xi \in \mathbb{R}^d : \; P\xi = 0\} = \mathcal{N}(P),$$

the null space of P, and let $v = \dim(V) = d - p$.

Definition 2.2. *Fix an orthonormal basis $\{\vec{\alpha}_1, \ldots, \vec{\alpha}_v\}$ of V. The* surface measure *$d\sigma$ on V is defined as follows:*

$$\int_V F \; d\sigma = \int_{\mathbb{R}^v} F\left(\sum_{j=1}^{v} x_j \vec{\alpha}_j\right) \; dx_1 \cdots dx_v$$

for every $F \in C_c(\mathbb{R}^d)$.

Note that this definition is independent of the choice of basis. Indeed, if $\{\vec{\beta}_1, \ldots, \vec{\beta}_v\}$ is another orthonormal basis of V, then the mapping $(x_1, \ldots, x_v) \mapsto (y_1, \ldots, y_v)$ given by $\sum x_j \vec{\alpha}_j = \sum y_j \vec{\beta}_j$ is a linear isometry, hence given by an orthogonal matrix which has determinant 1. Then $d\vec{x} = d\vec{y}$ and hence

$$\int_{\mathbb{R}^v} F\left(\sum_{j=1}^{v} x_j \vec{\alpha}_j\right) \; dx_1 \cdots dx_v = \int_{\mathbb{R}^v} F\left(\sum_{j=1}^{v} y_j \vec{\beta}_j\right) \; dy_1 \cdots dy_v.$$

Lemma 2.3. *For any $g \in C_c(\mathbb{R}^d)$ and any $\Psi \in \mathcal{S}(\mathbb{R}^p)$ with $\Psi(0) \neq 0$,*

$$\lim_{\varepsilon \to 0^+} \int_{\mathbb{R}^d} g(y_1, y_2) \frac{1}{\varepsilon^p} \widehat{\Psi}\left(\frac{y_2}{\varepsilon}\right) \, dy_1 dy_2 = \Psi(0) \int_{\mathbb{R}^{d-p}} g(y_1, 0) \, dy_1.$$

Here, $y = (y_1, y_2) \in \mathbb{R}^d$, with $y_1 \in \mathbb{R}^{d-p}$ and $y_2 \in \mathbb{R}^p$.

Proof. Fix $\kappa > 0$. Our goal is to show

$$\left| \int_{\mathbb{R}^d} g(y_1, y_2) \frac{1}{\varepsilon^p} \widehat{\Psi}\left(\frac{y_2}{\varepsilon}\right) \, dy_1 dy_2 - \Psi(0) \int_{\mathbb{R}^{d-p}} g(y_1, 0) \, dy_1 \right| < \kappa$$

for all $\varepsilon > 0$ sufficiently small. Then for every $\varepsilon > 0$, we have by definition of $\Psi(0)$

$$\begin{aligned}
&\left| \int_{\mathbb{R}^d} g(y_1, y_2) \frac{1}{\varepsilon^p} \widehat{\Psi}\left(\frac{y_2}{\varepsilon}\right) \, dy_1 dy_2 - \Psi(0) \int_{\mathbb{R}^{d-p}} g(y_1, 0) \, dy_1 \right| \\
= &\left| \int_{\mathbb{R}^d} g(y_1, y_2) \frac{1}{\varepsilon^p} \widehat{\Psi}\left(\frac{y_2}{\varepsilon}\right) \, dy_1 dy_2 - \int_{\mathbb{R}^p} \int_{\mathbb{R}^{d-p}} g(y_1, 0) \frac{1}{\varepsilon^p} \widehat{\Psi}\left(\frac{y_2}{\varepsilon}\right) \, dy_1 dy_2 \right| \\
\leq &\iint |g(y_1, y_2) - g(y_1, 0)| \frac{1}{\varepsilon^p} \left| \widehat{\Psi}\left(\frac{y_2}{\varepsilon}\right) \right| \, dy_1 dy_2.
\end{aligned}$$

We now partition our region of integration into where $|y_2|$ is small and where it is large. By uniform continuity of g on its compact support K, we may choose $\eta > 0$ sufficiently small so that

$$\sup_{y_1 \in K} |g(y_1, y_2) - g(y_1, 0)| < \frac{\kappa}{2|\Psi(0)|} \tag{2.3}$$

if $|y_2| \leq \eta$. On this region,

$$\begin{aligned}
&\iint_{|y_2| \leq \eta} |g(y_1, y_2) - g(y_1, 0)| \frac{1}{\varepsilon^p} \left| \widehat{\Psi}\left(\frac{y_2}{\varepsilon}\right) \right| \, dy_1 dy_2 \\
< &\iint \frac{\kappa}{2|\Psi(0)|} \frac{1}{\varepsilon^p} \left| \widehat{\Psi}\left(\frac{y_2}{\varepsilon}\right) \right| \, dy_1 dy_2 = \frac{\kappa}{2}.
\end{aligned}$$

Since $\widehat{\Psi}$ is integrable, we can make the tail integral as small as we would like. In particular, there exists $\varepsilon > 0$ sufficiently small relative to η so that

$$\int_{|y_2|>\eta/\varepsilon} |\widehat{\Psi}(y_2)|\ dy_2 < \frac{\kappa}{4\|g\|_\infty \mathrm{diam}(K)}. \tag{2.4}$$

Then for this ε,

$$\begin{aligned} &\iint_{|y_2|>\eta, y\in K} |g(y_1,y_2)-g(y_1,0)|\frac{1}{\varepsilon^p}\left|\widehat{\Psi}\left(\frac{y_2}{\varepsilon}\right)\right|\ dy_1dy_2 \\ &\le 2\|g\|_\infty \mathrm{diam}(K)\int_{|y_2|>\eta/\varepsilon} |\widehat{\Psi}(y_2)|\ dy_2 < \frac{\kappa}{2}. \end{aligned}$$

As this inequality holds for every $\kappa > 0$, the result follows. $\square$

Proposition 2.4. *For any P as above, there exists a constant $C_P > 0$ with the property that for any $\Phi \in \mathcal{S}(\mathbb{R}^p)$ with $\Phi(0) = 1$, the limit*

$$\lim_{\varepsilon\to 0^+}\int_{\mathbb{R}^d} F(\xi)\frac{1}{\varepsilon^p}\widehat{\Phi}\left(\frac{P\xi}{\varepsilon}\right)\ d\xi$$

exists and equals $C_P\int_V F\ d\sigma$.

Proof. Fix $\Phi \in \mathcal{S}(\mathbb{R}^p)$ with $\Phi(0) = 1$. Let $\{\vec{\alpha}_1,\dots,\vec{\alpha}_v\}$ be an orthonormal basis of V. Extend this to an orthonormal basis of $\mathbb{R}^d$, say

$$\{\vec{\alpha}_1,\dots,\vec{\alpha}_v,\vec{\alpha}_{v+1},\dots,\vec{\alpha}_d\}.$$

Given any function $H \in C_c(\mathbb{R}^d)$, we define $G_H : \mathbb{R}^d \to \mathbb{R}$ as follows:

$$G_H(x_1,\dots,x_d) = H\left(\sum_{j=1}^d x_j\vec{\alpha}_j\right).$$

Notice $G_H \in C_c(\mathbb{R}^d)$ as well. Then by definition,

$$\begin{aligned} \int_V F\ d\sigma &= \int_{\mathbb{R}^v} F\left(\sum_{j=1}^v x_j\vec{\alpha}_j\right)\ dx_1\cdots dx_v \\ &= \int_{\mathbb{R}^v} G_F(x_1,\dots,x_v,0,\dots,0)\ dx_1\cdots dx_v. \end{aligned}$$

By Lemma 2.3 with $g = G_F$, $y_1 = (x_1, \dots, x_v)$, and $y_2 = (x_{v+1}, \dots, x_d)$, we have

$$\begin{aligned}
&\int_V F\ d\sigma \\
&= \frac{1}{\Psi(0)} \lim_{\varepsilon\to 0^+} \int_{\mathbb{R}^d} G_F(x_1, \dots, x_v, x_{v+1}, \dots, x_d) \frac{1}{\varepsilon^p} \widehat{\Psi}\left(\frac{x_{d+1}}{\varepsilon}, \dots, \frac{x_d}{\varepsilon}\right)\ d\vec{x},
\end{aligned}$$

for any $\Psi \in \mathcal{S}(\mathbb{R}^d)$ with $\Psi(0) \neq 0$, where $\vec{x} = (x_1, \dots, x_d)$. By the definition of G_F, this gives

$$\int_V F\ d\sigma = \frac{1}{\Psi(0)} \lim_{\varepsilon\to 0^+} \int_{\mathbb{R}^d} F(\xi) \frac{1}{\varepsilon^p} \widehat{\Psi}\left(\frac{x_{v+1}}{\varepsilon}, \dots, \frac{x_d}{\varepsilon}\right)\ d\vec{x}, \tag{2.5}$$

for any $\Psi \in \mathcal{S}(\mathbb{R}^d)$ with $\Psi(0) \neq 0$, where we denote $\xi = \sum_{j=1}^d x_j \vec{\alpha}_d$. Now, let Q be the $p \times p$ matrix defined by

$$Q \begin{pmatrix} x_{v+1} \\ \vdots \\ x_d \end{pmatrix} = \sum_{j=v+1}^d x_j P \vec{\alpha}_j.$$

Since P is of full rank and acting on basis vectors, Q is non-singular. Recall $V = \{\xi :\ P\xi = 0\}$ and $\{\vec{\alpha}_1, \dots, \vec{\alpha}_v\}$ is a basis for V, so $P\vec{\alpha}_j = 0$ for $1 \leq j \leq v$. Then

$$\begin{pmatrix} x_{v+1} \\ \vdots \\ x_d \end{pmatrix} = Q^{-1} \sum_{j=v+1}^d x_j P \vec{\alpha}_j = Q^{-1} \sum_{j=1}^d x_j P \vec{\alpha}_j = Q^{-1} P \xi. \tag{2.6}$$

Define $\Psi \in \mathcal{S}(\mathbb{R}^d)$ by

$$\widehat{\Psi}(\xi) = \widehat{\Phi}(Q\xi).$$

Then by (2.5) and (2.6),

$$\int_V F\ d\sigma = \frac{1}{\Psi(0)} \lim_{\varepsilon\to 0^+} \int_{\mathbb{R}^d} F(\xi) \frac{1}{\varepsilon^p} \widehat{\Psi}\left(\frac{Q^{-1} P \xi}{\varepsilon}\right)\ d\vec{x}$$

$$= \frac{1}{\Psi(0)} \lim_{\varepsilon \to 0^+} \int_{\mathbb{R}^d} F(\xi) \frac{1}{\varepsilon^p} \widehat{\Phi}\left(\frac{P\xi}{\varepsilon}\right)\, d\vec{x}.$$

Finally,

$$\begin{aligned}
\Psi(0) &= \int_{\mathbb{R}^p} \widehat{\Psi}(\xi)\, d\xi \\
&= \int_{\mathbb{R}^p} \widehat{\Phi}(Q\xi)\, d\xi \\
&= \frac{1}{|Q|} \int_{\mathbb{R}^p} \widehat{\Phi}(\xi)\, d\xi \\
&= \frac{1}{|Q|} \Phi(0) = \frac{1}{|Q|}
\end{aligned}$$

and so the result follows with $C_P = |Q|$. Note that Q is also independent of the choice of basis, and is a function only of P. $\square$

We are now in a position to prove Proposition 2.1.

Proof of Proposition 2.1. For $\Phi \in \mathcal{S}(\mathbb{R}^m)$ with $\Phi(0) = 1$, the dominated convergence theorem gives

$$\Lambda(f_1, \ldots, f_k) = \lim_{\varepsilon \to 0^+} \int_{\mathbb{R}^m} \left(\prod_{j=1}^k f_j(A_j \vec{x}) \right) \Phi(\vec{x}\varepsilon)\, d\vec{x}.$$

Applying the Fourier inversion formula in $\mathbb{R}^n$, $g(y) = \int_{\mathbb{R}^n} e^{2\pi i y \cdot \xi} \widehat{g}(\xi)\, d\xi$, we get

$$\begin{aligned}
&\Lambda(f_1, \ldots, f_k) \\
= &\lim_{\varepsilon \to 0^+} \int_{\mathbb{R}^m} \prod_{j=1}^k \left[\int_{\mathbb{R}^n} e^{2\pi i A_j \vec{x} \cdot \xi_j} \widehat{f_j}(\xi_j)\, d\xi_j \right] \Phi(\vec{x}\varepsilon)\, d\vec{x} \\
= &\lim_{\varepsilon \to 0^+} \int_{\vec{\xi} = (\xi_1, \ldots, \xi_k) \in (\mathbb{R}^n)^k} \prod_{j=1}^k \widehat{f_j}(\xi_j) \left[\int_{\mathbb{R}^m} e^{2\pi i \vec{x} \cdot \mathbb{A}^t \xi}\, \Phi(\vec{x}\varepsilon)\, d\vec{x} \right] d\vec{\xi} \\
= &\lim_{\varepsilon \to 0^+} \int_{\vec{\xi} \in (\mathbb{R}^n)^k} \prod_{j=1}^k \widehat{f_j}(\xi_j) \frac{1}{\varepsilon^p} \widehat{\Phi}\left(\frac{\mathbb{A}^t \xi}{\varepsilon}\right) d\vec{\xi}
\end{aligned}$$

$$= C_{\mathbb{A}^t} \int_S \prod_{j=1}^{k} \widehat{f}_j(\xi_j)\ d\sigma(\xi),$$

for some constant $C_{\mathbb{A}^t}$. This last step follows from Proposition 2.4, with $p = m$, $d = nk$, $V = S$, $P = \mathbb{A}^t$, and $F = \prod \widehat{f}_j$. $\square$

Proposition 2.5. *Let $g \in \mathcal{S}(\mathbb{R}^m)$, $f_j \in C_c^\infty(\mathbb{R}^n)$. Then the integral*

$$\Theta(g; f_1, \ldots, f_k) := \int_{\mathbb{R}^m} g(\vec{x}) \prod_{j=1}^{k} f_j(A_j \vec{x})\ d\vec{x}$$

is absolutely convergent, and admits the representation

$$\Theta(g; f_1, \ldots, f_k) = \int_{(\mathbb{R}^n)^k} \widehat{g}(-\mathbb{A}^t \vec{\xi}) \prod_{j=1}^{k} \widehat{f}_j(\vec{\xi}_j)\ d\vec{\xi}.$$

Proof. By Fourier inversion,

$$\Theta(g; f_1, \ldots, f_k) = \int_{\mathbb{R}^m} g(\vec{x}) \prod_{j=1}^{k} \left[\int_{\mathbb{R}^n} e^{2\pi i A_j \vec{x} \cdot \vec{\xi}_j} \widehat{f}_j(\vec{\xi}_j)\ d\vec{\xi}_j \right]\ d\vec{x},$$

which is absolutely convergent since $\widehat{f}_j, g \in \mathcal{S}(\mathbb{R}^m)$. Then by Fubini's Theorem,

$$\begin{aligned}\Theta(g; f_1, \ldots, f_k) &= \int_{(\mathbb{R}^n)^k} \prod_{j=1}^{k} \widehat{f}_j(\vec{\xi}_j) \left[\int_{\mathbb{R}^m} g(x) e^{2\pi i \vec{x} \cdot \mathbb{A}^t \vec{\xi}}\ d\vec{x} \right]\ d\vec{\xi} \\ &= \int_{(\mathbb{R}^n)^k} \widehat{g}(-\mathbb{A}^t \vec{\xi}) \prod_{j=1}^{k} \widehat{f}_j(\vec{\xi}_j)\ d\vec{\xi},\end{aligned}$$

where the last line follows by the definition of the Fourier transform. $\square$

2.2 Extension of the multilinear form to measures

With S as in (2.2), denote $S^{\perp} = \{\tau \in (\mathbb{R}^n)^k : \tau \cdot \xi = 0 \text{ for all } \xi \in S\}$ and fix a $\tau \in S^{\perp}$. We will use the variable

$$\eta = (\eta_1, \dots, \eta_k) = \xi + \tau = (\xi_1, \dots, \xi_k) + (\tau_1, \dots, \tau_k) \in S + \tau, \tag{2.7}$$

where $\xi \in S$, and $\eta_j, \xi_j, \tau_j \in \mathbb{R}^n$. Define

$$\Lambda_\tau^*(g_1, \dots, g_k) = \int_{S+\tau} \prod_{j=1}^{k} g_j(\eta_j) \; d\sigma(\eta), \tag{2.8}$$

initially defined for $g_j \in C_c(\mathbb{R}^n)$ so that the integral is absolutely convergent. If $g_j = g$ for all j, we write $\Lambda_\tau^*(g)$ instead of $\Lambda_\tau^*(g, \dots, g)$. We will use Λ^* in place of Λ_0^*. Our next proposition shows that we may extend this multilinear form to continuous functions with appropriate decay.

In applications, g_j will be $\widehat{\mu}$, with μ as in Theorem 1.16. While defining the multilinear form Λ for measures requires only the use of Λ_0^* (so that the integration is on S as in Proposition 2.1), the proof of our main result will rely crucially on estimates on Λ_τ^* uniform in τ.

