Group Actions in Geometric and Arithmetic Combinatorics

By

JAMES WILLIAM EDWARD WHEELER

School of Mathematics
UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Science.

School of Mathematics, September 2021

Word count: 41000

Abstract

This project seeks to further and combine results and tools from the theory of growth in finite groups with state of the art methods of arithmetic and geometric combinatorics. This is a modern area of research at the crossroads of pure mathematics, with connections to computer science and coding and complexity theory, unified by the general theme of pseudorandomness. A significant progress in this area began in the 2000s after the foundational work of Helfgott, followed by Bourgain, Gamburd, Sarnak, and others. The growth phenomenon appears to be inherently connected with the renowned Sum-Product conjecture of Erdős and Szemerédi, towards which there has been a lot of progress in the past 15 years.

More specifically the project has two main connected threads, first, the project considers specific families of groups, such as those of upper-triangular matrices, to uncover and categorise the structures therein that pose obstruction to growth and establish quantitative estimates for growth in their absence. The nature of these obstructions much depends on the field, where the matrix elements come from: analysing various scenarios to this effect is a specific novel feature of this project. Partially this scope of questions furthers the earlier results by Breuillard, Green and Tao, Gill and Helfgot, Murphy and Petridis and others.

Growth in groups, and especially the concept of energy arising in its study are immediately related to geometric incidence theory estimates, arising in connection of these groups' actions on homogeneous spaces. This constitutes the other thread of the project, focusing on Möbius hyperbolae (which have connections to the two dimensional special linear group). The project improves on earlier results due to Bourgain, Solymosi and Tardos, Shkredov and others by using a special set of tools both from growth in groups and geometric incidence theory.

Acknowledgements

First and foremost I must thank my supervisor, Misha Rudnev, without whom such a thesis could not have been completed, and whose insight, help and encouragement have made this journey both enjoyable and enlightening. I am also indebted to the co-authors of my papers, Brendan Murphy, Misha Rudnev and Audie Warren. A less obvious source of immense help and encouragement has been from my academic siblings, in particular Sophie Stevens and Peter Bradshaw who listened to my ramblings and distracted me with their own research. Following a similar theme, the combinatorics department and Bristol school of Mathematics and other PhD students here at Bristol have helped make this experience of creating a thesis.

Turning to the non-academic people who deserve thanks and who have helped me stay sane and grounded in reality whilst undertaking this thesis, the obvious is my family, in particular recalling those who have asked what I have been doing and been very patient as I tried to explain my mathematics to people with little to no mathematical background. The second is the various people I have met through Scouting providing both challenges and activities in a completely different vein from my research.

I finish by thanking my sources of funding, EPSRC & the University of Bristol, without their support the past four years would not have been possible.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: .....JAMES WHEELER..... DATE: ....17TH SEPTEMBER 2021....

Publications

Some parts of this thesis appear in the following publications:

- *Growth in Some Finite Three-Dimensional Matrix Groups* [113], joint work with Brendan Murphy.
- *Incidence Bounds with Möbius Hyperbolae in Positive Characteristic* [140], joint work with Misha Rudnev.
- *Incidences of Möbius Transformations in* $\mathbb{F}_p$ [197], joint work with Audie Warren.

Sections containing reproduction are clearly indicated in the text.

Table of Contents

List of Tables

List of Figures

Notation and Conventions

Before we start the thesis proper we shall make a note of the main notation that shall be used to allow for easy reference. To this end, we also include the fundamental results of the Cauchy-Schwartz inequality and Hölder's inequality.

Notation

We will use the following notation. $\mathbb{F}$ is a field, $\mathbb{F}_q = \mathbb{F}_{p^r}$, for p a prime, is the finite field of size q which is itself the r^{th} power of the prime p. $\mathbb{Z}$ is the ring of integers, $\mathbb{R}$ is the field of real numbers and $\mathbb{C}$ is the complex numbers. We take $\mathbb{F}^* := \mathbb{F} \setminus \{0\}$ to be the non zero elements of the field $\mathbb{F}$.

Given sets A and B, we define the sumset and productset as

$$A+B := \{a+b : a \in A, b \in B\} \text{ and } AB := \{ab : a \in A, b \in B\}.$$

We also note we can define corresponding sets for any binary operation, in particular later we will also use the product set of sets of matrices under matrix multiplication and product set of functions under composition. We may wish to have longer, say k-fold sums and products. For products A^k is fairly unambiguously the productset of A times itself k times. For addition we need to distinguish from dilates as such we will stick to the following notations

$$kA = \{a_1 + \cdots + a_k : a_i \in A\} \qquad \text{and} \qquad k \cdot A = \{ka : a \in A\}.$$

When we wish to ensure the identity and inverses are included in our set we have the following notation for $A \subseteq G$, where G is a group with identity e,

$$A_{(n)} := \left(A \cup A^{-1} \cup \{e\}\right)^n.$$

For the group generated by A we will write $\langle A \rangle$. We can see this as the limit as n tends to infinity of the above $A_{(n)}$ illustrating the above's use when we wish to consider objects close to a group. Further following this idea, the diameter of a group G and a set of generators A (so $G = \langle A \rangle$), $diam(\Gamma(G,A)) = \min\{k : A_{(k)} = G\}$.

We write $\mathbb{1}_A$ as the indicator function of A, that is

$$\mathbb{1}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}.$$

If instead of A in the subscript we instead have a formula then it returns 1 if the formula is true and 0 otherwise. We use the following notation, $r_{A+A}(x)$, to count the number of representations of x in $A+A$, that is

$$r_{A+A}(x) = \sum_{a,a' \in A} \mathbb{1}_{x=a+a'}.$$

Again we may change what is in the subscript and change the definition as expected for functions such as r_{AA}, r_{A+B}, or $r_{AA^{-1}AA^{-1}}$.

We shall define the number of incidences between a point set P and another set of lines or curves, say L, by

$$\mathcal{I}(P,L) := \left|\{(p,l) \in P \times L : p \in l\}\right|.$$

Note that for brevity and clarity of any algebra within proofs we shall often refer to the number of incidences we wish to bound in a proof by σ. When we do this we will write out what σ is in full at the start of a given proof to define the objects the incidences occur between.

We define a k-rich object (for example line, curve or plane) as an object containing at least k points. We may refer to a k-rich point as a point that is passed through by at least k objects. For a few applications, we will instead use a different definition of k-rich where there is both an upper and lower bound on the number of points on the object, for example in Corollary 5.9 we use the definition that a transform is 2^k-rich if *between* 2^k and $2^{k+1}-1$ points of P lie on the transform. When we use a different definition of a k-rich object such as this in the thesis we shall make it clear. For notational ease, if L is the set of objects we shall use L_k to refer to the subset of k-rich objects.

Will will refer to the set of all distances determined by a point by

$$\Delta(P) = \{|x-y| : x,y \in P\}.$$

If we need to specify the distance (predominately Euclidean or Minkowski distances in this thesis although others such as taxi-cab distance, Hausdorff distance or other general metrics could be used) we shall add subscripts, for instance using $\Delta_{1,1}(P)$ for Minkowski distances in the plane. To differentiate from pinned distances and the related set of distances determined by a specific point x and the point set P we shall use $\Delta_{pin}(P)$ and $\Delta_x(P)$. We can then define pinned distances as

$$\Delta_{pin}(P) := \max_x \Delta_x(P).$$

Asymptotic Notation

We will make use of Vinogradov notation which allows us to ignore constants. That is if there exists some constant C such that as x tends to infinity $f(x) \geqslant Cg(x)$, we write

$$f(x) \gg g(x).$$

Similarly we use $\ll$ if the reverse is true. If we have $f(x) \ll g(x)$ and $f(x) \ll g(x)$ we write $f(x) \sim g(x)$. We also use Big O notation which, defined through use of the above, is

$$f(x) = O(g(x)) \iff f(x) \ll g(x).$$

We may also wish to hide log terms as well as constants in our bounds, in this case, we will use the notation $\lesssim$ and $\gtrsim$.

Matrix Groups

We will use several matrix families, particularly in Chapter 4 but also Chapter 5. For ease of reference, we will list them here

- We let $T_n(\mathbb{F})$ denote the group of invertible $n \times n$ upper triangular matrices over a field $\mathbb{F}$.
- Let $U_n(\mathbb{F}_q)$ denote the unipotent subgroup of $T_n(\mathbb{F}_q)$, comprising matrices with 1's on the diagonal, the unitriangular matrices. We could appeal to symmetry to change results for upper triangular matrices to lower triangular matrices instead. We say that a matrix g in $T_n(\mathbb{F}_q)$ is $\mathbb{F}^*$-*potent* if $g = \lambda u$ where $\lambda \in \mathbb{F}^*$ and $u \in U_n(\mathbb{F}_q)$.unipotent
- Diagonal matrices have zeros in all entries bar the main diagonal. We will notate them as $D_n(\mathbb{F})$ for the $n \times n$ matrices over the field $\mathbb{F}$ or as $\text{Diag}(a_1, \ldots, a_n)$ for a specific matrix when we do not wish to write it out in full.
- For the whole set of $d \times d$ matrices we shall use $mat(d)$ or $mat(d, \mathbb{F})$ if the field is ambiguous.
- The Heisenberg group is a specific group of unitriangular (triangular matrices with ones on the diagonal)
$$H = H(\mathbb{F}) := \left\{ \begin{pmatrix} 1 & g_1 & g_3 \\ 0 & 1 & g_2 \\ 0 & 0 & 1 \end{pmatrix} : g_1, g_2, g_3 \in \mathbb{F} \right\}.$$
- The special linear $SL_n(\mathbb{F})$ is the group of $n \times n$ matrices over the field $\mathbb{F}$ with determinant 1. The general linear group $GL_n(\mathbb{F})$ is the group of invertible $n \times n$ matrices.
- The Affine group is a specific subgroup of the 2×2 upper triangular matrices defined as follows
$$\text{Aff}(\mathbb{F}) := \left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} : a, b \in \mathbb{F}, a \neq 0 \right\}.$$
- The Euclidean group, E_n, is the group of isometries of Euclidean space, that is the group generated by reflections, rotations and translations. The Orthogonal group, $O(d)$, is the subgroup generated by those isometries which also fix a point (the reflections and rotations), the Special Orthogonal group $SO(d)$ is those with determinant one. The Poincare group, $IO(k, n-k)$, and Lorentz subgroup, $O(1,3)$, are related groups for Minkowski space described in Appendix A.

We also note the following definitions for the commutator $[g,h] = g^{-1}h^{-1}gh$, noting we can also use this notation with sets for the set of commutators as $[A,B] = A^{-1}B^{-1}AB$.

A nilpotent group of step s has a central series (or lower central series or upper central series) terminating in s steps. In the following, we will use the Lower Central Series defined as

$$G = G_0 \rhd G_1 \rhd \cdots \rhd G_s = \{1\}$$

with $G_i := [G_{i-1}, G]$. Note that the series need not terminate in the identity if G is not nilpotent.

We define the span of a set X over a field $\mathbb{F}$ as follows

$$\mathrm{Span}_{\mathbb{F}}(X) := \sum_i f_i x_i$$

where $f_i \in \mathbb{F}$ and $x_i \in X$.

Simple Results

During the thesis, we will constantly make use of the following standard results. We start with the cornerstone of combinatorics, the Cauchy-Schwartz inequality.

Theorem 0.1. *[Cauchy-Schwartz inequality] Let A and B be finite sets then*

$$\left(\sum_1^n a_i b_i\right)^2 \leqslant \left(\sum_1^n a_i^2\right)\left(\sum_1^n b_i^2\right).$$

A more general result is Hölder's inequality.

Theorem 0.2. *[Hölder's inequality] Let $1/p + 1/q = 1$ then*

$$\sum_{k=1}^n |a_k b_k| \leqslant \left(\sum_{k=1}^n |a_k|^p\right)^{1/p} \left(\sum_{k=1}^n |b_k|^q\right)^{1/q}.$$

When p=q=2 this is Theorem 0.1 again.

The other major tool of combinatorialists is pigeonholing, which is given n items and $m < n$ categories, at least 2 items are in the same category (named for when the objects are pigeons being put into pigeonholes). A more refined version that is used a lot in this thesis is dyadic pigeonholing.

Theorem 0.3. *[Dyadic pigeonhole principle] Let A be a finite set of real numbers with $1 < a \leqslant \alpha$ for all $a \in A$. then there exists $A' \subset A$ such that for all $a' \in A'$, $k < a' \leqslant 2k$ and*

$$|A'|k \gg \frac{\sum_{a \in A} a}{\log \alpha}.$$

Chapter 1

Introduction

This is a thesis in the field of combinatorics. Whilst not one of the ubiquitous areas of mathematics anyone could name such as algebra, geometry, or calculus, it is in some ways a fundamental piece whilst also serving as the meeting ground of many areas (for example geometry, number theory, probability and many others). At its most simplistic combinatorics is counting (sharing its etymological roots with combinations and the German Kombinatorik). However, as anyone who has more than a passing acquaintance with mathematics will tell you, just because the problem sounds simple (which counting does, it was probably the first thing any of us did with mathematics) does not mean its answer is. These questions of counting in combinatorics have lead to areas including graph theory, finite geometry, sum-product results and matroid theory.

More specifically this is a thesis in arithmetic combinatorics and incidence geometry (a sub-area of the geometric combinatorics in the title of this thesis). These are two related fields as shall become obvious during this thesis. We shall start by describing each of these topics, providing some of the key results and open questions in the area as well as the way they lead to the new results in this thesis.

Starting with starting with Arithmetic combinatorics we concern ourselves with questions regarding counting objects which come from arithmetic operations. The primary examples are that of the sumset and productset $A+A$ and AA defined as

$$A+A := \{a_1+a_2 : a_1, a_2 \in A\} \quad \text{and} \quad AA := \{a_1a_2 : a_1, a_2 \in A\}.$$

In particular we tend to want to say how large these sets are with respect to the size of the original set A.

Arithmetic combinatorics can be seen as the intersection of combinatorics, number theory, and harmonic analysis, with the main goal of understanding the operations of addition and multiplication and how they interact. The most famous result of arithmetic combinatorics is the Erdős-Szemerédi sum-product theorem which states, informally, that the size of at least one of the sumset and the productset is large with respect to the original set. This result and its related body of research will be covered in Chapter 2. There is an associated conjecture as to the correct exponents to formalise this “large”. This theorem and associated conjecture have lead to

its own branch of mathematics where the exponent is improved and other related questions. The Erdős-Szemerédi sum-product theorem is a result set in a ring as is other results in arithmetic combinatorics (or more usually fields), similar questions can be asked instead in just abelian groups, alternatively seen as what if we only care about addition, these questions form the branch of additive combinatorics, much of this field is covered in Tao and Vu's homonymous book [185]. The other major branch of these style of results is for non-commutative groups, that is asking for bounds on the size of $A * A$ for other operations $*$. Of particular interest in this thesis are bounds for when A is a subset of certain matrix groups and $*$ is matrix multiplication. This will be considered in Chapter 4. As matrix multiplication consists of both addition and multiplication, due to the Erdős-Szemerédi sum-product theorem, one should expect such objects, $A * A$, to be large with respect to A unless A is in someway avoiding one of the operations. An example would be the 2×2 upper unitriangular matrices where matrix multiplication is equivalent to addition of the upper right entry. The aim is to use the properties of groups and the simplification of only having to deal with one operation to gain further knowledge whilst still containing a sum-product style result when, with the matrix groups as an example, you consider a particular entry.

Progress has been made on this style of questions by using incidence results which we shall look at shortly and then in more detail in Chapter 3, and the polynomial method. Incidence geometry has also linked in with Erdős distinct distances problem states that every set of points in the plane has a nearly-linear number of distinct distances, this was solved by Guth and Katz using a variation of the polynomial method. We will consider this link more in Chapter 5 along with the connection between circles and hyperbola.

Turning next to incidence geometry this name is more self-explanatory, we are aware of what geometry is and may even have our favourite geometries and so considering the fundamental action in combinatorics is counting it is not a far leap to try and count the incidences between geometric items. That is to ask questions such as

- Given n points and m lines (or similar objects such as curves) how many incidences can I make?
- What arrangement of lines provides the most intersections?
- What is the minimum number of lines to create a certain number of incidences?
- What can be said about the configuration of such geometric objects given constraints on the number of their incidences?
- What changes when we move to other objects besides lines such as curves or planes?

Brass, Moser, and Pach's book, Research Problems in Discrete Geometry [16], surveys many of these problems.

As should become obvious through this thesis these two areas of arithmetic combinatorics and incidence geometry are intrinsically linked. As such another question is why and to what extent are these fields linked?