Proposition 2.6. *Let $\{A_1, \dots, A_k\}$ be a non-degenerate collection of $n \times m$ matrices in the sense of Definition 1.15. Assume that $nk/2 < m < nk$ and $g_1, \dots, g_k : \mathbb{R}^n \to \mathbb{C}$ are continuous functions satisfying*

$$|g_j(\kappa)| \le M(1 + |\kappa|)^{-\beta/2}, \qquad \kappa \in \mathbb{R}^n, \tag{2.9}$$

for some $\beta > 2(nk - m)/k$. Then the integral defining $\Lambda_\tau^* = \Lambda_\tau^*(g_1, \dots, g_k)$ is absolutely convergent for every $\tau \in S^{\perp}$. Indeed,

$$\sup_{\tau \in S^{\perp}} \Lambda_\tau^*(|g_1|, \dots, |g_k|) \le C$$

where C depends only on n, k, m, M, and $\mathbb{A}$.

Since $S + \tau$ for $\tau \in S^{\perp}$ includes all possible translates of S, Proposition

2.6 gives a uniform upper bound on the integral of $\prod g_j$ over all affine copies of S.

The proof of this proposition is based on a lemma which requires some additional notation. Let $0_{a\times b}$ denote the $a \times b$ matrix consisting of 0's . Let $I_{n\times n}$ denote the $n \times n$ identity matrix . Define the $nk \times n$ matrix E_j $(1 \le j \le k)$ by

$$E_j^t = (0_{n\times n} \ \cdots \ I_{n\times n} \ \cdots \ 0_{n\times n}),$$

where $I_{n\times n}$ is in the jth block. For any $\vec{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n) \in \{0,1\}^n$, we denote $I_{n\times n}(\vec{\varepsilon}) = \mathrm{diag}(\varepsilon_1, \ldots, \varepsilon_n)$. For a subset $J' \subseteq \{1, \ldots, n\}$ and index $j \in \{1, \ldots, k\}$, define the $nk \times n$ matrix

$$E_j(J')^t = (0_{n\times n} \ \cdots \ I_{n\times n}(\vec{\varepsilon}) \ \cdots \ 0_{n\times n}),$$

where $I_{n\times n}(\vec{\varepsilon})$ is in the jth block and $\varepsilon_i = 1$ if and only if $i \in J'$. Finally, for $\xi \in S$ we define

$$\xi_j(J') = E_j(J')\xi.$$

Lemma 2.7. *Let $\{A_1, \ldots, A_k\}$ be a non-degenerate collection of $n \times m$ matrices in the sense of Definition 1.15. Let $J = \{j_1, \ldots, j_r\} \subseteq \{1, \ldots, k\}$ and $J' \subseteq \{1, \ldots, n\}$ be collections of distinct indices with $\#(J') = n'$, defined by (1.5). Then the projection of $(\xi_{j_1}, \ldots, \xi_{j_{r-1}}, \xi_{j_r}(J'))$ on S is a coordinate system on S as defined in (2.2). In particular, there exists a constant $C = C(J, J', j, \mathbb{A})$ such that $d\sigma(\xi) = C \ d\xi_{j_1} \cdots d\xi_{j_{r-1}} d\xi_{j_r}(J')$, where $\sigma(\xi)$ is the Lebesgue measure on S.*

Proof. It suffices to prove

$$S' = \{\xi \in S : \ \xi_{j_1} = \xi_{j_2} = \cdots = \xi_{j_{r-1}} = \xi_{j_r}(J') = 0\}$$

has dimension 0.

We will examine $S' = \{\xi \in S : \ \xi_1 = \xi_2 = \cdots = \xi_{r-1} = \xi_r(J') = 0\}$; the other cases are similar. S' is the subspace defined by $\xi \in (\mathbb{R}^n)^k$ satisfying

$$\begin{pmatrix} A_1^t & A_2^t & \cdots & A_{r-1}^t & A_r^t & A_{r+1}^t & \cdots & A_k^t \\ I_{n\times n} & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & I_{n\times n} & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & I_{n\times n} & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & I_{n\times n}(\vec{1}-\vec{\varepsilon}) & 0 & \cdots & 0 \end{pmatrix} \xi = 0,$$

where $\varepsilon_i = 1$ if and only if $i \notin J'$, and $\vec{1} = (1, \ldots, 1)$. For $\dim S' = 0$, we need the kernel of the above $(m + nr) \times nk$ matrix to have dimension 0, so we need the rank of said matrix to be nk. Notice $I_{n\times n}$ is of rank n and we have $r - 1$ of these, and $I_{n\times n}(\vec{1} - \vec{\varepsilon})$ has rank $n' = nk - m - n(r - 1)$, so it suffices for

$$\operatorname{rank}\begin{pmatrix} (A_r(\vec{\varepsilon}))^t & A_{r+1}^t & \cdots & A_k^t \end{pmatrix} = nk - n(r-1) - n' = m.$$

A similar condition holds if we examine $S' = \{\xi \in S : \xi_{i_1} = \xi_{i_2} = \cdots = \xi_{i_{r-1}} = \xi_{i_r}(J') = 0\}$ in general, and upon taking the transpose we arrive at the sufficient condition

$$\operatorname{rank}\begin{pmatrix} \mathbb{A}_I \\ A_{i_{k-(r-1)}}(\vec{\varepsilon}) \end{pmatrix} = m \tag{2.10}$$

for $I = \{i_1, \ldots, i_{k-r}\} \subseteq \{1, \ldots, k\}$ a set of distinct indices, and $\vec{\varepsilon} \cdot \vec{1} = n - n'$. Notice this means the matrix is of full rank. (2.10) follows from the non-degeneracy assumption in Definition 1.15. $\square$

Proof of Proposition 2.6. Fix $\tau \in S^\perp$. We use Sym_k to denote the symmetric group on k elements. For a permutation $\theta \in \mathrm{Sym}_k$, we define the region

$$\Omega_\theta = \{\eta \in S + \tau : |\eta_{\theta(1)}| \leq |\eta_{\theta(2)}| \leq \cdots \leq |\eta_{\theta(k)}|\}$$

so that

$$S + \tau = \bigcup_{\theta \in \mathrm{Sym}_k} \Omega_\theta.$$

For simplicity and without loss of generality, we will examine the case of $\theta = \mathrm{id}$, the identity permutation, and write Ω_{id} as Ω; the other cases are analogous. It then suffices to show the convergence of the integral

$$I = \int_\Omega \prod_{j=1}^k |g_j(\eta_j)| \ d\sigma \lesssim \int_\Omega \prod_{j=1}^k (1 + |\eta_j|)^{-\beta/2} \ d\sigma.$$

Let $L = \prod_{j=1}^{r-1}(1 + |\eta_j|)^{-\beta/2}$. By Hölder's inequality over $k - (r-1)$ terms,

$$\begin{aligned}
I &\lesssim \int_\Omega \prod_{j=1}^k (1 + |\eta_j|)^{-\beta/2} \ d\sigma \\
&= \int_\Omega L \prod_{i=r}^k (1 + |\eta_i|)^{-\beta/2} \ d\sigma \\
&= \int_\Omega \prod_{i=r}^k \left[L^{1/(k-(r-1))} (1 + |\eta_i|)^{-\beta/2} \right] \ d\sigma \\
&\leq \prod_{i=r}^k \left[\int_{\substack{S+\tau \\ |\eta_1| \leq \cdots \leq |\eta_{r-1}| \leq |\eta_i|}} L (1 + |\eta_i|)^{-\frac{\beta}{2}(k-(r-1))} \ d\sigma \right]^{1/(k-(r-1))} \\
&\leq \int_{\substack{S+\tau \\ |\eta_1| \leq \cdots \leq |\eta_r|}} \prod_{j=1}^{r-1} (1 + |\eta_j|)^{-\beta/2} (1 + |\eta_r|)^{-\frac{\beta}{2}(k-(r-1))} \ d\sigma.
\end{aligned}$$

In the last inequality, we see that for each i, the integral is the same with a different dummy variable, so we collect the terms under the single index $i = r$. Recalling our decomposition of η into ξ and τ as per (2.7) and since τ is fixed, we arrive at

$$I \lesssim \int_{\substack{S \\ |\xi_1+\tau_1| \leq \cdots \leq |\xi_r+\tau_r|}} \prod_{j=1}^{r-1} (1 + |\xi_j + \tau_j|)^{-\beta/2} (1 + |\xi_r + \tau_r|)^{-\frac{\beta}{2}(k-(r-1))} \ d\sigma(\xi).$$

Suppose $\eta_r = (\eta_{r,1}, \ldots, \eta_{r,n})$, and for any permutation $\pi \in \mathrm{Sym}_n$, let

$$\eta_r^\pi = (\eta_{r,\pi(1)}, \ldots, \eta_{r,\pi(n')}).$$

As in the beginning of this section, we may partition our current region of integration into a finite number of regions of the form $\{|\eta_{r,\pi(1)}| \geq \cdots \geq |\eta_{r,\pi(n)}|\}$ for $\pi \in \mathrm{Sym}_n$. Then $|\eta_r| \leq n|\eta_{r,\pi(1)}|$, so that $|\eta_r| \approx |\eta_r^\pi|$.

By Lemma 2.7, $d\sigma(\xi) = C\ d\xi_1 \cdots d\xi_{r-1} d\xi_r^\pi$, hence

$$\begin{aligned}
I &\lesssim \int_{\substack{S \\ |\xi_1+\tau_1| \leq \cdots \leq |\xi_r+\tau_r|}} \prod_{j=1}^{r-1} (1 + |\xi_j + \tau_j|)^{-\beta/2} \\
&\qquad \cdot (1 + |\xi_r^\pi + \tau_r^\pi|)^{-\frac{\beta}{2}(k-(r-1))}\ d\xi_1 \cdots d\xi_{r-1} d\xi_r^\pi \\
&\lesssim \int_{\mathbb{R}^{n'}} \prod_{j=1}^{r-1} \left[\int_{\substack{\xi_j+\tau_j \in \mathbb{R}^n \\ |\xi_j+\tau_j| \leq |\xi_r^\pi + \tau_r^\pi|}} (1 + |\xi_j + \tau_j|)^{-\beta/2}\ d\xi_j \right] \\
&\qquad \cdot (1 + |\xi_r^\pi + \tau_r^\pi|)^{-\frac{\beta}{2}(k-(r-1))}\ d\xi_r^\pi.
\end{aligned}$$

Translating ξ_j by τ_j,

$$\begin{aligned}
I &\lesssim \int_{\mathbb{R}^{n'}} \prod_{j=1}^{r-1} \left[\int_{\substack{\xi_j \in \mathbb{R}^n \\ |\xi_j| \leq |\xi_r^\pi|}} (1 + |\xi_j|)^{-\beta/2}\ d\xi_j \right] (1 + |\xi_r^\pi|)^{-\frac{\beta}{2}(k-(r-1))}\ d\xi_r^\pi \\
&\lesssim \int_{\mathbb{R}^{n'}} \left[\int_0^{|\xi_r^\pi|} (1+\rho)^{-\beta/2+n}\ d\rho \right]^{r-1} (1 + |\xi_r^\pi|)^{-\frac{\beta}{2}(k-(r-1))}\ d\xi_r^\pi \\
&\lesssim \int_{\mathbb{R}^{n'}} (1 + |\xi_r^\pi|)^{(n-\beta/2)(r-1) - \frac{\beta}{2}(k-(r-1))}\ d\xi_r^\pi \\
&= \int_{\mathbb{R}^{n'}} (1 + |\xi_r^\pi|)^{n(r-1)-\beta k/2}\ d\xi_r^\pi,
\end{aligned}$$

where the Jacobian in making the spherical change of coordinates $\rho = |\xi_j|$ above is independent of τ_j. This last expression is finite (with a bound independent of τ) when

$$\beta k/2 - n(r-1) > n' = nk - m - n(r-1),$$

which holds since $2(nk-m)/k < \beta < n$ and $m > nk/2$. $\square$

2.3 Counting geometric configurations in sparse sets

In this section, we will show that the multilinear form Λ^* defined in (2.8) is effective in counting non-trivial configurations supported on appropriate sparse sets.

Proposition 2.8. *Suppose $nk/2 < m < nk$ and $2(nk-m)/k < \beta < n$. Let $\{A_1, \ldots, A_k\}$ be a collection of $n \times m$ matrices that are non-degenerate in the sense of Definition 1.15. Let μ be a positive, finite, Radon measure μ with*

$$\sup_{\xi \in \mathbb{R}^n} |\widehat{\mu}(\xi)|(1+|\xi|)^{\beta/2} \leq C. \tag{2.11}$$

Then there exists a non-negative, finite, Radon measure $\nu = \nu(\mu)$ on $[0,1]^m$ such that

(a) $\nu(\mathbb{R}^m) = \Lambda^*(\widehat{\mu})$.

(b) $\operatorname{supp} \nu \subseteq \{x \in \mathbb{R}^m : \ A_1 x, \ldots, A_k x \in \operatorname{supp} \mu\}$.

(c) For any subspace $V \subseteq \mathbb{R}^m$ with $\dim V < m$, $\nu(V) = 0$.

2.3.1 Existence of candidate ν

Fix a non-negative $\phi \in \mathcal{S}(\mathbb{R}^m)$ with $\int \phi = 1$ and let $\phi_\varepsilon(y) = \varepsilon^{-n}\phi(\varepsilon^{-1}y)$. Let $\mu_\varepsilon = \mu * \phi_\varepsilon$. Notice $\widehat{\phi} \in \mathcal{S}(\mathbb{R}^m)$ since $\phi \in \mathcal{S}(\mathbb{R}^m)$, so

$$|\widehat{\mu}_\varepsilon(\xi)| = |\widehat{\mu}(\xi)\widehat{\phi}(\varepsilon\xi)| \leq C(1+|\xi|)^{-\beta/2} \tag{2.12}$$

with $C = \|\widehat{\phi}\|_\infty$ independent of ε. Furthermore, $\widehat{\phi}(\varepsilon\xi) \to \widehat{\phi}(0) = \int \phi = 1$ as $\varepsilon \to 0$, hence

$$\widehat{\mu}_\varepsilon(\xi) \to \widehat{\mu}(\xi) \text{ pointwise as } \varepsilon \to 0. \tag{2.13}$$

We prove that the multilinear form Λ^*_τ satisfies a weak continuity property, in the following sense:

Lemma 2.9. *$\Lambda_\tau^*(\widehat{\mu}_\varepsilon) \to \Lambda_\tau^*(\widehat{\mu})$ as $\varepsilon \to 0$ for every fixed $\tau \in S^\perp$.*

Proof. Fix $\tau \in S^\perp$. By definition,

$$\Lambda_\tau^*(\widehat{\mu}_\varepsilon) = \int_{S+\tau} \prod_{j=1}^{k} \widehat{\mu}_\varepsilon(\xi_j)\ d\xi.$$

By (2.12), $|\widehat{\mu}_\varepsilon(\eta)| \leq C(1+|\eta|)^{-\beta/2} =: g(\eta)$ uniformly in ε, so

$$\prod_{j=1}^{k} |\widehat{\mu}_\varepsilon(\xi_j)| \leq C \prod_{j=1}^{k} g(\xi_j).$$

By Proposition 2.6, $\Lambda_\tau^*(g)$ is finite, so by (2.13) and the dominated convergence theorem,

$$\Lambda_\tau^*(\widehat{\mu}_\varepsilon) \to \Lambda_\tau^*(\widehat{\mu}).$$

□

For $F \in C([0,1]^m)$ such that $\widehat{F} \in \mathcal{S}(\mathbb{R}^m)$, define the linear functional ν by

$$\langle \nu, F \rangle = \lim_{\varepsilon \to 0} \int_{\mathbb{R}^m} F(\vec{x}) \prod_{j=1}^{k} \mu_\varepsilon(A_j \vec{x})\ d\vec{x}. \tag{2.14}$$

We will prove in Lemma 2.10 below that the limit exists and extends as a bounded linear functional on $C([0,1])$. Clearly, $\langle \nu, F\rangle \geq 0$ if $F \geq 0$. By the Riesz representation theorem, there exists a non-negative, finite, Radon measure ν that identifies this linear functional; namely $\langle \nu, F \rangle = \int F\ d\nu$.

Lemma 2.10. *There exists a non-negative, bounded, linear functional ν on $C([0,1])$, that is,*

$$|\langle \nu, F \rangle| \leq C \|F\|_\infty \tag{2.15}$$

for some positive constant C independent of $F \in C([0,1])$, which agrees with (2.14) if $\widehat{F} \in \mathcal{S}(\mathbb{R}^m)$.

Proof. Assume the limit (2.14) exists for $F \in C([0,1])$, $\widehat{F} \in \mathcal{S}(\mathbb{R}^m)$. Then

$$\begin{aligned}
|\langle \nu, F\rangle| &\le \lim_{\varepsilon\to 0} \int_{\mathbb{R}^m} |F(\vec{x})| \prod_{j=1}^{k} \mu_\varepsilon(A_j\vec{x})\, d\vec{x} \\
&\le \|F\|_\infty \lim_{\varepsilon\to 0} \Lambda(\mu_\varepsilon) \\
&= \|F\|_\infty \lim_{\varepsilon\to 0} \Lambda^*(\widehat{\mu}_\varepsilon) \\
&\le C\|F\|_\infty,
\end{aligned}$$

where the last line follows by Proposition 2.6, with a constant C independent of ε. Thus, (2.15) holds.

It remains to prove that $\langle \nu, F\rangle$ is well-defined. We will prove this by showing that the limit in (2.14) exists for $F \in C([0,1]^m)$ with $\widehat{F} \in \mathcal{S}(\mathbb{R}^m)$ and use density arguments to extend the functional to all of $C([0,1]^m)$. Applying Proposition 2.5 with $g = F$, $f_1 = \ldots = f_k = \mu_\varepsilon$, we obtain

$$\langle \nu, F\rangle = \lim_{\varepsilon\to 0} \Theta(F;\mu_\varepsilon,\ldots,\mu_\varepsilon) = \lim_{\varepsilon\to 0} \int_{\mathbb{R}^{nk}} \widehat{F}(-\mathbb{A}^t\xi) \prod_{j=1}^{k} \widehat{\mu}_\varepsilon(\xi_j)\, d\xi. \tag{2.16}$$

By (2.13),

$$\widehat{F}(-\mathbb{A}^t\xi) \prod_{j=1}^{k} \widehat{\mu}_\varepsilon(\xi_j) \to \widehat{F}(-\mathbb{A}^t\xi) \prod_{j=1}^{k} \widehat{\mu}(\xi_j)$$

pointwise, and by (2.12),

$$\left| \widehat{F}(-\mathbb{A}^t\xi) \prod_{j=1}^{k} \widehat{\mu}_\varepsilon(\xi_j) \right| \le C|\widehat{F}(-\mathbb{A}^t\xi)| \prod_{j=1}^{k} (1+|\xi_j|)^{-\beta/2}.$$

Existence of the limit in (2.16) will follow from the dominated convergence theorem, if we prove $|\widehat{F}(\mathbb{A}^t\xi)| \prod_{j=1}^{k}(1+|\xi_j|)^{-\beta/2} \in L^1(\mathbb{R}^{nk})$. To this end, let $g(t) = (1+|t|)^{\beta/2}$ for $t \in \mathbb{R}^n$. Then

$$\int_{\mathbb{R}^{nk}} |\widehat{F}(\mathbb{A}^t\xi)| \prod_{j=1}^{k} (1+|\xi_j|)^{-\beta/2}\, d\xi = \int_{\mathbb{R}^m} |\widehat{F}(\kappa)| \int_{\mathbb{A}^t\xi=\kappa} \prod_{j=1}^{k} g(\xi_j)\, d\sigma(\xi) d\kappa$$

$$
\begin{aligned}
&= \int_{\mathbb{R}^m} |\widehat{F}(\kappa)| \Lambda^*_{\tau(\kappa)}(g)\ d\kappa \\
&\le C \int_{\mathbb{R}^m} |\widehat{F}(\kappa)|\ d\kappa < \infty.
\end{aligned}
$$

Here $\tau(\kappa)$ is the unique vector in $S^\perp$ such that $\{\mathbb{A}^t\xi = \kappa\} = S + \tau(\kappa)$. We have used Proposition 2.6 to bound $\Lambda^*_{\tau(\kappa)}$ in the last displayed inequality above, and used the fact that $\widehat{F} \in \mathcal{S}(\mathbb{R}^m)$ to deduce that $\widehat{F} \in L^1(\mathbb{R}^m)$. By the dominated convergence theorem, the limit in (2.14) exists.

To extend ν to all of $C([0,1]^m)$, fix $F \in C([0,1]^m)$. Extend F to $\widetilde{F} \in C_c(\mathbb{R}^m)$ so that $F = \widetilde{F}$ on $[0,1]^m$. We will reuse F to mean $\widetilde{F}$ for convenience. Get a sequence of functions $F_n \in C^\infty([0,1]^m)$ with $\widehat{F_n} \in \mathcal{S}(\mathbb{R}^m)$ such that $\|F - F_n\|_\infty \to 0$. By the preceding proof,

$$
|\langle \nu, F_n - F_m\rangle| \le C\|F_n - F_m\|_\infty \to 0
$$

as $n, m \to \infty$ since F_n is Cauchy in supremum norm. Thus the sequence of scalars $\langle \nu, F_n\rangle$ is Cauchy and hence converges. Define

$$
\langle \nu, F\rangle = \lim_{n\to\infty} \langle \nu, F_n\rangle.
$$

Clearly,

$$
|\langle \nu, F\rangle| = \lim_{n\to\infty} |\langle \nu, F_n\rangle| \le C \lim_{n\to\infty} \|F_n\|_\infty = C\|F\|_\infty.
$$

□

The proof of Lemma 2.10 yields the following corollary which will be used later in the sequel:

Corollary 2.11. *For $F \in C([0,1]^m)$ with $\widehat{F} \in \mathcal{S}(\mathbb{R}^m)$,*

$$
\langle \nu, F\rangle = \int_{\mathbb{R}^{nk}} \widehat{F}(-\mathbb{A}^t\xi) \prod_{j=1}^{k} \widehat{\mu}(\xi_j)\ d\xi.
$$

2.3.2 Proof of Proposition 2.8.(a)

Proof. We have

$$\nu(\mathbb{R}^m) = \langle \nu, 1\rangle = \lim_{\varepsilon\to 0}\int_{\mathbb{R}^m}\prod_{j=1}^{k}\mu_\varepsilon(A_j x)\,dx = \lim_{\varepsilon\to 0}\Lambda(\mu_\varepsilon) = \lim_{\varepsilon\to 0}\Lambda^*(\widehat{\mu}_\varepsilon) = \Lambda^*(\widehat{\mu})$$

by Lemma 2.9, with $\tau = 0$. $\square$

2.3.3 Proof of Proposition 2.8.(b)

Proof. Define

$$X := \{x \in \mathbb{R}^m :\ A_1 x, \dots, A_k x \in \operatorname{supp}\mu\}.$$

Since $\operatorname{supp}\mu$ is closed, X is closed. Let F be any continuous function on $\mathbb{R}^m$ with $\operatorname{supp} F$ disjoint from X, then $\operatorname{dist}(\operatorname{supp} F, X) > 0$. In order to prove that ν is supported on X, we aim to show that $\langle \nu, F\rangle = 0$. To this end, let us define

$$\begin{aligned} X_N &:= \{\vec{x} \in \mathbb{R}^m :\ \operatorname{dist}(A_j\vec{x}, \operatorname{supp}(\mu)) \le 1/N \text{ for every } 1 \le j \le k\} \\ &= \bigcap_{j=1}^{k}\{\vec{x} \in \mathbb{R}^m :\ \operatorname{dist}(A_j\vec{x}, \operatorname{supp}(\mu)) \le 1/N\}. \end{aligned}$$

Then $X \subseteq X_N$ for every N, and $X = \bigcap_{N=1}^{\infty} X_N$. Furthermore,

$$X_N^c = \bigcup_{j=1}^{k}\{\vec{x} \in \mathbb{R}^m :\ \operatorname{dist}(A_j\vec{x}, \operatorname{supp}(\mu)) > 1/N\}$$

is an open set for every $N \ge 1$, with

$$\operatorname{supp}(F) \subseteq X^c = \bigcup_{N=1}^{\infty} X_N^c.$$

Introducing a smooth partition of unity subordinate to $\{X_N^c\}_N$, we can write $F = \sum_N F_N$ where each $F_N \in C_c^\infty(\mathbb{R}^m)$ with $\operatorname{supp}(F_N) \subseteq X_N^c$. Note that since supp (F) is a compact subset of X^c, it follows from the definition of a partition of unity that the infinite sum above is in fact a finite sum, so there is no issue of convergence.

Let $\mu_\varepsilon^{A_j}(\vec{x}) := \mu_\varepsilon(A_j\vec{x})$. To compute

$$\begin{aligned}\langle \nu, F_N\rangle &= \lim_{\varepsilon\to 0}\int_{\mathbb{R}^m} F_N(\vec{x})\prod_{j=1}^k \mu_\varepsilon(A_j\vec{x})\; d\vec{x}\\ &= \lim_{\varepsilon\to 0}\int_{\mathbb{R}^m} F_N(\vec{x})\prod_{j=1}^k \mu_\varepsilon^{A_j}(\vec{x})\; d\vec{x}\end{aligned}$$

for a fixed $N \geq 1$, we observe that

$$\begin{aligned}\operatorname{supp}(\mu_\varepsilon^{A_j}) &\subseteq \{\vec{x}\in\mathbb{R}^m :\ \operatorname{dist}(A_j\vec{x}, \operatorname{supp}(\mu)) \leq \varepsilon\}\\ &\subseteq \{\vec{x}\in\mathbb{R}^m :\ \operatorname{dist}(A_j\vec{x}, \operatorname{supp}(\mu)) \leq 1/N\}\end{aligned}$$

if $\varepsilon \leq 1/N$. Thus the product $\prod_{j=1}^k \mu_\varepsilon^{A_j}(\vec{x})$ is supported on X_N, whereas F_N is supported on X_N^c. This implies

$$\int_{\mathbb{R}^m} F_N(\vec{x})\prod_{j=1}^k \mu_\varepsilon^{A_j}(\vec{x})\; d\vec{x} = 0$$

for all $\varepsilon \leq 1/N$, so that $\langle \nu, F_N\rangle = 0$ for every $N \geq 1$. Therefore, $\langle \nu, F\rangle = 0$ as claimed. □

2.3.4 Proof of Proposition 2.8.(c)

Proof. It suffices to prove the proposition for $\dim V = v = m-1$, since smaller subspaces have even less measure. Let P_V denote the projection onto V. Fix $v_0 \in V$ and define

$$V_{\delta,\gamma} = \{x\in\mathbb{R}^m :\ |v_0 - P_V x| \leq \gamma, \operatorname{dist}(x, V) = |P_{V^\perp} x| \leq \delta\}.$$

It suffices to prove $\nu(V_{\delta,\infty}) \to 0$ as $\delta \to 0$. If ϕ_δ is any smooth function with

$$\phi_\delta = \begin{cases} 1 & \text{on } V_{\delta,1}, \\ 0 & \text{on } \mathbb{R}^m \setminus V_{\delta,2}, \end{cases} \tag{2.17}$$

then $\nu(V_{\delta,\infty}) \leq \int \phi_\delta \, d\nu = \langle \nu, \phi_\delta \rangle$, so we aim to show that $\langle \nu, \phi_\delta \rangle \to 0$ as $\delta \to 0$.

Fix bases $\{\mathfrak{a}_1, \ldots, \mathfrak{a}_{m-1}\}$ and $\{\mathfrak{b}\}$ for V and $V^\perp$ respectively, such that $\{\mathfrak{a}_1, \ldots, \mathfrak{a}_{m-1}, \mathfrak{b}\}$ forms an orthonormal basis of $\mathbb{R}^m$. Thus, for any $x \in \mathbb{R}^m$, there is a unique decomposition

$$\begin{gathered} x = u + w, \text{ with } u = \sum_{j=1}^{m-1} a_j \mathfrak{a}_j \in V, w = b\mathfrak{b} \in V^\perp, \\ \text{where } \vec{a} = (a_1, \ldots, a_{m-1})^t \in \mathbb{R}^{m-1}, b \in \mathbb{R}. \end{gathered} \tag{2.18}$$

Without loss of generality, we may assume ϕ_δ as in (2.17) to be variable-separated as

$$\phi_\delta(x) = \phi_V(\vec{a})\phi_{V^\perp}(\delta^{-1}b), \tag{2.19}$$

where $\phi_V \in C_c^\infty(\mathbb{R}^{m-1})$ is supported on $\{\vec{a} : \, |\sum_{j=1}^{m-1} a_j \mathfrak{a}_j - v_0| \leq 2\}$ and $\phi_{V^\perp} \in C_c^\infty(\mathbb{R})$ is supported on $\{b : \, |b| \leq 2\}$.

By Corollary 2.11,

$$\langle \nu, \phi_\delta \rangle = \int_{\mathbb{R}^{nk}} \widehat{\phi}_\delta(-\mathbb{A}^t \xi) \prod_{j=1}^{k} \widehat{\mu}(\xi_j) \, d\xi. \tag{2.20}$$

We will show that this integral tends to 0 as $\delta \to 0$. The estimation of this integral relies on an orthogonal decomposition of $\mathbb{R}^{nk}$ into specific subspaces, which we now describe. Let

$$W = \{\xi \in \mathbb{R}^{nk} : \, \mathbb{A}^t \xi \cdot x = 0 \text{ for all } x \in V\}. \tag{2.21}$$

Then S is clearly a subspace of W, as $\mathbb{A}^t \xi = 0$ if $\xi \in S$. It is also not difficult to see that $\dim W = nk - v = nk - (m-1)$. The proof of this has been relegated to Lemma 2.12 below. A consequence of this fact is that

$$\dim W \cap S^{\perp} = 1 \tag{2.22}$$

since S is $(nk-m)$-dimensional. Now write $\xi \in \mathbb{R}^{nk}$ as

$$\xi = \zeta + \eta + \lambda, \text{ where } \zeta \in S, \eta \in W \cap S^{\perp}, \lambda \in W^{\perp},$$

so that $d\xi = d\sigma_{S+\eta+\lambda}(\xi)\ d\eta\ d\lambda$. Here $d\sigma_{S+\eta+\lambda}$ denotes the surface measure on $S+\eta+\lambda$, as defined in Definition 2.2. We will soon show, in Lemmas 2.13 and 2.14 below, that the two factors of the integrand in (2.20) obey the size estimates:

$$\left|\prod_{j=1}^{k} \widehat{\mu}(\xi_j)\right| \lesssim (1 + |\mathbb{A}^t(\eta+\lambda)\cdot\mathfrak{b}|)^{-\varepsilon} G(\xi), \tag{2.23}$$

and

$$|\widehat{\phi}_\delta(-\mathbb{A}^t\xi)| \leq \delta C_M (1+|\lambda|)^{-M}(1+\delta|\mathbb{A}^t(\eta+\lambda)\cdot\mathfrak{b}|)^{-M} \tag{2.24}$$

for any $M \geq 1$. Here, $G(\xi) = \prod_{j=1}^{k} g(\xi_j)$, where $g(\xi_j) = (1+|\xi_j|)^{-\beta/2+\varepsilon}$ and $\varepsilon > 0$ is chosen sufficiently small so that

$$\beta - 2\varepsilon > 2(nk-m)/k. \tag{2.25}$$

Notice this is possible since $\beta > 2(nk-m)/k$.

Assuming (2.23) and (2.24) temporarily, the estimation of (2.20) proceeds as follows.

$$\begin{aligned}|\langle \nu, \phi_\delta\rangle| &\leq \int_{\mathbb{R}^{nk}} \left|\widehat{\phi}_\delta(-\mathbb{A}^t\xi)\prod_{j=1}^{k}\widehat{\mu}(\xi_j)\right|\ d\xi \\ &\lesssim \int_{W^{\perp}} \left[\int_{W\cap S^{\perp}} \left[\int_{S+\eta+\lambda} G(\xi)\ d\sigma_{S+\eta+\lambda}(\xi)\right] J(\eta,\lambda)\ d\eta\right](1+|\lambda|)^{-M}\ d\lambda,\end{aligned}$$

where

$$J(\eta,\lambda) = \delta C_M(1+|\mathbb{A}^t(\eta+\lambda)\cdot\mathfrak{b}|)^{-\varepsilon}(1+\delta|\mathbb{A}^t(\eta+\lambda)\cdot\mathfrak{b}|)^{-M}.$$

We claim that

$$\int_{S+\eta+\lambda} G(\xi)\ d\sigma(\xi) \le C \tag{2.26}$$

and

$$\sup_{\lambda\in W^\perp}\int_{W\cap S^\perp} J(\eta,\lambda)\ d\eta \le C_M\delta^{\varepsilon/2}. \tag{2.27}$$

These two estimates yield, for $M \ge \dim(W^\perp)+1$,

$$\begin{aligned}|\langle\nu,\phi_\delta\rangle| &\lesssim C_M\delta^{\varepsilon/2}\int_{W^\perp}(1+|\lambda|)^{-M}\ d\lambda\\ &\lesssim C_M\delta^{\varepsilon/2}\to 0\end{aligned}$$

as $\delta\to 0$, as required.

It remains to establish the estimates in (2.26) and (2.27). For the former, we observe that the left hand side of the inequality is $\Lambda^*_{\eta+\lambda}(g)$, so the desired conclusion follows from Proposition 2.6 and our choice (2.25) of ε.

To prove (2.27), we recall (2.22) so we may parametrize $\eta = sw_0$ for some fixed unit vector $w_0 \in W\cap S^\perp\setminus\{0\}$, with $d\eta = ds$. To confirm that $J(\eta,\lambda)$ has decay in η, we need to verify that $\mathbb{A}^tw_0\cdot\mathfrak{b}\neq 0$. Indeed, if $\mathbb{A}^tw_0\cdot\mathfrak{b} = 0$, then $\mathbb{A}^tw_0\in V$ since $\mathfrak{b}\in V^\perp$. Since w_0 also lies in W given by (2.21), this implies $\mathbb{A}^tw_0\cdot\mathbb{A}^tw_0 = 0$, so that $\mathbb{A}^tw_0 = 0$. But the last equation says $w_0\in S$, whereas $w_0\in S^\perp$ by assumption. This forces $w_0 = 0$, a contradiction as w_0 is a unit vector.