The above questions for arithmetic combinatorics and incidence geometry can be asked over various fields. Historically the real numbers were the original area of study before being extended to complex numbers although now such questions have been asked over finite fields. This has come with several differences. First, there are fewer tools to use as the study of finite fields is newer than that of Euclidean geometry and we cannot rely on the topological properties of the real numbers (such as the fact that they are ordered). A second difference is the presence of finite subfields which will form a major theme of awkward cases to deal with, this is part of the reason why results for finite fields tend to first be discovered for finite fields $\mathbb{F}_p$ before being generalised to arbitrary finite fields $\mathbb{F}_q$.

Incidence results find applications in sum-product results (such as Elekes' improvement in the sum-product conjecture of Erdos and Szemerdi), in Kakeya problems (those dealing with arranging lines in different directions) which in turn lead to randomness extractors, and Slyvester-Gallei problems (bounding the dimension of the span based on local dependencies).

1.1 Structure of Thesis

Given we have a starting idea of the area of mathematics we will explore in this thesis, I will now lay out the general journey we will take in this thesis. We will start with introductions to these areas of mathematics, the key results and the lay of the land with regards to the current frontiers of knowledge and questions. This will set us up to understand my own contributions in the papers Growth in Some Finite Three-Dimensional Matrix Groups [113], On Incidence bounds with Möbius hyperbolae in positive characteristic [140], and Incidences of Möbius transformations in $\mathbb{F}_p$ [197] the first being joint work with Brendan Murphy, the second joint with Misha Rudnev and the last joint work with Audie Warren. We will study the first of these in detail during Chapter 4 and the second two during Chapter 5.

In particular, the structure of this thesis will be as follows:

- Chapter 1 is this chapter and has given a brief overview and introduction to the field of combinatorics in which this thesis sits, provided some of the key questions and problems in this field, and is currently providing the structure of the rest of the thesis.

- We start with introductions. Chapter 2 will focus on arithmetic combinatorics and growth. It will look at where the current results are and what direction they are moving. It will also introduce the tools that will be used later in the thesis to obtain the new results. In particular, this chapter will cover sum-product results and the history of such results, the concepts of growth, approximate groups and energy (both standard energies and higher-order energies) and the connections between these areas. These will tie into later chapters, particularly Chapter 4 as well as some of the corollaries arising in Chapter 5, where we have results about growth in matrix groups as well as energy bounds for such groups. Whilst not directly related to sum-product results, it arises when considering matrix multiplication as this consists of both addition and multiplication of the entries. We prove further energy

results in Chapter 5 as well as expander corollaries, both concepts introduced in this Chapter 2.

- Chapter 3 will do a similar role but focus on incidence geometry. It will introduce the major historic results for both the reals and finite fields (both in the small and large, with respect to the field characteristic p, cases) and give an overview of the history behind some of these results and where the current understanding lies. It will introduce some tools we will use later as well as some related open questions. This chapter will, as well as surveying many incidence results, connect these concepts back to sum-product results of Chapter 2, introduce the concept of k-rich objects which will be used in the proofs of the new results of later chapters and consider some of the related questions such as Erdős distinct distances, unit distances and other distance problems.

- Chapter 4 will focus on the results from the (joint with Brendan Murphy) paper Growth in Some Finite Three-Dimensional Matrix Groups [113]. The paper and thus this chapter studies the growth of product sets in some finite three-dimensional matrix groups. This culminates in three main results. We prove two results about the group of 2×2 upper triangular matrices over arbitrary finite fields: a product set estimate using techniques from multiplicative combinatorics (Theorem 4.6), and an energy estimate using incidence geometry (Theorem 4.9). We also prove an energy result for the Heisenberg group (Theorem 4.10) using the same basic method as for the energy result over upper triangular 2×2 matrices. The energy method gives better quantitative results, but only applies to small sets. The first result can also be seen as a classification of approximate subgroups of the 2×2 upper triangular matrices over general finite fields. This chapter will also touch on future work towards the $n \times n$ case, in particular, the 3×3 case which I worked on with Brendan but also talk about Brendan and others further work to solving the general question. This is contained in a paper of Murphy, Pyber, Szabó and Eberhard [36] seeking to fully generalise the results of Gill and Helfgott [60] (Theorem 4.2) and Breuillard and Green [20] (Theorem 4.4) to arbitrary finite fields and understand the extra difficulties such a generalisation encounters.

- Chapter 5 will then focus on incidence results, we will prove incidence bounds (Theorem 5.3 and 5.4) between a Cartesian point set and an arbitrary collection of translates of the hyperbola $y = 1/x$ in $\mathbb{F}_p$ and $\mathbb{R}$. The start of this Chapter will be based on the joint paper of myself and Misha Rudnev, Incidence bounds with Möbius hyperbolae in positive characteristic [140]. This paper proves new incidence bounds between a plane point set, which is a Cartesian product, and a set of translates H of the hyperbola $xy = \lambda \neq 0$, over a field of asymptotically large positive characteristic p. These bounds improve recent bounds by Shkredov, which are based on using explicit incidence estimates in the early terminated procedure of repeated applications of the Cauchy-Schwarz inequality, underlying many qualitative results related to growth and expansion in groups. The improvement – both quantitative, plus the bounds can deal with a general H, rather than a Cartesian product – is mostly due to a non-trivial "intermediate" bound (Theorem 5.5) on the number of k-rich Möbius hyperbolae in positive characteristic. We will introduce the Minkowski distance

and show how it is linked to and used in the above incidence bound, observing that a certain energy-type quantity in the context of H can be bounded via the L^2-moment of the Minkowski distance in H (this is done in Subsection 5.4.2) and can therefore fetch the corresponding estimates apropos of the Erdős distinct distance problem. We will consider some applications including some sum-product results and some on the Erdős single distance problem as well as consider the weakness in the method stopping a complete answer to this open problem. The chapter will finish by turning to the joint work of myself and Audie Warren contained in the paper Incidences of Möbius transformations in $\mathbb{F}_p$ [197]. This covers a generalisation of the intermediate Theorem 5.5 to Theorem 5.8, along with some asymmetric versions of previous results from earlier in the chapter. It concludes with a selection of applications including, but not limited to, a Beck's theorem style result (Corollary 5.9) and an expander result (Corollary 5.4). We also include these results proofs.

- We will finish with Appendix A which covers some of the ideas that would be needed to convert Theorem 5.7 to a result about Minkowski distances rather than using it as is once passing to a field extension. In particular, it covers Minkowski isometeries (as the analogue of Euclidean isometeries used in [110] to prove Theorem 5.7) and the Blaschke-Grünwald kinematic mapping (a map to move from the group of isometries into projective space, $\mathbb{PF}^3$, thus allowing the use of geometric tools) giving further background and insight than given or needed in Chapter 5.

Chapter 2

Arithmetic Combinatorics, Structure, and Approximate Groups

This chapter is to serve as a primer of the arithmetic combinatorics to follow. It will predominately focus on the structure of sets and the conditions on when the sets can exhibit growth. Of particular note will be the sumset and the product set defined in the introduction. We will start with the Erdős-Szemerédi theorem [49] and its related conjecture which started this branch of research. This result deals with rings (in applications, more often fields) requiring both addition and multiplication, when dealing with hard problems our first instinct is often to simplify the question. What if we only deal with one operation, say addition? This leads to results in abelian groups rather than rings. This section of additive combinatorics will comprise our second section with results such as Freiman's theorem and the Plünnecke-Ruzsa inequality which will be the main tool we take from this. Once we are considering groups we can consider other operations, in particular why restrict ourselves to abelian groups? This requires a change of assumption from small doubling to small tripling and will be discussed in Section 2.3.1 before generalising the tools we gained from additive combinatorics. This sets the basis for Chapter 4 where we have new results for growth in certain matrix groups. This, in some sense, brings us back full circle with both addition and multiplication playing a part in what happens between entries. We shall also introduce the concept of approximate groups which will give us a different way of thinking about such problems and make it easier to use the comprehensive machinery group theory has in tackling our problems. We finish this chapter by exploring the concept of energy which provides another measure of structure (as well as a method for exploring the structure between two different sets) and will be a major tool in Chapters 4 and 5 where we will both prove new energy results and use energy results as stepping stones towards other new results.

2.1 Sum-Product Phenomena

A thesis touching on arithmetic combinatorics would not be complete without the Erdős-Szemerédi theorem [49] and its related conjecture. As said above this is where we will start with the statement of the theorem.