We now set $c_0 := \mathbb{A}^tw_0\cdot\mathfrak{b}$ which is nonzero by the discussion in the preceding paragraph. Making a linear change of variable $t = sc_0 + \mathbb{A}^t\lambda\cdot\mathfrak{b}$, with Jacobian $ds = dt/c_0$, we proceed to estimate the integral in (2.27) by partitioning the region of integration as follows,

$$\int_{W\cap S^\perp} J(\eta,\lambda)\ d\eta$$

$$
\begin{aligned}
&= \delta C_M \int_{\mathbb{R}} (1+|t|)^{-\varepsilon}(1+\delta|t|)^{-M} \frac{dt}{c_0} \\
&= \delta C_M \int_{|t|\leq \delta^{-1/2}} (1+|t|)^{-\varepsilon}(1+\delta|t|)^{-M}\ dt \\
&\qquad + \delta C_M \int_{|t|>\delta^{-1/2}} (1+|t|)^{-\varepsilon}(1+\delta|t|)^{-M}\ dt \\
&\lesssim \delta C_M \int_{|t|\leq \delta^{-1/2}} 1\ dt + \delta^{\varepsilon/2}\delta C_M \int_{|t|>\delta^{-1/2}} (1+\delta|t|)^{-M}\ dt \\
&\lesssim \delta^{1/2} C_M + \delta^{\varepsilon/2}\delta C_M \int_{\mathbb{R}} (1+\delta|\eta|)^{-M}\ d\eta \\
&\approx \delta^{1/2} C_M + \delta^{\varepsilon/2} C_M \lesssim \delta^{\varepsilon/2} C_M.
\end{aligned}
$$

This completes the proof of (2.27) and hence the proof of the proposition. □

Now we prove the three lemmas required earlier for this proof.

Lemma 2.12. *Define* W *as in (2.21). Then* $\dim W = nk - v = nk - (m-1)$.

Proof. As before, let P_V denote the projection onto V. By (2.21),

$$
\begin{aligned}
W &= \{\xi \in \mathbb{R}^{nk} :\ \mathbb{A}^t \xi \cdot x = 0 \text{ for all } x \in V\} \\
&= \{\xi \in \mathbb{R}^{nk} :\ \mathbb{A}^t \xi \cdot P_V x = 0 \text{ for all } x \in \mathbb{R}^m\} \\
&= \{\xi \in \mathbb{R}^{nk} :\ P_V^t \mathbb{A}^t \xi \cdot x = 0 \text{ for all } x \in \mathbb{R}^m\} \\
&= \{\xi \in \mathbb{R}^{nk} :\ P_V^t \mathbb{A}^t \xi = 0\} \\
&= \mathcal{N}(P_V^t \mathbb{A}^t).
\end{aligned}
$$

Writing $\mathbb{R}^m$ as $V \oplus V^\perp$, the dimension of $\mathbb{A}(V)$ must be equal to $\dim V = m-1$ as $\mathbb{A}$ is of full rank and hence an isomorphism from $\mathbb{R}^m$ to the range of $\mathbb{A}$. Then $\mathbb{A}P_V$ is an isomorphism from V to the range of $\mathbb{A}P_V$, so $\operatorname{rank}(P_V^t \mathbb{A}^t) = \dim(V) = m-1$, and thus $\dim W = nk - (m-1)$. □

Lemma 2.13. *With* $\mu, \xi, \eta, \lambda, \beta$ *defined as in the proof of Proposition 2.8(c), we have*

$$
\left| \prod_{j=1}^k \widehat{\mu}(\xi_j) \right| \lesssim (1 + |\mathbb{A}^t(\eta+\lambda)\cdot \mathfrak{b}|)^{-\varepsilon} \prod_{j=1}^k (1+|\xi_j|)^{-\beta/2+\varepsilon},
$$

for any $\varepsilon > 0$.

Proof. The decay condition (2.11) on $\widehat{\mu}$ gives

$$\left|\prod_{j=1}^{k} \widehat{\mu}(\xi_j)\right| \leq C \prod_{j=1}^{k}(1+|\xi_j|)^{-\varepsilon} \prod_{j=1}^{k}(1+|\xi_j|)^{-\beta/2+\varepsilon},$$

for any $\varepsilon > 0$. We have

$$|\eta+\lambda| \leq |\xi| \leq k \max_{1\leq j\leq k} |\xi_j|,$$

and by Cauchy-Schwarz

$$|\mathbb{A}^t(\eta+\lambda)\cdot \mathfrak{b}| = |(\eta+\lambda)\cdot \mathbb{A}\mathfrak{b}| \leq |\eta+\lambda||\mathbb{A}\mathfrak{b}|.$$

Since $\mathbb{A}\mathfrak{b}$ is fixed, $|\mathbb{A}^t(\eta+\lambda)\cdot\mathfrak{b}| \lesssim |\eta+\lambda|$, and so

$$\begin{aligned}\left|\prod_{j=1}^{k} \widehat{\mu}(\xi_j)\right| &\lesssim (1+|\eta+\lambda|)^{-\varepsilon} \prod_{j=1}^{k}(1+|\xi_j|)^{-\beta/2+\varepsilon} \\ &\lesssim (1+|\mathbb{A}^t(\eta+\lambda)\cdot\mathfrak{b}|)^{-\varepsilon} \prod_{j=1}^{k}(1+|\xi_j|)^{-\beta/2+\varepsilon}.\end{aligned}$$

□

Lemma 2.14. *With $\phi_\delta, \xi, \zeta, \eta, \lambda, \beta$ defined as in the proof of Proposition 2.8(c), we have*

$$|\widehat{\phi_\delta}(-\mathbb{A}^t\xi)| \leq \delta C_M (1+|\lambda|)^{-M}(1+\delta|\mathbb{A}^t(\eta+\lambda)\cdot\mathfrak{b}|)^{-M},$$

for any $M \in \mathbb{R}$.

Proof. Since $\zeta \in S$ implies $\mathbb{A}^t\zeta = 0$, we have

$$\begin{aligned}\widehat{\phi_\delta}(-\mathbb{A}^t\xi) &= \widehat{\phi_\delta}(-\mathbb{A}^t(\zeta+\eta+\lambda)) \\ &= \widehat{\phi_\delta}(-\mathbb{A}^t(\eta+\lambda))\end{aligned}$$

$$= \iint \phi_V(\vec{a})\phi_{V^\perp}(\delta^{-1}b)e^{2\pi i \mathbb{A}^t(\eta+\lambda)\cdot(u+w)}\, du\, dw.$$

By definition, $\eta \in W$ and $u \in V$ give $\mathbb{A}^t\eta \cdot u = 0$, and so

$$\begin{aligned}\widehat{\phi}_\delta(-\mathbb{A}^t\xi) &= \left[\int_V \phi_V(\vec{a})e^{2\pi i\mathbb{A}^t\lambda\cdot A_0\vec{a}}\, du\right]\left[\int_{V^\perp}\phi_{V^\perp}(\delta^{-1})e^{2\pi i b\mathbb{A}^t(\eta+\lambda)\cdot b\mathfrak{b}}\, dw\right]\\ &= \left[\int_{\mathbb{R}^{m-1}}\phi_V(\vec{a})e^{2\pi i A_0^t\mathbb{A}^t\lambda\cdot\vec{a}}\, d\vec{a}\right]\left[\int_{\mathbb{R}}\phi_{V^\perp}(\delta^{-1}b)e^{2\pi i b\mathbb{A}^t(\eta+\lambda)\cdot\mathfrak{b}}\, db\right],\end{aligned}$$

where we have used $u = \sum_{i=1}^{m-1} a_i\mathfrak{a}_i = A_0\vec{a}$ for some matrix A_0, $w = b\mathfrak{b}$, and $du\, dw = da\, db$.

The first factor is by definition $\widehat{\phi}_V(-A_0^t\mathbb{A}^t\lambda)$. Since $\widehat{\phi}_V \in \mathcal{S}(\mathbb{R}^{m-1})$, for every $M \in \mathbb{R}$ we have

$$|\widehat{\phi}_V(-A_0^t\mathbb{A}^t\lambda)| \leq C_M(1+|A_0^t\mathbb{A}^t\lambda|)^{-M}.$$

We claim $|A_0^t\mathbb{A}^t\lambda| \gtrsim |\lambda|$ for all $\lambda \in W^\perp$. Since $A_0^t\mathbb{A}^t$ is linear, it suffices to prove $A_0^t\mathbb{A}^t\lambda \neq 0$ for any $\lambda \in W^\perp$. If $\lambda \in W^\perp$, then by definition of $W^\perp$ there exists $x \in V \setminus \{0\}$ such that $(\mathbb{A}^t\lambda, x) \neq 0$. Then $x = A_0\vec{a}$ for some $\vec{a} \neq 0$, so

$$(A_0^t\mathbb{A}^t\lambda, \vec{a}) = (\mathbb{A}^t\lambda, A_0^t\vec{a}) \neq 0,$$

and hence $A_0^t\mathbb{A}^t\lambda \neq 0$. Then for every $M \in \mathbb{R}$,

$$|\widehat{\phi}_V(-A_0^t\mathbb{A}^t\lambda)| \leq C_M(1+|\lambda|)^{-M}.$$

The second factor is, upon scaling, $\delta\widehat{\phi}_{V^\perp}(-\delta\mathbb{A}^t(\eta+\lambda)\cdot\mathfrak{b})$. As $\widehat{\phi}_{V^\perp} \in \mathcal{S}(\mathbb{R})$, for every $M \in \mathbb{R}$ we have

$$|\delta\widehat{\phi}_{V^\perp}(-\delta\mathbb{A}^t(\eta+\lambda)\cdot\mathfrak{b})| \leq \delta(1+\delta|\mathbb{A}^t(\eta+\lambda)\cdot\mathfrak{b}|)^{-M},$$

completing the proof. □

Chapter 3

Estimates in the absolutely continuous case

In this chapter, we prove a quantitative lower bound on $\Lambda(\mu_1)$ in the case when μ_1 is absolutely continuous with bounded density. This provides a measure of control on the main term of the decomposition of $\Lambda(\mu)$ for a general measure μ. We we will restrict to the case when A_j is of the form $A_j = (I_{n\times n}\ B_j)$, where B_j are $n \times (m-n)$ matrices, for reasons that will become clear in Section 3.5. Set $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^{m-n}$, so that our configurations are of the form $\{x + B_1 y, \dots, x + B_k y\}$. With r defined as in (1.4), we will also assume $k - 1 \geq 2r$, or equivalently, $n\lceil (k+1)/2 \rceil \leq m$, the first inequality of condition (1.6) of the main result.

Proposition 3.1. *For every $\delta, M > 0$, there exists a constant $c(\delta, M) > 0$ with the following property: for every function $f : [0,1]^n \to \mathbb{R}$, $0 \leq f \leq M$, $\int f \geq \delta$, we have $\Lambda(f) \geq c(\delta, M)$.*

We will proceed as in the proof of Varnavides' Theorem given in [40]. The strategy will be to decompose $f = g + b$ into a "good" function g which is the major contribution and a "bad" function b whose contribution is negligible. This will be made precise in the following section.

3.1 Preliminaries

Proposition 3.2. *Let f be as in Proposition 3.1. Suppose $f = g + b$ where*

$$\|g\|_\infty, \|b\|_\infty \leq M; \quad \|g\|_1, \|b\|_1 = \delta.$$

Then

$$\Lambda(f) = \Lambda(g) + O(C(M,\delta)\|\widehat{b}\|_\infty).$$

Proof. We use the decomposition $f = g + b$ and the linearity of Λ to decompose $\Lambda(f)$ into 2^k pieces. The main piece will be $\Lambda(g)$ and the remaining pieces which constitute the error term have at least one copy of b. By the hypothesis and Hölder's inequality,

$$\|g\|_2^2, \|b\|_2^2 \leq M\delta.$$

We will apply Lemma 3.3 below to estimate each of the $2^k - 1$ summands in the error term, arriving at an upper bound of $(2^k - 1)\|\widehat{b}\|_\infty (M\delta)^r$. $\square$

We now prove the lemma required for the previous proposition.

Lemma 3.3. *Let f_j be as in Proposition 3.1. Assume moreover that $k-1 \geq 2r$ and $\|f_j\|_1 \leq 1$ for $1 \leq j \leq k$. Then*

$$|\Lambda(f_1, \ldots, f_k)| \leq M\|\widehat{f}_k\|_\infty \|f_r\|_1^{1/2} \|f_{2r}\|_1^{1/2} \prod_{\substack{j=1 \\ j \neq r}}^{2r-1} \|\widehat{f}_j\|_2.$$

We have a similar bound for permutations of $f_1, \ldots, f_k$.

Proof. Let us recall the Fourier representation of Λ from Proposition 2.1, which gives

$$|\Lambda(f_1, \ldots, f_k)| \leq \int_S \prod_{j=1}^{k} |\widehat{f}_j(\xi_j)| \, d\sigma(\xi).$$

Since $\|\widehat{f}_j\|_\infty \leq \|f_j\|_1 \leq 1$ for each j, reducing the number of factors in the product that appears in the last integrand only makes the integral larger. We use the hypothesis $k - 1 \geq 2r$ to drop $(k - 1 - 2r)$ of these factors and split the remaining $2r$ into two groups and apply the Cauchy-Schwarz inequality. Executing these steps leads to

$$\begin{aligned}
&|\Lambda(f_1, \ldots, f_k)| \\
&\leq \int_S \prod_{j=1}^{k} |\widehat{f}_j(\xi_j)| \; d\sigma(\xi) \\
&\leq \|\widehat{f}_k\|_\infty \int_S \prod_{j=1}^{r} |\widehat{f}_j(\xi_j)| \prod_{j=r+1}^{2r} |\widehat{f}_j(\xi_j)| \; d\sigma(\xi) \\
&\leq \|\widehat{f}_k\|_\infty \left(\int_S \prod_{j=1}^{r} |\widehat{f}_j(\xi_j)|^2 \; d\sigma(\xi) \right)^{1/2} \left(\prod_{j=r+1}^{2r} |\widehat{f}_j(\xi_j)|^2 \; d\sigma(\xi) \right)^{1/2}.
\end{aligned}$$

Both of the above integrals are estimated in the same way; we will focus only on the first. If $\xi_j = (\xi_{j,1}, \ldots, \xi_{j,n})$, let $\xi_j' = (\xi_{j,1}, \ldots, \xi_{j,n'})$ with n' as defined in Definition 1.5; notice this is the same as ξ_j^{id} as defined in Proposition 2.6. By Lemma 3.4 below, $\|\widehat{f}_r\|^2_{L^2_{\xi_r'}} \leq M\|f_r\|_1$, and so by Lemma 2.7,

$$\begin{aligned}
&\int_S \prod_{j=1}^{r} |\widehat{f}_j(\xi_j)|^2 \; d\sigma(\xi) \\
&= \int_S \prod_{j=1}^{r} |\widehat{f}_j(\xi_j)|^2 \; d\xi_r' d\xi_1 \cdots d\xi_{r-1} \\
&\leq \int_{\mathbb{R}^{n(r-1)}} M\|f_r\|_1 \prod_{j=1}^{r-1} |\widehat{f}_j(\xi_j)|^2 \; d\xi_1 \cdots d\xi_{r-1} \\
&= M\|f_r\|_1 \prod_{j=1}^{r-1} \|\widehat{f}_j\|_2^2.
\end{aligned}$$

The result follows. □

Lemma 3.4. *Let* $1 \leq j \leq k$. *If* $\xi_j = (\xi_{j,1}, \ldots, \xi_{j,n})$, *we denote* $\xi_j' = (\xi_{j,1}, \ldots, \xi_{j,n'})$ *and* $\xi_j'' = (\xi_{j,n'+1}, \ldots, \xi_{j,k})$. *Suppose*

(a) $|f_j| \leq M$,

(b) $\operatorname{supp} f_j \subseteq [0,1]^n$.

Then

$$\|\widehat{f_j}\|^2_{L^2_{\xi'}} \leq M\|f_j\|_1$$

uniformly for all ξ_j''.

Proof. Fix ξ_j'', and let $F(x', \xi_j'') = \int f_j(x', x'')e^{-2\pi i \xi_j'' \cdot x''}\ dx''$, where x', x'' are the dual variables to ξ_j', ξ_j'' respectively. Then (a) and (b) give $|F(x', \xi_j'')| \leq M$ for all x' so that

$$\|F\|_\infty \leq M.$$

We calculate

$$\|F\|_{L^1_{x'}} = \int |F(x', \xi_j'')|\ dx' \leq \iint |f_j(x', x'')|\ dx'dx'' = \|f_j\|_1.$$

By Hölder's inequality,

$$\|F\|^2_{L^2_{x'}} \leq \|F\|_\infty \|F\|_{L^1_{x'}} \leq M\|f_j\|_1.$$

Now,

$$\widehat{f_j}(\xi_j', \xi_j'') = \int F(x', \xi_j'')e^{-2\pi i \xi_j' \cdot x'}\ dx',$$

which is the Fourier transform of F in x'. Therefore by Plancherel's theorem in the x' variables,

$$\|\widehat{f_j}\|^2_{L^2_{\xi_j'}} = \|F\|^2_{L^2_{x'}} \leq M\|f_j\|_1.$$

$\square$

Before moving on, we shall prove an approximation identity that is useful in the following sections.

Lemma 3.5. *Suppose* $\|f_j\|_\infty, \|g_j\|_\infty \leq C$ *and* $\|f_j - g_j\|_p \leq \kappa$ *for* $1 \leq j \leq R$, *for some* $1 \leq p \leq \infty$. *Then*

$$\left\|\prod_{j=1}^R f_j - \prod_{j=1}^R g_j\right\|_p \leq RC^{R-1}\kappa.$$

Proof.

$$
\begin{aligned}
\left\|\prod_{j=1}^{R} f_j - \prod_{j=1}^{R} g_j\right\|_p &\le \left\|f_1 \cdot \prod_{j=2}^{R} f_j - g_1 \cdot \prod_{j=2}^{R} f_j\right\|_p \\
&\quad + \left\|g_1 \cdot f_2 \cdot \prod_{j=3}^{R} f_j - g_1 \cdot g_2 \cdot \prod_{j=3}^{R} f_j\right\|_p \\
&\quad + \cdots + \left\|\prod_{j=1}^{R-1} g_j \cdot f_R - \prod_{j=1}^{R-1} g_j \cdot g_R\right\|_p \\
&\le \left(\prod_{j=2}^{R} \|f_j\|_\infty\right) \|f_1 - g_1\|_p \\
&\quad + \|g_1\|_\infty \left(\prod_{j=3}^{R} \|f_j\|_\infty\right) \|f_2 - g_2\|_p \\
&\quad + \cdots + \left(\prod_{j=1}^{R-1} \|g_j\|_\infty\right) \|f_R - g_R\|_p \\
&\le RC^{R-1}\kappa.
\end{aligned}
$$

□

3.2 Almost periodic functions

In light of Proposition 3.2, our next goal will be to identify a large class of "good" functions g for which we can bound $\Lambda(g)$ from below. It turns out that almost periodic functions, defined analogously to [40], can be used for this purpose.

Definition 3.6. *1. A* character *is a function $\chi : [0,1]^n \to \mathbb{C}$ of the form $\chi(x) = e^{2\pi i v \cdot x}$ for some $v \in \mathbb{Z}^n$.*

*2. If $K \in \mathbb{N}$, then a K-*quasiperiodic function *is a function f of the form $\sum_{\ell=1}^{K} c_\ell \chi_\ell$ where each χ_ℓ are characters (not necessarily distinct), and c_ℓ are scalars with $|c_\ell| \le 1$.*

3. *If $\sigma > 0$, then $f : [0,1]^n \to \mathbb{C}$ is (σ, K)-*almost periodic *if there exists a K-quasiperiodic function f_{QP} such that $\|f - f_{QP}\|_{L^2([0,1]^n)} \leq \sigma$. We call f_{QP} a* K-quasiperiodic function approximating f within σ.

Lemma 3.7. *Let $K \in \mathbb{N}$, $M > 0$, $0 < \delta < 1$, and*

$$0 < \sigma \leq \frac{\delta^k}{4kM^{k-1}}. \tag{3.1}$$

Then there exists $c(K, \delta, M) > 0$ such that for any non-negative (σ, K)-almost periodic function f bounded by M and obeying $\int f \geq \delta$,

$$\Lambda(f) \geq c(K, \delta, M).$$

Proof. Our goal is to bound $\Lambda(f)$ from below by a multiple of $\|f\|_1^k$, which is known to be at least as large as δ^k. We will achieve this by approximating each factor in the integral defining Λ by f, on a reasonably large set with acceptable error terms.

To this end, let f_{QP} be a K-quasiperiodic function approximating f within σ, say $f_{QP}(x) = \sum_{\ell=1}^{K} c_\ell e^{2\pi i v_\ell \cdot x}$ and $\|f - f_{QP}\|_2 \leq \sigma$. Let $\varepsilon > 0$ be a small constant to be fixed later, and define

$$C_\varepsilon = \{y \in \mathbb{R}^{m-n} : \left|\left|\left| A_j^t v_\ell \cdot y \right|\right|\right| \leq \varepsilon, \text{ for all } 1 \leq j \leq k, 1 \leq \ell \leq K\}, \tag{3.2}$$

where $|||t|||$ denotes the distance of $t \in \mathbb{R}$ to the nearest integer. A pigeonhole argument gives that

$$|C_\varepsilon| \geq c(\varepsilon, K) > 0 \tag{3.3}$$

for some $c(\varepsilon, K)$ possibly dependent on k, m and n but independent of f. This non-trivial result will be shown in Corollary 3.9 following this proof.

Let T^a be the shift map $T^a f(x) := f(x + a)$. For $y \in C_\varepsilon$ and any $x \in [0,1]^n$,

$$|T^{B_j y} f_{QP}(x) - f_{QP}(x)| = \left| \sum_{\ell=1}^{K} c_\ell e^{2\pi i v_\ell \cdot x} (1 - e^{2\pi i v_\ell \cdot A_j y}) \right|$$

$$
\begin{aligned}
&\leq \sum_{\ell=1}^{K} |1 - e^{2\pi i A_j^t v_\ell \cdot y}| \\
&\leq \sum_{\ell=1}^{K} |A_j^t v_\ell \cdot y| \\
&\leq K\varepsilon. \qquad (3.4)
\end{aligned}
$$

The bound above leads to the following estimate:

$$
\begin{aligned}
\|T^{B_j y} f - f\|_{L^1_x} &\leq \|T^{B_j y} f - T^{B_j y} f_{QP}\|_{L^2_x} + \|T^{B_j y} f_{QP} - f_{QP}\|_{L^\infty_x} \\
&\quad + \|f_{QP} - f\|_{L^2_x} \\
&= 2\|f_{QP} - f\|_{L^2_x} + \|T^{B_j y} f_{QP} - f_{QP}\|_{L^\infty_x} \\
&\leq 2\sigma + K\varepsilon \\
&\leq \frac{\delta^k}{2kM^{k-1}} + K\varepsilon.
\end{aligned}
$$

In the sequence of inequalities above, we have used Hölder's inequality in the first step, triangle inequality in the second step, norm-invariance of the shift operator in the third step, and (3.4) in the last step.

We now choose $\varepsilon = \delta^k/(4kM^{k-1})$, so that

$$
\|T^{B_j y} f - f\|_{L^1_x} \leq \frac{3\delta^k}{4kM^{k-1}}. \tag{3.5}
$$

For every $1 \leq j \leq k$, the bound $\|T^{B_j y} f\|_{L^\infty_x} = \|f\|_\infty \leq M$ holds trivially, so by Lemma 3.5 with $C = M$, $f_j = f$, $g_j = T^{B_j y} f$, $R = k$, $p = 1$, and $\kappa = (3\delta^k)/(4kM^{k-1})$, we have

$$
\left\| \prod_{j=1}^{k} T^{B_j y} f - f^k \right\|_{L^1_x} \leq kM^{k-1} \frac{3\delta^k}{4kM^{k-1}} = \frac{3\delta^k}{4}
$$

using (3.5). On the other hand, the bounded non-negativity of f, the hypothesis $\int f \geq \delta$, and Hölder's inequality lead to

$$
\|f^k\|_1 \geq \|f\|_1^k \geq \delta^k,
$$

and so for $y \in C_\varepsilon$,

$$\left\| \prod_{j=1}^{k} T^{B_j y} f \right\|_{L^1_x} \geq \left\| f^k \right\|_1 - \left\| \prod_{j=1}^{k} T^{B_j y} f - f^k \right\|_{L^1_x} \geq \frac{\delta^k}{4}. \tag{3.6}$$

We now combine 3.3, the positivity of f and the above to obtain

$$\begin{aligned}
\Lambda(f) &= \int_{\mathbb{R}^{m-n}} \int_{\mathbb{R}^n} \prod_{j=1}^{k} f_j(x + B_j y)\ dx\ dy \\
&= \int_{\mathbb{R}^{m-n}} \left\| \prod_{j=1}^{k} T^{B_j y} f \right\|_{L^1_x} dy \\
&\geq \int_{C_\varepsilon} \left\| \prod_{j=1}^{k} T^{B_j y} f \right\|_{L^1_x} dy \\
&\geq \delta^k c(\varepsilon, K)/4 \\
&= c(K, \delta, M).
\end{aligned}$$

□

Lemma 3.8. *Given $0 < \varepsilon < 1$ and any integer $K \geq 1$, there exists a positive constant $c'(\varepsilon, K)$ such that*

$$|\{t \in [0,1] : \ \||tv_\ell\|| \leq \varepsilon \textit{ for all } 1 \leq \ell \leq K\}| \geq c'(\varepsilon, K),$$

for any choice of $v_1, \ldots, v_K \in \mathbb{Z}$.

Proof. Clearly it suffices to prove the lemma for the case when $\varepsilon \leq 1$. Let N be the unique integer that $N^{-1} < \varepsilon \leq (N-1)^{-1}$, and consider the partition of the unit cube $[0,1]^K$ into N^K disjoint cubes of side length N^{-1}. That is, the vertices of the cubes are at points of the form $N^{-1}\mathbb{Z}^K$ mod 1. Define

$$X_Q = \{t \in [0,1] : \ (tv_1, \ldots, tv_K) \bmod 1 \in Q\}.$$

Then since $|[0,1]| = 1$, there must exist a cube Q such that

$$|X_Q| \geq \frac{1}{N^K} \geq \left(\frac{\varepsilon}{2}\right)^K,$$

where we have used that $\varepsilon \leq (n-1)^{-1}$. For $t \in X_Q$,

$$(tv_1, \ldots, tv_K) \bmod 1 \in Q - Q \subseteq [-n^{-1}, n^{-1}]^K \subseteq [-\varepsilon, \varepsilon]^K,$$

and so $|||tv_\ell||| \leq \varepsilon$ for every $1 \leq \ell \leq K$. Notice

$$|X_Q - X_Q| \geq |X_Q| \geq \left(\frac{\varepsilon}{2}\right)^K,$$

and so by symmetry

$$|\{t \in [0,1] : |||tv_\ell||| \leq \varepsilon \text{ for all } 1 \leq \ell \leq K\}| \geq \frac{1}{2}\left(\frac{\varepsilon}{2}\right)^K.$$

$\square$

Corollary 3.9. *Given $0 < \varepsilon < 1$, and integers $k, K, m, n \in \mathbb{N}$, $m > n$, there exists a positive constant c depending on all of these quantities, for which the set*

$$C_\varepsilon = \{y \in \mathbb{R}^{m-n} : \left|\left|\left|A_j^t v_\ell \cdot y\right|\right|\right| \leq \varepsilon, \textit{ for all } 1 \leq j \leq k, 1 \leq \ell \leq K\}.$$

defined as in Lemma 3.7 obeys the size estimate

$$|C_\varepsilon| \geq c$$

for any choice of matrices $\{A_j\}$ and vectors $\{v_\ell\}$.

Proof. Let $A_j^t v_\ell(i)$ denote the ith component of $A_j^t v_\ell$. Let

$$D_i = \{y_i \in [0,1] : \left|\left|\left|A_j^t v_\ell(i) y_i\right|\right|\right| \leq \varepsilon/(m-n) \text{ for all } 1 \leq j \leq k, 1 \leq \ell \leq K\}.$$

By Lemma 3.8,

$$|D_i| \geq c'(\varepsilon/(m-n), K)^k$$

for $1 \leq i \leq m-n$. If $y = (y_1, \ldots, y_{m-n}) \in \prod_{i=1}^{m-n} D_i$, then

$$|||A_j^t v_\ell \cdot y||| = \left|\left|\left|\sum_{i=1}^{m-n} A_j^t v_\ell(i) y_i\right|\right|\right| \leq (m-n)\frac{\varepsilon}{m-n} = \varepsilon,$$

and so

$$C_\varepsilon \supseteq D_1 \times \cdots D_{m-n}.$$

Therefore,

$$|C_\varepsilon| \geq c'(\varepsilon/(m-n), K)^{k(m-n)} = c(\varepsilon, K).$$

□

3.3 Ubiquity of almost periodic functions

To make use of Lemma 3.7, we will approximate a general function f by an almost periodic function. In the following sequence of lemmas, we construct an increasingly larger family of σ-algebras with the property that any function measurable with respect to these will be almost periodic. We do this by an iterative random mechanism, the building block of which is summarized in the next result.

Lemma 3.10. *Let $0 < \varepsilon \ll 1$ and let χ be a character. Viewing $\mathbb{C}$ as $\mathbb{R}^2$, partition the complex plane $\mathbb{C} = \bigcup_{Q \in \mathbb{Q}_\varepsilon} Q$ into squares of side-length ε with corners lying in the lattice $\varepsilon\mathbb{Z}^2$. For $\omega \in [0,1]^2$, define $\mathcal{B}_{\varepsilon,\chi,\omega}$ to be the σ-algebra generated by the atoms*

$$\{\chi^{-1}(Q + \varepsilon\omega) : Q \in \mathbb{Q}_\varepsilon\}.$$

There exists ω such that

1. *$\|\chi - \mathbb{E}(\chi|\mathcal{B}_{\varepsilon,\chi,\omega})\|_\infty \leq C\varepsilon$.*

2. *For every $\sigma > 0$ and $M < 0$, there exists $K = K(\sigma, \varepsilon, M)$ such that every function f which is measurable with respect to $\mathcal{B}_{\varepsilon,\chi,\omega}$ with $\|f\|_\infty \leq M$ is (σ, K)-almost periodic.*

Proof. (1) follows from definition of $\mathcal{B}_{\varepsilon,\chi,\omega}$, for any ω.

To prove (2), it suffices to prove: for each integer $\ell > \ell_0$ for some sufficiently large ℓ_0, there exists a set $\Omega_\ell \subseteq [0,1]^2$, $|\Omega_\ell| > 1 - C2^{-\ell}\varepsilon$ with the following property. For $\omega \in \Omega_\ell$, there exists $K = K(\sigma, \varepsilon, M)$ such that every function f which is measurable with respect to $\mathcal{B}_{\varepsilon,\chi,\omega}$ with $\|f\|_\infty \le M$ is (σ, K)-almost periodic, for $\sigma = 2^{-\ell}$.

Indeed, taking $\Omega = \bigcap_{\ell > \ell_0} \Omega_\ell$, we have

$$|\Omega| > 1 - C \sum_{\ell > \ell_0} 2^{-\ell}\varepsilon \ge 1 - C2^{-\ell_0}\varepsilon$$

and so we may find $\omega \in \Omega$. Then if $\sigma > 0$, get $\ell > \ell_0$ with $2^{-\ell} \le \sigma$, and by the above, there exists $K = K(2^{-\ell}, \varepsilon, M)$ such that every function f which is measurable with respect to $\mathcal{B}_{\varepsilon,\chi,\omega}$ with $\|f\|_\infty \le M$ is $(2^{-\ell}, K)$-almost periodic. Thus, we fix $\sigma = 2^{-\ell}$.