Theorem 2.1. *[Erdős-Szemerédi] [49] There exist positive constants c and ε such that, for every finite and non-empty set of real numbers, A,*

$$\max(|AA|,|A+A|) \geqslant c|A|^{1+\varepsilon}.$$

The conjecture states that ε can be taken arbitrarily close to 1. The ideas behind this result are that sets with a small sumset are additive in nature (respectively small productsets relate to being multiplicative in structure) and a finite and non-empty set of real numbers cannot be both additive and multiplicative. This makes sense as sets with additive and multiplicative structure are fields (more precisely rings but as we deal with finite sets and a finite subring of a field is a subfield we shall stick to dealing with fields) and the reals have no finite subfields. It leaves the question are there sets a bit like fields? The conjecture says no. This result is the wellspring of this branch of mathematics and this idea that additive structure and multiplicative structure cannot coexist will be seen throughout this thesis. Later in Chapter 4 we prove a result for certain matrix groups which essentially states that subsets of this group must grow unless we have somehow killed off the addition or multiplication by being in cosets of subgroups that are abelian and so act like additive or multiplicative sets, specifically unipotent subgroups and tori or where interference from subfields cause an obstruction. In Chapter 5 we have several corollaries of our main result in the form of sum-product type results.

This section thus aims to ground our understanding of this sum-product result and serve as a basis to help our intuition of later problems to do with when we expect there to be obstructions to growth. To this end, we will start with some examples and the trivial results for this problem before briefly surveying the progress which has been made on this problem in the last forty or so years. Later sections will then consider related questions for subsets of groups rather than rings and thus with only a single operation, this is leading towards the new results in Chapter 4.

2.1.1 Examples

Starting with the intuition behind this result, we have three main examples, arithmetic progressions (that is $a,a+k,a+2k,a+3k,\dots$), geometric progressions ($a,a^k,ak^2,ak^3,\dots$), and random sets. Intuitively it is easy to see that the first two are heavily connected to addition and multiplication respectively. The first example is an example of a set with additive structure, the second multiplicative and the third neither additive nor multiplicative structure. We shall focus on the first example as the second can be transformed to this case by use of logarithms. We shall also consider some of the examples Erdős and Szemerédi provided in their original paper [49].

Taking our first example of an arithmetic progression, let

$$A = \{a,a+k,a+2k,\dots,a+(n-1)k\}.$$

We have that $|A| = n$. Starting with the easier of the sum and product sets

$$A+A = \{2a,2a+k,2a+2k,\dots,2a+(2n-2)k\}.$$

Hence $|A+A| = 2|A|-1$. The product set is more complicated. If we chose the simplest of arithmetic progressions

$$A_1 = \{1,2,\dots,n\}.$$

Then it is clear $A_1A_1 \subseteq \{1,2,\dots,n^2\}$ but estimating $|AA|$ precisely is known as the multiplication table problem, this has been studied by among others Kevin Ford [55], Erdős[46] [47], Tenenbaum [187], and Koukoulopoulos [93].

Asymptotically Erdős proved that for the number of distinct products in the multiplication table of the first N elements, $M(N)$, then $M(N)/N^2$ tends to 0 as N tends to infinity [46] then improving this in 1960 [47] to get that $M(N)$ is asymptotically $N^2/(\log N)^c$ with $c = 1-\frac{1+\log\log 2}{\log 2} \approx 0.86$ the Erdős–Tenenbaum–Ford constant. Ford [55] further improved this to $N^2/((\log N)^c(\log\log N)^{3/2})$.

We will note the following lemma from Nathanson and Tenenbaum [115] as our simple lower bound for the size of the productset of an Arithmetic progression

Lemma 2.1 (Lemma 5)**.** *[115] Let A, B be arithmetic progressions of length $n \geqslant 2$, then*

$$|AB| \geqslant \left(\frac{n}{\log n}\right)^2.$$

This shows, together with the prior, that $|A+A| \approx |A|$ and up to log terms $|AA| \approx |A|^2$ and so arithmetic progressions have a near-maximal product set but minimal sumset.

We have similar results for the geometric progression but with $|AA| \approx |A|$ and $|A+A| \approx |A|^2$. Taking our geometric progression as

$$A = \{a, ar, ar^2, \dots, ar^{n-1}\}$$

then $AA = \{a^2, a^2r, a^2r^2, \dots, a^2r^{2n-2}\}$ and so $|AA| = 2|A|-1$. This can be seen slightly quicker if one notes that by taking logarithms we are in the prior example. To look at $A+A$, first note that $A+A \subseteq \{2a,\dots,2ar^{n-1}\}$. We will consider the r-nary representation, this gives us that each pair of elements have a distinct sum so $|A+A| = \binom{|A|}{2}$.

For an understanding of a general set, we consider a random set. Take K to be an integer, large enough that $n^4/K \to 0$ as $n \to \infty$ (this ensures we should be expecting to get most of the set and thus be almost in the arithmetic progression case) and let $A \subset \{1,\dots,K\}$ such that $\mathbb{P}(a \in A) = n/K$. Then the expected size of A is n. We then calculate the expected size of the product and sumsets as

$$\mathbb{E}(|A+A|) = \binom{n}{2} \quad \text{and} \quad \mathbb{E}(|AA|) = \binom{n}{2}.$$

This should be expected as even small perturbations to any one of a,b,c,d destroy both of the following relations; $a+b = c+d$ and $ab = cd$. Note these relations shall be of import when we look at energy in Section 2.5. The probabilistic arguments for the above can be found in, among other places, Tao and Vu's book [185].

This is telling us that we should not expect additive or multiplicative structure in a set. It also tells us that the product set and sumset can both be maximal and so any hope for a set with a small sumset and productset would require lots of structure.

We also consider the example Erdős and Szemerédi provided in their original paper [49] to provide the upper bound of

$$\max\{|A+A|,|A\cdot A|\} \ll |A|^2 e^{-\frac{c\log|A|}{\log\log|A|}}.$$

This example shows why the $o(1)$ is required in the conjecture and that ε cannot be one, just arbitrarily close. To construct the example take

$$A = \left\{ \prod_{i=1}^{2j} p_i^{\varepsilon_i} : p_i \text{ prime }, p_i < (\log x)^3, \varepsilon_i \in \{0,1\} \right\},$$

for some large x and $2j$ the largest even integer not exceeding $\frac{\log x}{3\log\log x}$. This is the set of square-free numbers spanned by the first segment of primes. Note that x is larger than any element of A. By calculation Erdős and Szemerédi show $|A| = x^{2/3+o(1)}$ which comes from calculating the number of primes less than $(\log x)^3$ (which we shall call s) choose $2j$. As $a < x$ for all $a \in A$, we have $a + a' < 2x < |A|^{3/2+o(1)}$ and so the size of the sumset is below the desired bound. Turning next to the productset, we start by introducing the notation $aa' = Q^2 L$ where Q is the highest common factor of a and a' and L L is a product of relatively prime square-free integers (noting that Q and L are coprime as aa' must be cube free and a and a' are square-free). To continue we split into two cases when the number of distinct prime factors of the greatest common factor of a and a' is greater than j, that is Q has more than j distinct prime factors, this is shown to be small. In essence, the reasoning for this is once Q has most of the prime factors, both a and a' must have these prime factors and so have fewer degrees of freedom. Considering the extreme example when Q has $2j$ distinct primes, then the only option is $a = a'$, the next step down when Q has $2j-1$ distinct prime factors we have 1 degree of freedom for each of a and a' so they can be equal to Q or of the form Qp for a choice of prime p. This leads to a bound in this case by calculation of

$$|A|\log|A|\binom{2j}{j}\binom{s}{j} < |A|x^{1/3+o(1)} < |A|^{3/2+o(1)}.$$

All that remains is to deal with the case where Q does not have too many distinct prime factors. Erdős and Szemerédi then noted that after counting the distinct prime factors of L and Q (also noting that as the number of distinct prime factors of Q, $\nu(Q)$ is less than $j < \frac{\log x}{6\log\log x}$, then L has $2j - \nu(Q)$ distinct prime factors), $Q^2 L$ can be written in at least $\binom{2j}{j}$ ways as the product of two elements of A. This is because both a must then be the product of Q and a section of the primes of L and a' the product of the other Q and the rest of L as a and a' are square free. The number of representations of $Q^2 L$ is then the number of ways of splitting the primes in L between a and a'. Hence this case contribute less than $|A|^2 2^{\frac{-\log x}{3\log\log x}}$ and the bound is proven.