Let N be the number of atoms of $\mathcal{B}_{\varepsilon,\chi,\omega}$. Recall that the atoms are of the form $\chi^{-1}(Q + \varepsilon\omega)$ for $Q \in \mathbb{Q}_\varepsilon$. Since the image of χ lies in the unit circle $\mathbb{S}^1$, the preimage of $(Q + \varepsilon\omega)$ under χ is non-empty if and only if $(Q + \varepsilon\omega) \cap \mathbb{S}^1 \ne \varnothing$. Since $\operatorname{diam}(Q + \varepsilon\omega) = \sqrt{2}\varepsilon$, this holds if and only if $(Q + \varepsilon\omega)$ lies in the $(\sqrt{2}\varepsilon)$-thickened unit circle,

$$\mathbb{S}^1_{[\sqrt{2}\varepsilon]} = \{x \in \mathbb{C} : \operatorname{dist}(x, \mathbb{S}^1) \le \sqrt{2}\varepsilon\}.$$

It is easy to calculate $|\mathbb{S}^1_{[\sqrt{2}\varepsilon]}| = \pi 4\sqrt{2}\varepsilon$ and $|Q + \varepsilon\omega| = \varepsilon^2$ for every $Q \in \mathbb{Q}_\varepsilon$, so since the squares $Q + \varepsilon\omega$ are disjoint,

$$N \le \frac{|\mathbb{S}^1_{\sqrt{2}\varepsilon}|}{|Q + \varepsilon\omega|} = \frac{\pi 4\sqrt{2}}{\varepsilon}.$$

In particular, $\mathcal{B}_{\varepsilon,\chi,\omega}$ has at most $C\varepsilon^{-1}$ atoms; we use this fact to reduce the proof of Lemma 3.10 to a simpler form. Namely, it suffices to prove that for f an indicator function of one of those atoms, say

$$f(x) = f_{Q,\omega}(x) := \mathbb{1}_{\chi^{-1}(Q+\varepsilon\omega)}(x) = \mathbb{1}_Q(\chi(x) - \varepsilon\omega),$$

there exists a $C(\sigma,\varepsilon)$-quasiperiodic function $g_{Q,\omega}$ such that $\|f_{Q,\omega}-g_{Q,\omega}\|_2 \leq C^{-1}\sigma\varepsilon$ with probability $1-C\sigma\varepsilon^{-1}$. (Here and below, $C(\sigma,\varepsilon)$ will denote a constant which may change from line to line, but always depends only on M,σ,ε, in particular remains independent of f.) Indeed, if this were the case, then any measurable f may be written as

$$f(x)=\sum_{i\in I} c_i f_{Q_i,\omega}(x)$$

where $f_{Q_i,\omega}(c)=\mathbb{1}_{Q_i}(\chi(x)-\varepsilon\omega)(x)$, $Q_i\in\mathbb{Q}_\varepsilon$ are distinct, and $\#I\leq C\varepsilon^{-1}$. Notice $\|f\|_\infty \leq M$ and the fact that $f_{Q_i,\omega}$ have disjoint support means $|c_i|\leq M$ for $i\in I$. Letting $g_{Q_i,\omega}$ be a $C(\sigma/M,\varepsilon)$-quasiperiodic function approximating $f_{Q_i,\omega}$ to within $C^{-1}M^{-1}\sigma\varepsilon$, we have that $g=\sum_{i\in I}c_i g_{Q_i,\omega}$ is $C(\sigma,\varepsilon)$-quasiperiodic (repeating $g_{Q_i,\omega}$ at most M times if necessary) and

$$\|f-g\|_2\leq\sum_{i\in I}|c_i|\cdot\|f_{Q_i,\omega}-g_{Q_i,\omega}\|_2\leq\sigma.$$

Thus we restrict to the case when f is the indicator function of one of those atoms, which we denote f_ω for ease.

$$f(x)=f_\omega(x):=\mathbb{1}_Q(\chi(x)-\varepsilon\omega).$$

Let $0\leq h\leq 1$ be a continuous function on $[-2,2]^2$ which is a good L^2-approximation for $\mathbb{1}_Q$; precisely,

$$\|\mathbb{1}_Q-h\|^2_{L^2([-2,2]^2)}<\frac{\sigma^3\varepsilon^3}{10CM^2}. \tag{3.7}$$

By the Weierstrass Approximation Theorem, there exists a polynomial P such that

$$\|h-P\|_{L^\infty([-2,2]^2)}<\left(\frac{\sigma^3\varepsilon^3}{10CM^2}\right)^{1/2}. \tag{3.8}$$

Let $h_\omega(x):=h(\chi(x)-\varepsilon\omega)$, and $g_\omega(x)=P(\chi(x)-\varepsilon\omega)$; notice g_ω can be written as a linear combination of at most $C(\sigma,\varepsilon)$ characters, with coefficients at most $C(\sigma,\varepsilon)$. Repeating characters if necessary, we may reduce the coefficients to be less than 1 and so g_ω is $C(\sigma,\varepsilon)$-quasiperiodic. It should also be noted that $\|g_\omega\|_\infty \le ||g_\omega - h_\omega||_\infty + ||h_\omega||_\infty \le 2$.

It remains to show $\|f_\omega - g_\omega\|_2 \le C^{-1}M^{-1}\sigma\varepsilon$ with probability at least $1 - C\sigma\varepsilon^{-1}$. Define

$$F(\omega) = \|f_\omega - g_\omega\|_2^2 = \int_{[0,1]^n} |f_\omega(x) - g_\omega(x)|^2 \; dx.$$

By an application of Cauchy-Schwarz inequality on the integrand, combined with (3.7) and Tonelli's Theorem, we obtain

$$\begin{aligned}
\|F\|_{L^1_\omega} &= \int_{[0,1]^2}\int_{[0,1]^n} |f_\omega(x) - g_\omega(x)|^2 \; dx \; d\omega \\
&\le 2\int_{[0,1]^2}\int_{[0,1]^n} \left[|f_\omega(x) - h_\omega(x)|^2 + |h_\omega(x) - g_\omega(x)|^2\right] dx\, d\omega \\
&< \frac{\sigma^3\varepsilon^3}{2CM^2} + \int_{[0,1]^n}\int_{[0,1]^2} |h_\omega(x) - g_\omega(x)|^2 \; d\omega \; dx \\
&< \frac{\sigma^3\varepsilon^3}{2CM^2} + \int_{[0,1]^n}\int_{[0,1]^2} |h(\chi(x) - \varepsilon\omega) - P(\chi(x) - \varepsilon\omega)|^2 \; d\omega \; dx.
\end{aligned}$$

Using the change of variables $\omega' = \chi(x) - \varepsilon\omega$ for fixed $x \in [0,1]^n$, we have $d\omega' = \varepsilon^2 \; d\omega$. Notice ω' belongs to $[0,\varepsilon]^2$ shifted by $\chi(x)$, so is contained in $[-2,2]^2$. By (3.8),

$$\begin{aligned}
\|F\|_1 &\le \frac{\sigma^3\varepsilon^3}{2CM^2} + \int_{[0,1]^n}\int_{\omega'\in[-2,2]^2} |h(\omega') - P(\omega')|^2 \; \frac{d\omega'}{\varepsilon^2} \; dx \\
&\le \frac{\sigma^3\varepsilon^3}{2CM^2} + \int_{[0,1]^n}\int_{\omega'\in[-2,2]^2} |h(\omega') - P(\omega')|^2 \; \frac{d\omega'}{\varepsilon^2} \; dx \\
&= \frac{\sigma^3\varepsilon^3}{2CM^2} + \int_{[0,1]^n} \frac{1}{\varepsilon^2}\|h - P\|^2_{L^2([-2,2]^2)} \; dx \\
&\le \frac{\sigma^3\varepsilon^3}{2CM^2} + \int_{[0,1]^n} \frac{1}{\varepsilon^2}\left(\frac{\sigma^3\varepsilon^3}{10CM^2}\right) \; dx \\
&< \frac{1}{\varepsilon^2}\left(\frac{\sigma^3\varepsilon^3}{CM^2}\right).
\end{aligned}$$

By Markov's inequality,

$$\begin{aligned}
&|\{\omega \in [0,1]^2 : \|f_\omega - g_\omega\|_2^2 > C^{-2}M^{-2}\sigma^2\varepsilon^2\}| \\
&= |\{\omega \in [0,1]^2 : F(\omega) > C^{-2}M^{-2}\sigma^2\varepsilon^2\}| \\
&\leq \frac{C^2M^2\|F\|_1}{\sigma^2\varepsilon^2} \leq \frac{C^2M^2\frac{1}{\varepsilon^2}\left(\frac{\sigma^3\varepsilon^3}{CM^2}\right)}{\sigma^2\varepsilon^2} = C\sigma\varepsilon^{-1}.
\end{aligned}$$

Thus,

$$|\{\omega \in [0,1]^2 : \|f_\omega - g_\omega\|_2^2 \leq C^{-1}M^{-1}\sigma\varepsilon\}| \geq 1 - C\sigma\varepsilon^{-1},$$

as required. □

The above proof also gives the following result.

Corollary 3.11. *Let $0 < \varepsilon \ll 1$ and let χ be a character. Then the σ-algebra $\mathcal{B}_{\varepsilon,\chi,\omega}$ described in the statement of Lemma 3.10 can be chosen to have the additional property that for every atom $\chi^{-1}(Q + \varepsilon\omega)$, $Q \in \mathbb{Q}_\varepsilon$, there exists a K-quasiperiodic function $g_{Q,\omega}$ that obeys $\|g_{Q,\omega}(\cdot) - \mathbb{1}_Q(\chi(\cdot) - \varepsilon\omega)\|_2 < \sigma$ for every $\sigma > 0$, and in addition $\|g_{Q,\omega}\|_\infty \leq 2$.*

We can concatenate the σ-algebras from Lemma 3.10. If $\mathcal{B}_1, \ldots, \mathcal{B}_R$ are σ-algebras, denote by $\mathcal{B}_1 \vee \cdots \vee \mathcal{B}_R$ the smallest σ-algebra which contains all of them.

Corollary 3.12. *Let $0 < \varepsilon_1, \ldots, \varepsilon_R \ll 1$ and let $\chi_1, \ldots, \chi_R$ be characters. Let $\mathcal{B}_{\varepsilon_1,\chi_1}, \ldots, \mathcal{B}_{\varepsilon_R,\chi_R}$ be the σ-algebras arising from Lemma 3.10. Then for every $\sigma > 0$, there exists $K = K(R, \sigma, \varepsilon_1, \ldots, \varepsilon_R)$ such that every function f which is measurable with respect to $\mathcal{B}_{\varepsilon_1,\chi_1} \vee \cdots \vee \mathcal{B}_{\varepsilon_R,\chi_R}$ with $\|f\|_\infty \leq M$ is (σ, K)-almost periodic.*

Proof. Since there are at most $C(R, \varepsilon_1, \ldots, \varepsilon_R)$ atoms in $\mathcal{B}_{\varepsilon_1,\chi_1} \vee \cdots \vee \mathcal{B}_{\varepsilon_R,\chi_R}$, it suffices to prove the claim in the case when f is the indicator function of a single atom. Then f is the product of R indicator functions $f_1, \ldots, f_R$, where f_j is the indicator function of an atom from $\mathcal{B}_{\varepsilon_j,\chi_j}$. Let g_j be a $K(\sigma/(R2^{R-1}), \varepsilon_j)$-quasiperiodic function approximating f_j to within

$\sigma/(R2^{R-1})$ as provided in Corollary 3.12; notice $\|g_j\|_\infty \leq 2$. Then $g = \prod g_j$ is a K-quasiperiodic function where $K = \prod K(\sigma/(R2^{R-1}), \varepsilon_j)$ depends only on $R, \sigma, \varepsilon_1, \ldots, \varepsilon_R$. Finally, by Lemma 3.5 with $C = 2$, $p = 2$, and $\kappa = \sigma/(R2^{R-1})$, we have

$$\left\| \prod_{j=1}^{R} f_j - \prod_{j=1}^{R} g_j \right\|_2 \leq \sigma.$$

□

3.4 Proof of Proposition 3.1.

We will need two more auxiliary results, analogous to [40, Lemma 2.10 and 2.11].

Lemma 3.13. *Let b be a function bounded by M with $\|\widehat{b}\|_\infty \geq \sigma > 0$. Then there exists $0 < \varepsilon \ll \sigma$, a character χ, and an associated σ-algebra $\mathcal{B}_{\varepsilon,\chi}$ (as defined earlier in this section) such that*

$$\|\mathbb{E}(b|\mathcal{B}_{\varepsilon,\chi})\|_2 \geq C^{-1}\sigma.$$

Proof. Since $\|\widehat{b}\|_\infty \geq \sigma$, there exists a character χ such that

$$\left| \int_{[0,1]^n} b(x)\chi(x)\,dx \right| \geq \frac{\sigma}{2}. \tag{3.9}$$

On the other hand, the σ-algebra $\mathcal{B}_{\varepsilon,\chi}$ is generated by the atoms $\{\chi^{-1}(Q + \varepsilon\omega); Q \in \mathbb{Q}_\varepsilon\}$ for some ω in the unit square. On each atom, χ can vary by at most $C\varepsilon$, hence

$$\|\chi - \mathbb{E}(\chi|\mathcal{B}_{\varepsilon,\chi})\|_\infty \leq C\varepsilon. \tag{3.10}$$

Since b is bounded by M, (3.9) and (3.10) yield

$$\int_{[0,1]^n} b(x)\mathbb{E}(\chi|\mathcal{B}_{\varepsilon,\chi})(x)\,dx \geq \frac{\sigma}{2} - MC\varepsilon.$$

Conditional expectation being self-adjoint, the inequality above may be rewritten as

$$\int_{[0,1]^n} \mathbb{E}(b|\mathcal{B}_{\varepsilon,\chi})(x)\chi(x)\,dx \geq \frac{\sigma}{2} - MC\varepsilon.$$

Recalling that χ is bounded above by 1, the desired result now follows by choosing ε sufficiently small relative to σ and M, and applying Cauchy-Schwarz inequality to the integral on the left. $\square$

Lemma 3.14. *Let $F : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ be an arbitrary function, let $0 < \delta \leq 1$, and let $f \geq 0$ be a function bounded by M with $\int f \geq \delta$. Let σ satisfy (3.1). Then there exists a K with $0 < K \leq C(F, \delta)$ and a decomposition $f = g + b$ where $g \geq 0$ is a bounded (σ, K)-almost periodic function with $\int g \geq \delta$, and b obeys the bound*

$$\|\widehat{b}\|_\infty \leq F(\delta, K). \tag{3.11}$$

The proof of Lemma 3.14 is exactly identical to [40, 2.11].

Proof of Proposition 3.0.8. Let $F : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ be a function to be chosen later. Decompose $f = g + b$ as in Lemma 3.14. By Lemma 3.7,

$$\Lambda(g) \geq c(K, \delta, M).$$

By Proposition 3.2, (3.11), and the above inequality,

$$\Lambda(f) \geq c(K, \delta, M) + O(C(M, \delta)F(\delta, K)).$$

By choosing F sufficiently small and since $K \leq C(F, \delta)$, we get

$$\Lambda(f) \geq c(\delta, M)$$

as required. $\square$

3.5 Quantitative Szemerédi bounds fail for general $\mathbb{A}$

At the beginning of this chapter, we restricted to the case when A_j is of the form $A_j = (I_{n\times n}\ B_j)$ where B_j are $n \times (m-n)$ matrices. The reason for this is that generic A_i will not provide a lower bound on Λ when $\int f = \delta$, even when satisfying the non-degeneracy condition.

In the case $n = 2, k = 3, m = 4$, consider the function $f = \mathbb{1}_{B((0,1),\delta)}$, the indicator of the ball centered at $(0, 1)$ with radius $\delta \leq 1/3$. Define

$$A_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

$$A_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$A_3 = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

It is clear that (2.10) holds for these matrices. In the integral defining Λ, we consider the conditions for $\vec{x} = (x_1, \ldots, x_4)$ to be in the support of $\prod_{i=1}^{3} f(A_i\vec{x})$. The first term of the product gives $f(A_1\vec{x}) = f(x_1, x_2)$, and so in particular, $|x_2 - 1| < \delta$ which implies

$$|x_2| > 1 - \delta. \tag{3.12}$$

Similarly, considering the second term yields in particular

$$|x_3| < \delta, \tag{3.13}$$

while the third term gives

$$|x_2 + x_3| < \delta. \tag{3.14}$$

On the other hand, (3.12) and (3.13) give

$$|x_2 + x_3| \geq |x_2| - |x_3| > 1 - 2\delta \geq \delta.$$

Then the support of $\prod_{i=1}^{4} f(A_i\vec{x})$ is empty, and $\Lambda = 0$.

Chapter 4

Proof of the main result

The preceding section gave a quantitative lower bound on the Λ quantity in the case of absolutely continuous measures with bounded density. This suggests the strategy of decomposing the measure μ as $\mu = \mu_1 + \mu_2$ where μ_1 is absolutely continuous with bounded density, and μ_2 gives negligible contribution. In light of the Fourier form of Λ, the key property of μ_2 here will be having good bounds on the Fourier transform.

Let $\phi \in \mathcal{S}(\mathbb{R}^n)$ be a non-negative function supported on $B(0,1)$ with $\int \phi = 1$. For any positive integer N, define $\phi_N(x) = N^n \phi(Nx)$. Let $N \gg 1$ be a large constant to be determined later, and let

$$\mu_1(x) = \mu * \phi_N(x).$$

Clearly, $\mu_1 \geq 0$ is a C^∞ function of compact support with $\int d\mu_1 = 1$. Since ϕ_N is supported on $B(0, N^{-1})$,

$$\begin{aligned}
|\mu_1(x)| &\leq \int_{B(x,N^{-1})} |\phi_N(x-y)| \, d\mu(y) \\
&= \int_{B(x,N^{-1})} N^n |\phi(N(x-y))| \, d\mu(y) \\
&\leq C N^n \mu(B(x, N^{-1})) \\
&\leq C N^{n-\alpha}
\end{aligned}$$

where the last inequality follows by the ball condition (a). Then $|\mu_1(x)| \leq M = Ce$ if $N = e^{1/(n-\alpha)}$, which tends to infinity as $\alpha \to n^-$.