2.1.2 Trivial Results

To help with intuition it can be useful to look at what can be said trivially. Considering sumsets (as product sets work similarly as do more esoteric examples) it is easy to see that

$$2|A|-1 \leqslant |A+A| \ll |A|^2.$$

So if $|A+A| \approx |A|$ then we have lots of structure, if $|A+A| \approx |A|^2$ we have no structure. So far, so simple. The above fits into what we stated in our examples, this also shows we cannot expect to do better than these in the general case.

2.1.3 Brief History of the Progress on ε

Progress towards proving the sum-product conjecture has been through gradual increases in ε. In the original, Erdős and Szemerédi [49] proved a qualitative result over the integers. The result was quantified by Nathanson [114] with methods refined by Ford [54] who also noted the method worked for subsets of the real numbers. The first major milestone was due to Elekes [42] who managed 5/4 (so $\varepsilon = 1/4$). This was achieved through clever use of Szemerédei-Trotter theorem (this theorem will be discussed in Chapter 3), we shall show this proof below as an example of the connections between arithmetic combinatorics and incidence geometry. In 2005 Solymosi [161] used a different method which also works for complex numbers. The next major milestone was also due to Solymosi [162] who proved a bound of $n^{4/3}$ by use of multiplicative energy. This 4/3 has had incremental progress built upon on it by among others Konyagin and Shkredov [90] [89], Rudnev, Shkredov, and Stevens [138], and Shakan [147] before the current record due to Rudnev and Stevens [139], with $\frac{4}{3} + \frac{2}{1167}$.

We include the proof of Elekes below which makes use of incidence geometry. This shows a concrete connection between this chapter and the following chapter.

Theorem 2.2 (Elekes [42])**.** *Let $A \subseteq \mathbb{R}$ be a finite set. Then*

$$|A+A|^2|AA|^2 \gg |A|^5.$$

Hence

$$\max(|A+A|, |AA|) \gg |A|^{5/4}.$$

Proof. We chose a set of points and a set of lines with the aim of using the Szemerédi-Trotter theorem (Theorem 3.1) as follows:

$$P := (A+A) \times (AA) \qquad \text{and} \qquad L := \{l_{ab} : a, b \in A\}$$

where l_{ab} is the line given by the equation $y = a(x-b)$. Note each line l_{ab} passes through at least $|A|$ points as for all $c \in A$ the point $(b+c, ac) \in P$ and is on the line l_{ab}. The number of incidences between the points and lines is then at least

$$|A|^3 = |A||L| < \mathcal{I}(P,L).$$

Next we appeal to Szemerédi-Trotter (Theorem 3.1) which tells us that

$$\mathcal{I}(P,L) \ll |P|^{2/3}|L|^{2/3} + |P| + |L|,$$

and hence

$$|A|^3 \ll (|A+A||AA|)^{2/3}|A|^{4/3} + |A+A||AA|.$$

If the first term dominates then we are done, if the second dominates then we have

$$|A|^3 \ll |A+A||AA|,$$

and so

$$\max(|A+A|,|AA|) \gg |A|^{3/2}$$

which is even better.

□

2.1.4 Standard Restrictions

As well as restrictions to growth due to the structure of the sets involved we also have the less interesting restrictions to being almost contained within a larger group or subfield. These do not arise in the case of addition and multiplication in the integers nor finite fields of prime order (baring almost having everything), but in general finite fields subfields can be an issue and with examples of matrices there are many subgroups such as upper triangular and diagonal matrices which can occur.

This style of restrictions means that results have to either restrict the size of $|A|$ (equivalently having the characteristic of the field sufficiently large with respect to the cardinality of A) or other similar restrictions or results can consider so-called statistical terms. Examples of such results can be found due to Vinh.

Theorem 2.3 (Theorem 3)**.** *[194] Let P be a collection of points and L be a collection of lines in $\mathbb{F}_q^2$. Then we have*

$$|\{(p,l) \in P \times L : p \in l\}| \leqslant \frac{|P||L|}{q} + q^{1/2}\sqrt{|P||L|}.$$

Which leads to the following in a similar way that Szemerédi-Trotter led to Elekes' result [42].

Theorem 2.4 (Theorem 4)**.** *[194] Let $A \subseteq \mathbb{F}_q$ with q an odd prime power.*

$$\max(|A+A|,|AA|) \geqslant \frac{2|A|^2}{q^{1/2} + \sqrt{q + \frac{4|A|^3}{q}}}.$$

For $q^{1/2} \ll |A| \ll q^{2/3}$ then

$$\max(|A+A|,|AA|) \geqslant c\frac{|A|^2}{q^{1/2}}.$$

For $q^{2/3} \ll |A| \ll q$ then

$$\max(|A+A|,|AA|) \geqslant c(q|A|)^{1/2}.$$

These results improved on similar results due to Hart, Iosevich and Solymosi [69].

2.1.5 Other Fields

There is no good reason to only consider Theorem 2.1 and its related conjecture only in the real numbers. As such the problem has been generalised in many other settings, in particular, we have already stated that Solymosi's [160] result holds in the complex numbers, as does an earlier result due to Chang [24] which also applied to the quaternion algebra. Solymosi and Wong [168] also have a result for quaternions. Results for finite-dimensional division algebras[1] over $\mathbb{R}$ and semi-simple commutative Banach algebras[2] over $\mathbb{R}$ or $\mathbb{C}$ are due to Chang [23]. A continuous version has been studied (related to the Erdős-Volkmann ring conjecture [50]) and first proven by [39] with a quantitative version due to Bourgain [9].

Of more concern to this thesis is instead finite fields with most of the upcoming results dealing with these fields. These can be seen in two different cases; when the field is of prime order, and general finite fields. In the first case, difficulties arise from the lack of order and less developed tools compared to the real case as well as the potential to fill our entire field leading to further restriction on the sum and product sets that they are bounded by p, as such results tend to bound the size of A by some power of p. The later case of subfields causes additional problems. Early results in this direction include [13] [14] [15].

The goal for an explicit ε to bound $|A+A|+|AA|$ by $|A|^{1+\varepsilon}$ (assuming $|A|$ is sufficiently small and precisely how small varies from result to result) starts with $\varepsilon = 1/14$ due to Garaev [57] who first quantified the sum-product estimate explicitly based on the method of Bourgain, Katz and Tao [14]. This was refined by several papers [82] [12] [95] before the bound of $|A|^{1+1/11-o(1)}$ was reached by Rudnev [130] at what seemed the limit of this method. This result was matched up to a log factor in $\mathbb{F}_q$ by Li and Roche-Newton [96]. Garaev [58] proved that $|A+A|+|AA| \gg |A|^{1+\delta/2}$ if $\delta < 1/3$ and $|A| < |\mathbb{F}_p|^{1-\delta}$.

The next big result was due to Oliver Roche-Newton, Misha Rudnev, and Ilya D. Shkredov [128] who proved, in the spirit of Elekes [42], the following

$$|A \pm A||AA| = \Omega\left(|A|^{6/5}\right).$$

It is also of note that this applies in $\mathbb{F}_q$ as well with suitable constraints. This was again improved incrementally by [148] [136] [27].

Most recently Mohammadi and Stevens [103] attained the exponent 5/4 in $\mathbb{F}_p$ leaving the state of the art result as follows

Theorem 2.5 (Theorem 2)**.** *[103] Let $\mathbb{F}$ be a field of characteristic $p \neq 2$. Let $A \subset \mathbb{F}$. If $p > 0$ suppose in addition that $|A| \ll p^{1/2}$. Then*

$$\max\{|A \pm A|, |A * A|\} \gtrsim |A|^{5/4}$$

where $* \in \{\times, \div\}$. Moreover, this result applies to all four choices of binary operator.

[1] An algebra in which division, except by zero, is always possible.

[2] Algebras that are also Banach spaces and so is a normed space which is complete, that is every Cauchy sequence's limit is within the space itself.

It is of note that this exponent is the same as Elekes' result [42] for the reals, especially when considering the next milestone result in the reals due to Solymosi [162] relies on the ordering of the reals which does not exist for finite fields.

Another category of rings we can ask sum-product style problems over is that of matrix rings. We will consider these in greater depth in Chapter 4. We can consider these as rings with both a multiplication and addition operation defined (that of matrix multiplication and addition of corresponding entries) or as a group with the operation of matrix multiplication which contains both standard addition and multiplication and thus of interest, we consider results for groups in the next two sections, abelian groups in Section 2.2 and non-commutative groups such as these in Section 2.3 and Chapter 4.

Some results in the first of these two directions include Chang [25] who showed for $\mathrm{Mat}(d)$ the set of $d \times d$ matrices over $\mathbb{R}$ the sum-product result stated below as well as another for $\mathrm{Sym}(d)$ the subset of symmetric matrices.