Focusing now on μ_2, we will prove that

$$\left|\widehat{\mu}_2(\xi)\right| \lesssim N^{-\frac{\varepsilon\beta}{2}}(1+|\xi|)^{-\frac{\beta}{2}(1-\varepsilon)} \tag{4.1}$$

for some constant $\varepsilon > 0$ to be chosen later. Since $\int \phi = 1$ and $\phi \in \mathcal{S}(\mathbb{R}^n)$,

$$|1-\widehat{\phi}(\xi)| = |\widehat{\phi}(0)-\widehat{\phi}(\xi)| = \left|\int_0^1 \frac{d}{dt}\widehat{\phi}(t\xi)\ dt\right| \leq \int_0^1 |\xi \cdot \nabla\widehat{\phi}|\ dt \leq C|\xi|.$$

In particular, defining $\mu_2 = \mu - \mu_1$ we have

$$|\widehat{\mu}_2(\xi)| \lesssim |\widehat{\mu}(\xi)| \min(1, |\xi|N^{-1}).$$

Notice if $|\xi| \geq N$, then

$$\begin{aligned}
|\widehat{\mu}_2(\xi)| &\leq |\widehat{\mu}(\xi)| \\
&\lesssim (1+|\xi|)^{-\beta/2} \\
&= (1+|\xi|)^{-\varepsilon\beta/2}(1+|\xi|)^{-\beta/2(1-\varepsilon)} \\
&\lesssim N^{-\varepsilon\beta/2}(1+|\xi|)^{-\beta/2(1-\varepsilon)}.
\end{aligned}$$

On the other hand, if $|\xi| < N$, then we still have

$$\begin{aligned}
|\widehat{\mu}_2(\xi)| &\leq |\widehat{\mu}(\xi)| \\
&\lesssim (1+|\xi|)^{-\beta/2}|\xi|N^{-1} \\
&= (1+|\xi|)^{-\beta/2}|\xi|^{\varepsilon\beta/2}|\xi|^{1-\varepsilon\beta/2}N^{-1} \\
&\lesssim N^{-\varepsilon\beta/2}(1+|\xi|)^{-\beta/2(1-\varepsilon)}.
\end{aligned}$$

Now, decompose

$$\Lambda^*(\widehat{\mu}) = \Lambda^*(\widehat{\mu_1}) + \Lambda(\widehat{\mu_2}, \widehat{\mu_1}, \ldots, \widehat{\mu_1}) + \cdots + \Lambda(\widehat{\mu_2}).$$

By Proposition 3.1, $\Lambda^*(\widehat{\mu_1}) = \Lambda(\mu_1) > c(\delta, M)$. It remains to show the

Λ^* quantities containing at least one copy of μ_2 are negligible relative to $c(\delta, Ce)$. These quantities can be written as $\Lambda^*(g_1, \ldots, g_k)$ where for each $1 \leq j \leq k$, g_j is either $\widehat{\mu_1}$ or $\widehat{\mu_2}$ and at least one g_j is $\widehat{\mu_2}$. Without loss of generality, suppose $g_1 = \widehat{\mu}_2$, so that

$$|g_1(\eta_1)| \lesssim N^{-\varepsilon\beta/2}(1 + |\eta_j|)^{-\beta/2(1-\varepsilon)}$$

by the above estimate on $\widehat{\mu}_2$. For $j \geq 2$, we have

$$|g_j(\eta_j)| \lesssim (1 + |\eta_j|)^{-\beta/2(1-\varepsilon)}$$

by the above estimate on $\widehat{\mu}_2$ and the general Fourier decay condition (b) on μ. Then

$$\begin{aligned}\Lambda^*(g_1, \ldots, g_k) &= \int_S \prod_{j=1}^{k} g_j(\eta_j)\ d\sigma \\ &\leq N^{-\varepsilon\beta/2} \int_S \prod_{j=1}^{k} (1 + |\eta_j|)^{-\beta/2(1-\varepsilon)}\ d\sigma.\end{aligned}$$

Since $\beta > 2(nk - m)/k$, we may choose $\varepsilon > 0$ so that $\beta' = \beta(1 - \varepsilon) > 2(nk - m)/k$. Then by Proposition 2.6 with β' in place of β, the integral above is bounded by a constant independent of N. Then we may choose N sufficiently large that $\Lambda^*(g_1, \ldots, g_k) \leq 2^{-k}c(\delta, M)$, and so

$$\Lambda^*(\widehat{\mu}) \geq 2^{-k}c(\delta, M).$$

The result follows by Proposition 2.8.

Chapter 5

Examples

For a fixed choice of $n \geq 1$ and $k \geq 3$, let m be a multiple of n that satisfies (1.6). Non-degeneracy in the sense of Definition 1.15 in this case will be the condition

$$\operatorname{rank} \begin{pmatrix} A_{i_1} \\ \vdots \\ A_{i_{m/n}} \end{pmatrix} = \operatorname{rank} \begin{pmatrix} I_{n\times n} & B_{i_1} \\ \vdots & \vdots \\ I_{n\times n} & B_{i_{m/n}} \end{pmatrix} = m,$$

for $i_1, \ldots, i_{m/n} \in \{1, \ldots, k\}$ distinct. Reducing,

$$\operatorname{rank} \begin{pmatrix} I_{n\times n} & B_{i_1} \\ 0_{n\times n} & B_{i_2} - B_{i_1} \\ \vdots & \vdots \\ 0_{n\times n} & B_{m/n} - B_{i_1} \end{pmatrix} = m.$$

Since $I_{n\times n}$ is of rank n, it suffices for

$$\operatorname{rank} \begin{pmatrix} B_{i_2} - B_{i_1} \\ \vdots \\ B_{i_{m/n}} - B_{i_1} \end{pmatrix} = m - n, \tag{5.1}$$

for $i_1, \ldots, i_{m/n} \in \{1, \ldots, k\}$ distinct. Notice that while it is necessary to check (5.1) for every choice of m/n indices $i_1, \ldots, i_{m/n}$, we do not need to check for permutations of the indices—any permutation suffices.

Example 5.1 (Triangles). We now prove the claim in Corollary 1.18 that if a, b, c are three distinct points in the plane, then any set $E \subset \mathbb{R}^2$ obeying the assumptions of Theorem 1.16 with ε_0 small enough (depending on C and on a, b, c) must contain a similar copy of the triangle $\triangle abc$. Note that our proof allows for degenerate triangles where a, b, c are collinear.

Let θ be the angle between the line segments $\overline{ab}$ and $\overline{ac}$, measured counter-clockwise, and let $\lambda = \frac{|c-a|}{|b-a|}$. Permuting the points a, b, c if necessary, we may assume without loss of generality that $\theta \in (0, \pi]$. Then it suffices to prove that E contains a configuration of the form

$$x, \quad x+y, \quad x+\lambda y_\theta, \tag{5.2}$$

where y_θ is the vector y rotated by an angle θ counter-clockwise, for some $x, y \in \mathbb{R}^2$ with $y \neq 0$.

Fix $n = 2$, $k = 3$, and $m = 4$. Let $B_1 = 0_{2\times 2}$, By (5.1), non-degeneracy means that

$$\operatorname{rank}(B_j) = 2, \qquad \operatorname{rank}\Big(B_3 - B_2\Big) = 2,$$

for $j = 2, 3$. With $\theta \in (0, \pi]$ and $\lambda > 0$ as above, let

$$B_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad B_3 = \begin{pmatrix} \lambda\cos\theta & -\lambda\sin\theta \\ \lambda\sin\theta & \lambda\cos\theta \end{pmatrix}.$$

It is easy to check that non-degeneracy holds, and this collection of matrices corresponds to configurations of the form (5.2). Letting $V = \{0\}$, Theorem 1.16 asserts that any set $E \subset \mathbb{R}^2$ obeying its assumptions with ε_0 small enough must contain such a configuration, non-degenerate in the sense that $y \neq 0$. This proves Corollary 1.18.

Example 5.2 (Collinear triples). We prove that if a, b, c are three distinct collinear points in $\mathbb{R}^n$, then any set $E \subset \mathbb{R}^n$ obeying the assumptions of Theorem 1.16 with ε_0 small enough (depending on C and on a, b, c) must contain a non-degenerate similar copy of $\{a, b, c\}$.

Without loss of generality, suppose $|c-a| > |b-a|$ Let $\lambda = \frac{|c-a|}{|b-a|} > 1$. Then it suffices to prove that E contains a configuration of the form

$$x, \quad x+y, \quad x+\lambda y, \tag{5.3}$$

for some $x, y \in \mathbb{R}^n$ with $y \neq 0$.

Fix a positive integer n, $k = 3$, and $m = 2n$. Let $B_1 = 0_{n\times n}$, $B_2 = I_{n\times n}$, $B_3 = \lambda I_{n\times n}$. Similarly to Example 5.1, this system of matrices produces configurations of the form (5.3), and the non-degeneracy condition (5.1) becomes

$$\operatorname{rank}(B_j) = n, \qquad \operatorname{rank}\Big(B_3 - B_2\Big) = n,$$

for $j = 2, 3$, which is easy to check for B_j as above. Applying Theorem 1.16 with $V = \{0\}$ as before, we get the desired conclusion.

Example 5.3 (**Parallelograms).** We now prove Corollary 1.19, by proving a more general result.

Corollary 5.4. *Suppose that $E \subset \mathbb{R}^n$ satisfies the assumptions of Theorem 1.16, with ε_0 sufficiently small depending on C. Then E contains a configuration of the form $\{x, x+y_1, \ldots, x+y_{k-2}, x+y_1+\cdots+y_{k-2}\}$, where the k points are all distinct and $x, y_1, \ldots, y_{k-2} \in \mathbb{R}^n$.*

Note that there is no relationship imposed on k and n. Geometrically, such configurations can be viewed as corresponding to k of the vertices of a $(k-2)$-parallelotope sitting in n dimensions; an "origin" vertex, $(k-2)$ "generator" vertices, and the vertex diagonal from the origin vertex. In particular, parallelograms are the special case when $k = 4$.

Proof. Fix $n \geq 1$, $k \geq 4$, and $m = (k-1)n$. Let

$$\begin{aligned}
B_1 &= \begin{pmatrix} 0_{n\times n} & 0_{n\times n} & 0_{n\times n} & \cdots & 0_{n\times n} \end{pmatrix}, \\
B_2 &= \begin{pmatrix} I_{n\times n} & 0_{n\times n} & 0_{n\times n} & \cdots & 0_{n\times n} \end{pmatrix}, \\
B_3 &= \begin{pmatrix} 0_{n\times n} & I_{n\times n} & 0_{n\times n} & \cdots & 0_{n\times n} \end{pmatrix}, \\
&\vdots
\end{aligned}$$

$$\begin{aligned} B_{k-1} &= \begin{pmatrix} 0_{n\times n} & 0_{n\times n} & 0_{n\times n} & \cdots & I_{n\times n} \end{pmatrix}, \\ B_k &= B_2 + B_3 + \cdots + B_{k-1} \\ &= \begin{pmatrix} I_{n\times n} & I_{n\times n} & I_{n\times n} & \cdots & I_{n\times n} \end{pmatrix}. \end{aligned}$$

(5.1) tells us non-degeneracy will be the condition

$$\operatorname{rank}\begin{pmatrix} B_{i_1} \\ B_{i_2} \\ \vdots \\ B_{i_{k-2}} \end{pmatrix} = (k-2)n, \qquad \operatorname{rank}\begin{pmatrix} B_2 - B_k \\ B_3 - B_k \\ \vdots \\ B_{k-1} - B_k \end{pmatrix} = (k-2)n$$

for $i_1, i_2, \ldots, i_{k-2} \in \{2, 3, 4, \ldots, k\}$ distinct.

If $i_1, i_2, \ldots, i_{k-2} \in \{2, 3, 4, \ldots, k-1\}$, then by rearranging,

$$\operatorname{rank}\begin{pmatrix} B_{i_1} \\ B_{i_2} \\ \vdots \\ B_{i_{k-2}} \end{pmatrix} = \operatorname{rank}\begin{pmatrix} I_{n\times n} & 0_{n\times n} & \cdots & 0_{n\times n} \\ 0_{n\times n} & I_{n\times n} & \cdots & 0_{n\times n} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{n\times n} & 0_{n\times n} & \cdots & I_{n\times n} \end{pmatrix} = (k-2)n.$$

If one of the indices is k, say $i_{k-2} = k$, then replacing $B_{i_{k-2}}$ with $B_{i_{k-2}} - (B_{i_1} + \cdots + B_{i_{k-3}})$ reduces the matrix to the previous case as rank is invariant under elementary row operations. Finally, it is easy to see that

$$\operatorname{rank}\begin{pmatrix} B_2 - B_k \\ B_3 - B_k \\ \vdots \\ B_{k-1} - B_k \end{pmatrix} = \operatorname{rank}\begin{pmatrix} 0_{n\times n} & I_{n\times n} & \cdots & I_{n\times n} \\ I_{n\times n} & 0_{n\times n} & \cdots & I_{n\times n} \\ \vdots & \vdots & \ddots & \vdots \\ I_{n\times n} & I_{n\times n} & \cdots & 0_{n\times n} \end{pmatrix} = (k-2)n.$$

In this example, we may consider degeneracy to be the case where two or more points coincide; the exceptional subspaces are then of the form $\{y_{i_1,j_1} = y_{i_2,j_2}\}$, $1 \le i_1, i_2 \le k-1$, $1 \le j_1, j_2 \le n$, of which there are finitely many. The result follows by Theorem 1.16.

$\square$

Example 5.5 (Polynomial configurations). Finally, we prove Corollary 1.20. We will in fact prove a stronger statement, namely that the result in Corollary 1.20 holds in $\mathbb{R}^n$ for all $n \geq 3$, with (1.8) replaced by 4-point configurations defined below in Corollary 5.6.

As in Example 5.3, fix $n \geq 1$, $k = 4$, and $m = 3n$, and let $B_1 = 0_{n\times 2n}$. We will use a Vandermonde-style matrix for the remaining B_i. To make the notation less cumbersome, for a function

$$\begin{aligned} g : \mathbb{N}\times\mathbb{N} &\to \mathbb{R} \\ (i,j) &\mapsto g(i,j), \end{aligned}$$

we denote by $(g(i,j))_{a\times b}$ the $a \times b$ matrix whose entry in the ith row and jth column is given by $g(i,j)$.

Corollary 5.6. *Let $a_1, \ldots, a_{2n} > 1$ be distinct real numbers, and let $\eta + 1, d \in \mathbb{N}$. Consider the following matrices:*

$$\begin{aligned} B_2 &= (a_j^{\eta+(i-1)d})_{n\times 2n}, \\ B_3 &= (a_j^{\eta+(n+i-1)d})_{n\times 2n}, \\ B_4 &= (a_j^{\eta+(2n+i-1)d})_{n\times 2n}. \end{aligned}$$

Suppose that $E \subset \mathbb{R}^n$ obeys the assumptions of Theorem 1.16, with ϵ_0 small enough depending on C and a_i. Then E contains a configuration of the form

$$x, \quad x + B_2 y, \quad x + B_3 y, \quad x + B_4 y \tag{5.4}$$

for some $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^{2n}$ with $B_i y \neq 0$ for $i = 2, 3, 4$.

The proof of Corollary 5.6 will rely on two short lemmas.

Lemma 5.7. *Suppose $0 \leq \eta_1 < \eta_2 < \ldots < \eta_t$ are integers. Then for any choice of constants $c_1, c_2, \ldots, c_t$ that are not all zero, the polynomial*

$$P(x) = \sum_{i=1}^{t} c_i x^{\eta_i}$$

has fewer than t distinct positive roots.

Proof. We prove this with induction. For $t = 1$, it is clear that $c_1 x^{\eta_1}$ cannot have a positive root since $c_1 \neq 0$, so the base case is satisfied. We make the inductive hypothesis that the lemma holds for t, and check $t + 1$. Suppose to the contrary that there exist constants $c_1, c_2, \ldots, c_{t+1}$, not all zero, such that the polynomial

$$P(x) = \sum_{i=1}^{t+1} c_i x^{\eta_i}$$

has at least $t + 1$ distinct positive roots. But then

$$x^{-\eta_1} P(x) = c_1 + c_2 x^{\eta_2 - \eta_1} + \cdots + c_{t+1} x^{\eta_{t+1} - \eta_1},$$

so by Rolle's Theorem, the following polynomial has at least t distinct positive roots:

$$\begin{aligned} P_1(x) &:= \frac{d}{dx}(x^{-\eta_1} P(x)) \\ &= c_2(\eta_2 - \eta_1) x^{\eta_2 - \eta_1 - 1} + c_3(\eta_3 - \eta_1) x^{\eta_3 - \eta_1 - 1} + \cdots \\ &\quad + c_{t+1}(\eta_{t+1} - \eta_1) x^{\eta_{t+1} - \eta_1 - 1} \\ &= \sum_{i=1}^{t} c_{i+1}(\eta_{i+1} - \eta_1) x^{\eta_{i+1} - \eta_1 - 1} \end{aligned}$$

Since η_i were strictly increasing integers, $c_{i+1}(\eta_{i+1} - \eta_1)$ are not all zero, and $\eta_{i+1} - \eta_1 - 1 \geq 0$ are strictly increasing integers. This contradicts the induction hypothesis and completes the proof. $\square$

Lemma 5.8. *If*

$$A = (a_j^{\eta_i})_{t \times s}$$

where $a_1, a_2, \ldots, a_s$ are distinct, positive real numbers and $0 \leq \eta_1 < \eta_2 < \ldots < \eta_t$ are integers, then A has full rank.