Theorem 2.6. *[25] Let $A \subset Mat(d)$ and $|A| = N$. If $\det(a - a') \neq 0$, for all $a \neq a' \in A$, then*

$$|A + A| + |A \cdot A| > \phi(N)N,$$

where $\phi(N) \to \infty$ as $N \to \infty$.

For Symmetric matrices they instead have for an $\varepsilon > 0$ depending on the dimension

$$|A + A| + |AA| > |A|^{1+\varepsilon}$$

Solymosi and Vu [167] showed the following for a suitably restricted set of matrices.

Theorem 2.7. *[167] Let A be a finite κ-well-conditioned set of size d matrices with complex entries. Then we have*

$$|A + A| \times |AA| \geqslant \Omega_{\kappa,d}(|A|^{5/2}).$$

Consequently, we have

$$|A + A| + |AA| \geqslant \Omega_{\kappa,d}(|A|^{5/4}).$$

Where κ-well-conditioned is to do with the matrices being far from singular. Solymosi and Tao [163] showed an exponent of $1 + 1/(4 + \varepsilon)$ rather than 5/4 without the requirement of $k \times k$ matrices.

The reason why the matrices have to be restricted to well-conditioned ones can be seen by considering the example of $d \times d$ matrices of the form $I + xE_{i,j}$ where I is the identity matrix and $E_{i,j}$ is the matrix considering of zeros everywhere except the $\mathrm{i}^{\mathrm{th}}\mathrm{j}^{\mathrm{th}}$ entry which is instead 1. Taking A as a subset of all matrices of the form $I + xE_{1,d}$, then $|A + A| = |AA| = 2|A| - 1$.

Solymosi and Wong [168] later showed that a family of well condition matrices or the quaternions we have the sum-product estimate

$$\max(|A + A|, |AA|) \gg \frac{|A|^{4/3}}{(\log A)^{1/3}}.$$

In the case of diagonal matrices over the reals, which we note is equivalent to a sum-product result for subsets of $\mathbb{R}^d$, Mudgal [105] proved for all $\delta < 1/3 + 5/5277$, we have

$$|A+A| + |AA| \gg_d |A|^{1+\delta/d}.$$

We also note there has been study into the sum-product phenomenon in arbitrary rings [178].

2.2 Additive Combinatorics

As one would expect of an unproven conjecture of Erdős and Szemerédi, the sum-product conjecture can be considered hard. Following mathematicians instinct to ask related but simpler questions when confronted with something they cannot solve, this section covers the sentiment of if we cannot do two operations (addition and multiplication) together, what can we say for just one? Alternatively, this can be seen as taking a question of rings and asking it instead for a group. As the sum-product conjecture is asked for commutative operations this was originally asked for abelian groups which is what we will concern ourselves in this section. More concretely we will cover the main result here, that of Freiman's theorem (Theorem 2.8) as an understanding of what is known in this branch of mathematics before introducing various tools stemming from the Ruzsa triangle inequality aiming towards the Plünnecke-Ruzsa inequality (Theorem 2.13). In the following section, we shall lose the abelian restriction, discuss what issues this causes, and generalise the tools to non-commutative groups which in turn sets the foundation for the new results in Chapter 4 where we detail new results for growth in particular matrix groups.

During this section we will focus on the operation of addition (noting that multiplication is similar) and thus predominately concern ourselves with the two questions: If $|A+A| \leqslant K|A|$ what can be said about A? If $|A+A| \leqslant K|A|$ what can be said about $|A+A+A|$ and other iterated sumsets? The second we shall be answered by the Plünnecke-Ruzsa inequality (Theorem 2.13) in the next subsection. The first of these two is asking about additive structure, we have seen an example in Subsection 2.1.1. To fully answer the first question we shall use Freiman's theorem which will require the definition of generalised arithmetic progressions which is as follows.

Definition 2.1. *A* generalised arithmetic progression *of dimension d is*

$$\{x_0 + l_1x_1 + \cdots + l_dx_d : 0 \leq l_1 < L_1, \ldots, 0 \leq l_d < L_d\}$$

where $x_0, x_1, \ldots, x_d, L_1, \ldots, L_d \in \mathbb{Z}$.

Put simply a generalised arithmetic progression is like an arithmetic progression but we have more than one common difference (the x_i for $1 \leqslant i \leqslant d$ in the above definition). As an example with common difference 4 and 7 we could have

$$11, 15, 18, 19, 22, 23, 25, \ldots.$$

The size of a proper generalised arithmetic progression is the product $L_1L_2 \ldots L_d$. We could also see these generalised arithmetic progressions as the sumsets of multiple arithmetic

progressions with zero added in. That is in the above example we could see this as $(A_4 \cup \{0\}) + (A_7 \cup \{0\})$ where A_i is the arithmetic progression with difference i and suitable starting position (11 in the above example).

This definition along with Freiman's theorem [56] then allows us to answer what a set with a small sumset looks like.

Theorem 2.8. *[Freiman's Theorem] [56] If A is a finite subset of $\mathbb{Z}$ with $|A+A| \leq K|A|$, then A is contained in a generalised arithmetic progression of dimension at most $d(K)$ and size at most $f(K)|A|$, where $d(K)$ and $f(K)$ are constants depending only on K.*

Theorem 2.8 is an example of a structure theorem, an easier structure theorem is the Balog-Szemerédi-Gowers theorem which has the advantage of polynomial dependence on the doubling constant K. The Balog-Szemerédi-Gowers Theorem due to Balog and Szemerédi [5] was first proven using a regularity lemma before a more effective proof due to Gowers [61]. The theorem states the following.

Theorem 2.9 (BSG)**.** *Suppose that A,B are two additive sets with $E(A,B) > \eta|A|^{3/2}|B|^{3/2}$. Then there are sets $A' \subseteq A, B' \subseteq B$ such that $|A'| > c\eta^C|A|$, $|B'| > c\eta^C|B|$ and $|A'-B'| \leqslant C\eta^{-C}|A'|^{1/2}|B'|^{1/2}$.*

There is a non-commutative version, however, you have to be careful with your analogue of differences (and energy), the correct way is via "differences" $a^{-1}b$.

Freiman's theorem was generalised to any abelian group, G, by Green and Ruzsa [62] in the following theorem.

Theorem 2.10. *[62] Let $A \subseteq G$ satisfy $|A+A| \leqslant K|A|$. Then A is contained in a coset progression of dimension at most $d(K)$ and size at most $f(K)|A|$. We may take $d(K) = CK^4\log(K+2)$ and $f(K) = \exp(CK^4\log^2(K+2))$ for some absolute constant C.*

This result requires the definition of *coset progression of dimension d* as a subset of G of the form $P+H$, where $H \leqslant G$ is a subgroup, P is a proper progression of dimension d. The size of a coset progression is its cardinality. Tao [179] further generalised the result to solvable groups. Quantitatively, better results are known.

Note that we can do similar with multiplicative structure and geometric progressions (again this can also be seen by taking logarithms).

2.2.1 Ruzsa Calculus

There are several tools available to us in our quest to improve on sum-product like results in abelian groups. In this section, we make a note of the Ruzsa triangle inequality and the Plünnecke-Ruzsa inequality. We shall revisit these tools and move them to a non-commutative setting in Subsection 2.3.2 which in turn find use both in proving the properties of approximate groups introduced in Section 2.4 and later proofs in Chapter 4.

We start with the Ruzsa triangle inequality which states the following

Theorem 2.11 (Ruzsa triangle inequality)**.** *[143] If A, B, and C are finite subsets of an abelian group, then*

$$|A||B-C| \leqslant |A-B||A-C|.$$

An alternative formation requires Ruzsa distance which is defined as

$$d(A,B) = \log \frac{|A-B|}{\sqrt{|A||B|}}.$$

Using this we can reformulate Theorem 2.11 as follows

Theorem 2.12 (Ruzsa triangle inequality)**.** *If A, B, and C are finite subsets of an abelian group, then*

$$d(B,C) \leqslant d(A,B) + d(A,C).$$

I also provide a short proof.