Proof. Without loss of generality, $t \leq s$. It suffices to show the following submatrix has full rank:

$$A_t = (a_j^{\eta_i})_{t \times t}.$$

This holds if and only if $\det A_t \neq 0$. If to the contrary $\det A_t = 0$, then we can find constants $c_1, c_2, \ldots, c_t$ that are not all zero such that $\sum_i^t c_i R_i = \vec{0}$, where R_i is the ith row of A_t; considering the kth position this says $\sum_{i=1}^t c_i a_j^{\eta_i} = 0$ for $1 \leq j \leq t$. That is, the polynomial

$$P(x) = \sum_{i=1}^{t} c_i x^{\eta_i}$$

has at least the t distinct positive roots $x = a_j$ for $1 \leq j \leq t$. This contradicts Lemma 5.7, so we must have $\det A_t \neq 0$ and hence A has full rank. $\square$

Proof of Corollary 5.6. By Lemma 5.8,

$$\operatorname{rank}\begin{pmatrix} B_{i_1} \\ B_{i_2} \end{pmatrix} = 2n$$

for $i_1, i_2 \in \{2, 3, 4\}$ distinct. It remains to check

$$\operatorname{rank}\begin{pmatrix} B_2 - B_4 \\ B_3 - B_4 \end{pmatrix} = \operatorname{rank}\begin{pmatrix} (a_j^{\eta+(i-1)d} - a_j^{\eta+(2n+i-1)d})_{n\times 2n} \\ (a_j^{\eta+(n+i-1)d} - a_j^{\eta+(2n+i-1)d})_{n\times 2n} \end{pmatrix} = 2n. \quad (5.5)$$

For constants $c_1, \ldots, c_{2n}$, consider the polynomial

$$\begin{aligned} Q_{c_1,\ldots,c_{2n}}(x) = c_1(x^{\eta} - x^{\eta+2nd}) &+ c_2(x^{\eta+d} - x^{\eta+(2n+1)d}) + \cdots \\ &+ c_n(x^{\eta+(n-1)d} - x^{\eta+(3n-1)d}) + c_{n+1}(x^{\eta+nd} - x^{\eta+2nd}) \\ &+ \cdots + c_{2n}(x^{\eta+(2n-1)d} - x^{\eta+(3n-1)d}) \end{aligned} \quad (5.6)$$

If (5.5) fails to hold, then as in the proof of Lemma 5.8, there are constants $c_1, \ldots, c_{2n}$ not all 0 whose corresponding polynomial $Q(x) := Q_{c_1,\ldots,c_{2n}}(x)$ has at least the $2n$ distinct roots $a_1, \ldots, a_{2n}$, all of which are larger than 1. We may simplify $Q(x)$ as

$$Q(x) = c_1 x^{\eta}(1 - x^{2nd}) + c_2 x^{\eta+d}(1 - x^{2nd}) + \cdots + c_n x^{\eta+(n-1)d}(1 - x^{2nd})$$

$$
\begin{aligned}
&\quad + c_{n+1}x^{\eta+nd}(1-x^{nd}) + \cdots + c_{2n}x^{\eta+(2n-1)d}(1-x^{nd}) \\
&= (1-x^{nd})[c_1x^{\eta}(1+x^{nd}) + c_2x^{\eta+d}(1+x^{nd}) + \cdots \\
&\quad + c_nx^{\eta+(n-1)d}(1+x^{nd}) + c_{n+1}x^{\eta+nd} + \cdots + c_{2n}x^{\eta+(2n-1)d}] \\
&= (1-x^{nd})P(x),
\end{aligned}
$$

where

$$
\begin{aligned}
P(x) = c_1x^{\eta} &+ c_2x^{\eta+d} + \cdots + c_nx^{\eta+(n-1)d} \\
&+ (c_1+c_{n+1})x^{\eta+nd} + \cdots + (c_n+c_{2n})x^{\eta+(2n-1)d}.
\end{aligned}
$$

The roots of $Q(x)$ which are larger than 1 coincide with the roots of $P(x)$. Notice that not all of the coefficients of $P(x)$ are 0, since not all of $c_1, \ldots, c_{2n}$ are 0. Then by Lemma 5.7, $P(x)$ has fewer than $2n$ positive roots, a contradiction. Thus, (5.5) holds and $\{A_1, \ldots, A_k\}$ is non-degenerate. The result follows by applying Theorem 1.16, with $V_i = \{y \in \mathbb{R}^{2n} : \ B_iy = 0\}$ for $i = 2, 3, 4$. □

It should be noted that a nearly identical proof will work for general k. If k is of the form 2ℓ or $2\ell+1$ for $\ell \in \mathbb{N}$, then $m = (\ell+1)n$ and the matrices are populated as before, using distinct real numbers $a_1, \ldots, a_{\ell n} > 1$. The proof of the analogue of (5.5) relies on a polynomial similar to that in (5.6), and uses the fact that $1 - x^{nd}$ divides $1 - x^{and}$ for any $a \in \mathbb{N}$.

Chapter 6

Conclusion

6.1 Overview of the results

Many interesting results have been produced which are related to Szemerédi's theorem of k-term arithmetic progressions in sets of positive density in the integers. In Section 1.1.1, we discussed a few of the motivating results in the discrete case. This was followed by some continuous analogues in the Euclidean setting of Szemerédi-type problems. Of particular note was Theorem 1.8 on 3-term arithmetic progressions in subsets of $\mathbb{R}$, of which this thesis can be seen to be an extension via Theorem 1.16, in terms of dimension and included patterns.

Similar to [30], a Fourier-analytic representation of the Λ multilinear form was proved in Proposition 2.1. This form allowed us to use Fourier decay to control the size of Λ as well as provide a way to use Λ with measures. This extension of Λ to measures, Λ^*, was the main result of Section 2.2. Proposition 2.8 showed that $\Lambda^*(\widehat{\mu}) > 0$ is the necessary bound to prove existence of the appropriate patterns.

The key idea behind proving positivity of $\Lambda^*(\widehat{\mu})$ was decomposing μ as $\mu = \mu_1 + \mu_2$, where μ_1 is absolutely continuous with bounded density, and μ_2 is singular but obeys good Fourier bounds. The portion involving only μ_1 is handled by a quantitative lower bound given by Proposition 3.1, using a similar method as used by Tao in [40] to prove Varnavides' Theorem.

The other portions were shown to be relatively negligible in Chapter 4, to complete the proof.

Chapter 5 contained several examples of geometric configurations covered by the main result. Of interest are the instances where negative results have been proved in the absence of the Fourier decay condition. In particular, Example 5.1 shows a similar copy of a specific triangle is contained in sets satisfying the assumptions of Theorem 1.16; Maga's result (Theorem 1.12) shows that this is not the case if the Fourier decay condition is not satisfied. Likewise, Example 5.3 shows existence of parallelograms in sets satisfying the assumptions of Theorem 1.16, while Maga (Theorem 1.11) indicates this is false without Fourier decay. Finally, Example 5.2 provides a multidimensional generalization of Łaba and Pramanik's result (Theorem 1.8).

6.2 Possible directions for future research

In this section, we will discuss possible directions for future research, inherently touching upon some of the the limitations of the research.

6.2.1 Non-degeneracy

We begin by recalling our notion of non-degeneracy for a set of appropriately-sized matrices as given in Definition 1.15. Although such a constraint on the A_j allows us to use ξ_j as a basis for S as in Lemma 2.7, it is not necessarily a requirement for the result to hold. We thus ask what happens in the case that non-degeneracy fails.

Question 6.1. *Suppose $\{A_1, \ldots, A_k\}$ is a collection of degenerate matrices. Is it necessarily true that the result of Theorem 1.16 fails?*

6.2.2 Valid range of m

Along the lines of examining the hypotheses of Theorem 1.16, recall the restriction placed on the variables n, m, k given by (1.6). In particular, we required $m \geq n \lceil (k+1)/2 \rceil$, which was needed for the estimate made in

Lemma 3.3 to produce suitable bounds on $|\Lambda(f_1, \ldots, f_k)|$. Prior to this point, we needed only that $m > nk/2$ in order to prove the integrand in Λ was absolutely convergent in Proposition 2.6, which is strictly weaker for $n \geq 2$ and k even or $n \geq 3$ and k odd.

Question 6.2. *Can Theorem 1.16 hold with the weaker condition $nk/2 < m < nk$ instead of (1.6)?*

It is worth noting that in proving Proposition 2.6, we made use of the fact that

$$|g_j(\eta_j)| \lesssim (1 + |\eta_j|)^{-\beta/2}, \qquad \eta_j \in \mathbb{R}^n.$$

If $|\eta_j| \gtrsim |\eta|$ for all j, then the integrand in Λ has good decay as $|\eta| \to \infty$:

$$\left| \prod_{j=1}^{k} g_j(\eta_j) \right| \lesssim (1 + |\eta|)^{-k\beta/2},$$

so that the integral is convergent whenever $k\beta/2 > nk - m$. This holds for β sufficiently close to n if $kn/2 > nk - m$, or equivalently, $m > nk/2$. Thus, this is the best range of m we could hope for, using the method outlined in the previous chapters.

6.2.3 Linear configurations

Even having a lower bound of $nk/2$ for m precludes some nice geometric configurations in the theory. In general, m can be viewed as the number of "free variables" over which we have no control. Recall in Example 5.3 that the set-up for parallelograms in $\mathbb{R}^2$ was $n = 2$, $k = 4$, and $m = 3 \cdot 2 = 6$. The free variables correspond to a starting point $x \in \mathbb{R}^2$, and two other arbitrary points $x + y, x + z \in \mathbb{R}^2$ for a total of 6 entries. The fourth point is determined by the other three, via $x + y + z$. Another simple geometric configuration would be a square in $\mathbb{R}^2$, which has $2 \cdot 2 = 4$ free variables. Given

$$x, \quad x + \begin{pmatrix} y_1 \\ y_2 \end{pmatrix},$$

where $x \in \mathbb{R}^2$ and $y_1, y_2 \in \mathbb{R}$, a square is generated by the following vertices:

$$x, \quad x + \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad x + \begin{pmatrix} -y_2 \\ y_1 \end{pmatrix}, \quad x + \begin{pmatrix} y_1 - y_2 \\ y_1 + y_2 \end{pmatrix}.$$

Unfortunately, $m > nk/2$ is not satisfied in this case, and so Theorem 1.16 fails regardless of the answer to Question 6.2.

A similar issue occurs when we attempt to get all the vertices of a general parallelotope, rather than just the generator vertices and the diagonal vertex as described in the discussion following Corollary 5.4. Here, a p-parallelotope in $\mathbb{R}^n$ would be given by a general $n \geq 1$, $k = 2^p$, and $m = (p+1)n$. Note that $m > nk/2$ fails if $p \geq 3$.

Another nice, geometric configuration which cannot be covered are colinear k-tuples for $k \geq 4$, since $m = 2n$ in this case. In particular, 4-term arithmetic progressions will not satisfy the conditions of the theorem. As mentioned in the introduction, Gowers utilized higher order Fourier analysis to handle longer arithmetic progressions in the integers, which also appears in the work of Green and Tao [19] in showing primes contain arithmetic progressions of arbitrary length. Gowers and Wolf [17] show more recently that this notion can provide results involving more general configurations as well. This leads to the following questions.

Question 6.3. *Could higher order Fourier analysis be utilized in place of linear Fourier analysis to provide better bounds on m? In particular, can k-term arithmetic progressions be included in this theory?*

Question 6.4. *If the answer to the previous question is yes, can such a theory also include geometric configurations such as squares (in $\mathbb{R}^2$) and parallelotopes (in $\mathbb{R}^n$)?*

6.2.4 Non-linear configurations

Finally, we note that the analysis involved in proving Theorem 1.16 relied on linear configurations. Two simple, non-linear configurations are patterns of the form $\{x, x+y, x+y^2\}$ and regular polytopes.

Question 6.5. *Is there a version of Theorem 1.16 which applies to a pattern of the form* $\{x, x+y, x+y^2\}$*?*

Question 6.6. *Is there a version of Theorem 1.16 which applies to regular polytopes?*

Bibliography

[1] J. Arias de Reyna. Some results connected with a problem of Erdős. III. *Proc. Amer. Math. Soc.*, 89(2):291–292, 1983. → pages 2

[2] F. A. Behrend. On sets of integers which contain no three terms in arithmetical progression. *Proc. Nat. Acad. Sci. U. S. A.*, 32:331–332, 1946. → pages 3

[3] C. Bluhm. Random recursive construction of Salem sets. *Ark. Mat.*, 34(1):51–63, 1996. → pages 8

[4] C. Bluhm. On a theorem of Kaufman: Cantor-type construction of linear fractal Salem sets. *Ark. Mat.*, 36(2):307–316, 1998. → pages 8

[5] D. Borwein and S. Z. Ditor. Translates of sequences in sets of positive measure. *Canad. Math. Bull.*, 21(4):497–498, 1978. → pages 2

[6] J. Bourgain. Construction of sets of positive measure not containing an affine image of a given infinite structures. *Israel J. Math.*, 60(3):333–344, 1987. → pages 2

[7] V. Chan, I. Łaba, and M. Pramanik. Finite configurations in sparse sets. *ArXiv e-prints*, July 2013. → pages iii

[8] B. Cook, Á. Magyar, and T. Titichetrakun. A Multidimensional Szemerédi Theorem in the primes. *ArXiv e-prints*, June 2013. → pages 4

[9] P. Erdős. Remarks on some problems in number theory. *Math. Balkanica*, 4:197–202, 1974. Papers presented at the Fifth Balkan Mathematical Congress (Belgrade, 1974). → pages 1

[10] P. Erdős and S. Kakutani. On a perfect set. *Colloq. Math.*, 4:195–196, 1957. → pages 5

[11] P. Erdős and P. Turán. On Some Sequences of Integers. *J. London Math. Soc.*, S1-11(4):261. → pages 2

[12] K. J. Falconer. On a problem of Erdős on sequences and measurable sets. *Proc. Amer. Math. Soc.*, 90(1):77–78, 1984. → pages 1

[13] K. J. Falconer. *Fractal geometry.* John Wiley & Sons Inc., Hoboken, NJ, second edition, 2003. Mathematical foundations and applications. → pages 6

[14] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Analyse Math.*, 31:204–256, 1977. → pages 3

[15] H. Furstenberg and Y. Katznelson. An ergodic Szemerédi theorem for commuting transformations. *J. Analyse Math.*, 34:275–291 (1979), 1978. → pages 3

[16] W. T. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001. → pages 3

[17] W. T. Gowers and J. Wolf. The true complexity of a system of linear equations. *Proc. Lond. Math. Soc. (3)*, 100(1):155–176, 2010. → pages 70

[18] B. Green. Roth's theorem in the primes. *Ann. of Math. (2)*, 161(3):1609–1636, 2005. → pages 4

[19] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. *Ann. of Math. (2)*, 167(2):481–547, 2008. → pages 4, 70

[20] V. Harangi, T. Keleti, G. Kiss, P. Maga, A. Máthé, P. Mattila, and B. Strenner. How large dimension guarantees a given angle? *Monatsh. Math.*, 171(2):169–187, 2013. → pages 6

[21] J.-P. Kahane. Sur certains ensembles de Salem. *Acta Math. Acad. Sci. Hungar.*, 21:87–89, 1970. → pages 8

[22] J.-P. Kahane. *Some random series of functions*, volume 5 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 1985. → pages 8

[23] J.-P. Kahane and R. Salem. *Ensembles parfaits et séries trigonométriques.* Hermann, Paris, second edition, 1994. With notes

by Kahane, Thomas W. Körner, Russell Lyons and Stephen William Drury. → pages 8

[24] R. Kaufman. On the theorem of Jarník and Besicovitch. *Acta Arith.*, 39(3):265–267, 1981. → pages 8

[25] T. Keleti. A 1-dimensional subset of the reals that intersects each of its translates in at most a single point. *Real Anal. Exchange*, 24(2):843–844, 1998/99. → pages 6, 9

[26] T. Keleti. Construction of one-dimensional subsets of the reals not containing similar copies of given patterns. *Anal. PDE*, 1(1):29–33, 2008. → pages 9

[27] Y. Kohayakawa, T. Łuczak, and V. Rödl. Arithmetic progressions of length three in subsets of a random set. *Acta Arith.*, 75(2):133–163, 1996. → pages 4

[28] M. N. Kolountzakis. Infinite patterns that can be avoided by measure. *Bull. London Math. Soc.*, 29(4):415–424, 1997. → pages 2

[29] P. Komjáth. Large sets not containing images of a given sequence. *Canad. Math. Bull.*, 26(1):41–43, 1983. → pages 2

[30] I. Łaba and M. Pramanik. Arithmetic progressions in sets of fractional dimension. *Geom. Funct. Anal.*, 19(2):429–456, 2009. → pages 7, 8, 11, 67

[31] P. Maga. Full dimensional sets without given patterns. *Real Anal. Exchange*, 36:79–90, 2010. → pages 9

[32] A. Máthé. Sets of large dimension not containing polynomial configurations. *ArXiv e-prints*, January 2012. → pages 6

[33] P. Mattila. *Geometry of sets and measures in Euclidean spaces*, volume 44 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995. Fractals and rectifiability. → pages 7

[34] H. I. Miller. Some results connected with a problem of Erdős. II. *Proc. Amer. Math. Soc.*, 75(2):265–268, 1979. → pages 2

[35] K. F. Roth. On certain sets of integers. *J. London Math. Soc.*, 28:104–109, 1953. → pages 3

[36] R. Salem. On singular monotonic functions whose spectrum has a given Hausdorff dimension. *Ark. Mat.*, 1:353–365, 1951. → pages 8

[37] R. Salem and D. C. Spencer. On sets of integers which contain no three terms in arithmetical progression. *Proc. Nat. Acad. Sci. U. S. A.*, 28:561–563, 1942. → pages 3

[38] E. Szemerédi. On sets of integers containing no four elements in arithmetic progression. *Acta Math. Acad. Sci. Hungar.*, 20:89–104, 1969. → pages 3

[39] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975. Collection of articles in memory of Juriĭ Vladimirovič Linnik. → pages 3

[40] T. Tao. Arithmetic progressions and the primes. *Collect. Math.*, (Vol. Extra):37–88, 2006. → pages 38, 42, 52, 53, 67

[41] T. Tao and T. Ziegler. A multi-dimensional Szemerédi theorem for the primes via a correspondence principle. *ArXiv e-prints*, June 2013. → pages 4