Proof. To prove this we construct an injective function from $A \times (B-C)$ to $(A-B) \times (A-C)$. For each $x \in B-C$ we choose some $b_x \in B$ and $c_x \in C$ such that $x = b_x - c_x$. Then we define our injective function as

$$\phi : A \times (B-C) \hookrightarrow (A-B) \times (A-C)$$
$$(a,x) \mapsto (a-b_x, a-c_x).$$

As this ϕ is an injection we have that

$$|A||B-C| \leqslant |A-B||A-C|.$$

□

Note that we can follow the same idea as the above proof for a non-commutative group so that if A, B, and C are finite subsets of a non-abelian group, then

$$|A||B^{-1}C| \leqslant |AB||AC|. \tag{2.1}$$

A proof of this can be found in Lemma 2.3.4 of Tointon's book [190].

The Ruzsa triangle inequality is used to prove the Plünnecke-Ruzsa inequality which answers the question of what can be said about iterated sumsets such as $A+A+A$ or $A+A+A+A$ given a restricted sumset. In the abelian group case, it tells us that once the sumset is small, so too is the triple sumset and so on.

Theorem 2.13 (Plünnecke-Ruzsa inequality [123])**.** *If A and B are finite subsets of an abelian group and K is a constant so that $|A+B| \leq K|A|$, then for all non-negative integers m and n,*

$$|mB - nB| \leq K^{m+n}|A|.$$

The Plünnecke-Ruzsa inequality was first proved by Plünnecke [123] in 1970 before being rediscovered in 1989 by Ruzsa [141]. They both used techniques from graph theory however in 2012 Petridis [119] provided an elementary approach.

2.3 Non Commutative Groups

There is no reason to restrict ourselves to just addition and multiplication and thus we can consider other binary operations. In this section, we shall consider how subsets of non-commutative groups grow building on the previous section. This idea of considering how subsets grow arises in several branches of mathematics; additive combinatorics considers growth in abelian groups (as seen in the previous section), geometric group theory considers how $|A^k|$ grows as k is taken to infinity and Group theory itself considers the special case when $|AA| = |A|$. More details can be found in a survey of Helfgott [75].

This section shall start by referencing some results about growth in non-commutative groups although we shall go into greater detail in Chapter 4. We shall instead predominately concern ourselves with why such results require small tripling rather than small doubling as in the abelian case and providing the non-commutative versions of the Plünnecke-Ruzsa inequality.

This idea of studying non-commutative groups started with Helfgot [73] who looked at the group $SL_2(\mathbb{F}_p)$. This branch has been continued by, among others, Helfgot [74] ($SL_3(\mathbb{F}_p)$), Pyber and Szabó [125], and Breuillard, Green, and Tao [21] ($SL_n(\mathbb{F}_p)$), Helfgott and Seress [77] (symmetric and alternating groups), Petridis, Roche-Newton, Rudnev and Warren [121], who studied the affine group,

$$\text{Aff}(\mathbb{F}) := \left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} : a, b \in \mathbb{F}, a \neq 0 \right\},$$

as well as Murphy and the author [113] who studied upper triangular matrices, this last paper serving as the basis for Chapter 4 where will consider upper triangular matrices over arbitrary finite fields as well as the Heisenberg group over the same fields as a three dimensional Lie group. Tointon's book [190] surveys many of these ideas.

2.3.1 Small Tripling

Unlike in abelian cases where Theorem 2.13 states that knowing a set has small doubling means we have small tripling and so on, this is not necessarily the case in non-abelian groups. This is easiest to note by considering the following example, as we are not in the abelian case will use multiplicative notation.

For the example, we consider $A := H \cup \{x\}$ where H is a finite subgroup of our universal group G and $x \in G$ such that $H \cap xHx^{-1} = \{id\}$. That is x is a conjugating element of H such that x is not a member of the normaliser of H.

If we consider A^2 we see

$$A^2 = \{x^2\} \cup xH \cup Hx \cup H$$

so

$$|A^2| = 3|H| + 1 = 3|A| - 2 \leqslant 3|A|.$$

This shows that under doubling this A does not grow by much. However, note that

$$A^3 \supseteq HxH$$

and as $|HxH| = |H|^2$ so $|A^3|$ is greater than $K|A|$ for a constant K, and so our set of small doubling has large tripling unlike the abelian case of Plünnecke-Ruzsa. This explains why we required small tripling in our non-commutative Ruzsa calculus such as Lemma 2.2 and the results leading to its proof.

Tao [177] proved that this double coset is typical of the general case of sets A with small doubling but not small tripling. Explicitly he proved that given such an A there must exist such a double coset which is significantly larger than A.

2.3.2 Noncommutative Ruzsa Calculus

With our plan to prove results for non-commutative groups it would be helpful to have access to similar tools as in the abelian case. This subsection will seek to generalise much of Subsection 2.2.1 using the non-commutative Ruzsa triangle inequality (2.1).

Our big aim is to take the Plünnecke-Ruzsa inequality and find its non-commutative version. The case $m = n = 1$ generalises to arbitrary (potentially non abelian) groups fairly simply, if $|A^2| \leqslant K|A|$ then by the non commutative Ruzsa triangle inequality (2.1) taking the A, B, and C all as A gives use that $|AA^{-1}| \leqslant K^2|A|$ and similarly instead taking A, B, and C all as A^{-1} gives $|A^{-1}A| \leqslant K^2|A|$. However, for longer chains of A and A^{-1} we must instead require $|A^3| \leqslant K|A|$ as we saw in Subsection 2.3.1.

Sometimes it is useful for the subsets we are working with to contain inverses and the identity so they act more like a group. To this end we introduce the notation

$$A_{(k)} := \left(A \cup A^{-1} \cup \{e\}\right)^k.$$

The following lemma links small tripling ($|A^3| \ll |A|$) with the size of $A_{(k)}$. We note that this includes the non commutative Plünnecke-Ruzsa as the analogue of $nA - mA$ is $A^{\delta_1} \dots A^{\delta_{n+m}}$, where δ_i is either 1 or -1 and we have n ones and m minus ones, and $A^{\delta_1} \dots A^{\delta_{n+m}} \subseteq A_{(n+m)}$.

Lemma 2.2 (Plünnecke-Ruzsa)**.** *Suppose that A is a finite subset of an arbitrary group such that $|A^3| \leqslant K|A|$. Then*

$$|A_{(3)}| \leqslant 27K^3|A|,$$

for all $k \geqslant 3$

$$|A_{(k)}| \leqslant \left(\frac{|A_{(3)}|}{|A_{(1)}|}\right)^{k-1} |A_{(1)}|,$$

and hence, for all $k \geqslant 3$,

$$|A_{(k)}| \leqslant 27^k K^{3(k-1)}|A|.$$

Proof. This proof mostly consists of repeated applications of the non-commutative Ruzsa triangle inequality.

Starting with the first inequality we note that $A_{(3)}$ is the union of the twenty-seven sets XYZ where each of X, Y, and Z are chosen from the set $\{A, A^{-1}, \{e\}\}$. Some of these sets are easy to bound, $|AAA| \leqslant K|A|$ by assumption, $|\{e \cdot e \cdot e\}| = |\{e\}| = 1$ and $|A^{-1}A^{-1}A^{-1}| = |A^3| \leqslant K|A|$. For

the rest, we turn to the non-commutative Ruzsa triangle inequality. For ease, we tabulate them with the first three columns stating our choice of A, B, and C in Inequality (2.1) and the last the bound that this gives.

A	B	C	Bound
A	A^2	A	$\lvert A^{-2}A\rvert \leqslant K^2\lvert A\rvert$
A^{-1}	A^{-2}	A^{-1}	$\lvert A^2A^{-1}\rvert \leqslant K^2\lvert A\rvert$
A	A	AA^{-1}	$\lvert A^{-1}AA^{-1}\rvert \leqslant K^3\lvert A\rvert$
A^{-1}	A^{-1}	$A^{-1}A$	$\lvert AA^{-1}A\rvert \leqslant K^3\lvert A\rvert$

Table 2.1: Table of repeat use of Ruzsa triangle inequality in proof of Lemma 2.2.

Note also that as $(A^{-2}A)^{-1} = A^{-1}A^2$ and $(A^2A^{-1})^{-1} = AA^{-2}$ we also have $\lvert A^{-1}A^2\rvert \leqslant K^2\lvert A\rvert$ and $\lvert AA^{-2}\rvert \leqslant K^2\lvert A\rvert$. Combining these although gives

$$\lvert A_{(3)}\rvert \leqslant 27K^3\lvert A\rvert.$$

To prove the rest we again use Inequality (2.1) with A taken to be $A_{(1)}$, B taken as $A_{(2)}$, and C taken as $A_{(k-2)}$. Also note that $A_{(i)} = A_{(i)}^{-1}$. This gives

$$\lvert A_{(1)}\rvert\lvert A_{(2)}A_{(k-2)}\rvert \leqslant \lvert A_{(1)}A_{(2)}\rvert\lvert A_{(1)}A_{(k-2)}\rvert,$$

which after rearranging gives

$$\lvert A_{(k)}\rvert \leqslant \frac{\lvert A_{(3)}\rvert}{\lvert A_{(1)}\rvert}A_{(k-1)}.$$

Iteration of this step leads to the second bound in the lemma.

The last bound comes from noting $\lvert A\rvert \leqslant \lvert\lvert A_{(1)}\rvert \leqslant 2\lvert A\rvert + 1$ and plugging the first into the second. □

We also make use of the following lemma due to Ruzsa which allows us to bound the number of copies of $A^{-1}A$ needed to cover B given a few constraints. The essence of the proof is the greedy algorithm.

Lemma 2.3 (Ruzsa covering lemma)**.** *Let A, B be finite subsets of G. Suppose $\lvert AB\rvert \leqslant K\lvert A\rvert$, then there is a subset $X \subset B$ with $\lvert X\rvert \leqslant K$ and $B \subset A^{-1}AX$.*

We note that there is a commutative version of this result. We make use of this result specifically in Subsection 2.4.1 concerning the covering inherent in the definition of approximate subgroup but similar covering lemmas are a fundamental tool in combinatorics.

2.4 Approximate Group Theory

This section provides an alternative way of thinking about these problems. If $\lvert AA\rvert = \lvert A\rvert$ then A is a group so if $\lvert AA\rvert \leqslant K\lvert A\rvert$ then A is in some way close to being a group. This idea was formalised

with the definition of a K-approximate group due to Tao [177]. We note that whilst we should really talk about approximate subgroups we will normally use the term approximate group and leave that we are working inside a larger group as an accepted fact.

Definition 2.2. *Let G be a group and $K \geqslant 1$. A non-empty subset $A \subset G$ is a K-approximate subgroup of G if:*

- *It is symmetric, that is if $g \in A$ then $g^{-1} \in A$, and it contains the identity.*
- *There exists a subset $X \subset G$ of cardinality $|X| \leqslant K$ such that $A \cdot A \subset X \cdot A$.*

When $K = 1$, an approximate group is a subgroup. Considering one of our examples from earlier, if we consider an arithmetic progression $A = \{-na, \dots, -a, 0, a, 2a, 3a, \dots na\}$ then A is symmetric and contains the identity 0, we choose $X = \{-na, na\}$ then $AA \subseteq XA$ so A is a 2-approximate group. If a set has small tripling then $(A \cup A^{-1} \cup \{e\})^2$ is a $K^{O(1)}$-approximate group (by Lemma 2.2). As such, the question of when finite subsets of a group grow is equivalent to classifying approximate groups.

Research into classifying approximate groups has been undertaken by Freiman (albeit with different terminology) who classified approximate subgroups of the integers [56]. Green and Ruzsa [62] generalised Freiman's result to all abelian groups. Chapter 4 includes new results classifying approximate subgroups of 2×2 upper triangular matrices over general finite fields. Much of the next few subsections providing the background for approximate groups and the tools for their use later in this thesis can be found in the books of Tointon [190] and Tao and Vu [185].

2.4.1 Equivalence with Small Tripling

As stated above, the definition of an approximate group is linked to that of small tripling. In this subsection, we will make this connection explicate. For ease of reference, we will do this in the form of a proposition.

We have that small tripling implies an approximate group.

Proposition 2.1 (Proposition 2.5.5 [190])**.** *Let $A \subset G$ with $|A^3| \leqslant K|A|$, then $A_{(2)} := \left(A \cup A^{-1} \cup \{e\}\right)^2$ is an $O(K^9)$-approximate group.*

We note that if $A = A^{-1}$ and $|A^5| \leqslant K|A|$ then A^2 is an approximate group by Ruzsa's covering lemma (Lemma 2.3). In a similar vein Corollary 3.11. of Tao [177] says

Corollary 2.1. *Let A be a multiplicative set such that $|A^3| \leqslant K|A|$. Then the set $A_{(3)} := \left(A \cup A^{-1} \cup \{e\}\right)^3$ is a $O(K^O(1))$-approximate group. In particular, if A is symmetric and contains the identity, then A^3 is a $O(KO(1))$-approximate group.*

Proof. To start we note by Lemma 2.2 $A_{(3)} \leqslant 27K^3|A|$. we also have that $A_{(3)}$ is symmetric so all that remains is to find our X as in the definition. To do this we appeal to Lemma 2.3 after noting that by Lemma 2.2

$$|A_{(3)} \cdot A_{(3)}| = |A_{(6)}| \ll K^{O(1)}|A| \leqslant K^{O(1)}|A_{(3)}|.$$

□

We also show that once we have an approximate group then all our iterated product sets are bounded, that is

Lemma 2.4. *If A is a finite K-approximate group then*

$$|A^m| \leqslant K^{m-1}|A|.$$

Proof. Using the above definition of an approximate group we have a subset X associated with A. We are done by noting

$$A^m = A^{m-1}A \subseteq XA^{m-1} \subseteq \ldots X^{m-2}A^2 \subseteq X^{m-1}A.$$

□

2.4.2 Inheritance

Having now shown the connection between small tripling and approximate groups, we will next look at how inheritable the property of being an approximate group is. We note it is obvious that a group homomorphism still preserves approximate groups, we will also consider how approximate subgroups interact with intersections.

Proposition 2.2. *Let A be a K-approximate subgroup of G. Given $H \leqslant G$ a subgroup of G, then $A^2 \cap H$ is a K^3-approximate subgroup.*

Proof. $A^2 \cap H$ is still a symmetric set containing the identity, so we need only concern ourselves with finding the appropriate X, to do this we will use the Ruzsa covering lemma. For notation's sake, we will use X as the set from the definition of A as an approximate group and seek to define X' for $A^2 \cap H$. We have $A^2 \subseteq XA$, and every set of the form $xA \cap H$ is contained in $y(A^2 \cap H)$ for some $y \in xA \cap H$, then $(A^2 \cap H)^2 = A^4 \cap H \subset X'(A^2 \cap H)$ with $|X'| \leqslant K^3$. □

You can see the above proposition as a special case of the following we ask instead about the intersection of 2 approximate groups.

Proposition 2.3. *Let A be a K-approximate subgroup of G and B an K'-approximate subgroup, then $A^2 \cap B^2$ is a $(KK')^3$-approximate group.*

Proof. This time we will show $A^2 \cap B^2$ is contained in at most $K^3K'^3$ sets of the form $xA \cap yB$. We can follow the above argument twice to see that $(A^2 \cap B^2)^2 \subset A^4 \cap B^4$ is contained in at most $(KK')^3$ cosets of $A^2 \cap B^2$. □

This further solidifies our expectation that approximate subgroups should act like actual subgroups.

2.4.3 Subrings

These ideas can be extended to almost all algebraic objects, once we have approximate groups, approximate rings or fields seem an obvious next step. In some ways, the Erdős-Szemerédi theorem [49] (Theorem 2.1) is a statement about the lack of finite approximate subfields in the reals. That is an approximate subfield should have both additive structure and multiplicative structure and the theorem states that such sets do not exist. Further reading can be found in [180]. In this way of viewing the problem Bourgain, Katz, and Tao [14] essentially prove that $\mathbb{F}_p$ has no approximate subfields except those that are almost all the field or almost none of the field.

2.4.4 Moving from Group Theory to Approximate Group Theory

2.4.4.1 The Intuition

Stated in this way group theory is 1-approximate group theory. As such a valid question is which results can be generalised to arbitrary approximate groups and what changes when this happens? Group theory has many results for us to use and luckily the intuition is that many (this is stated of course with a warning that not all results do, that some results that should seemingly be simple to generalise do so with a bit of extra work which can later cause problems (for example the intersection of approximate subgroups need not be an approximate subgroup unlike the case of standard subgroups but the intersection of squares of subgroups are approximate subgroups), and finally the generalisation of a result is not always obvious in the way that there may be multiple equally valid options and the choice of which leads to non-trivial results is difficult) have an approximate equivalent.

This intuitively makes sense as we only relaxing (and not completely removing) one of the group axioms (that of closure) we still require our approximate groups to contain the identity and have inverses. Whilst we will not fully formalise this intuition (such formalisations can instead be found in the literature including [32], [17] and [190]) we will detail the results we use later in this thesis in the next subsection.

2.4.4.2 Sample Results

Some of these will be used later, particularly in Chapter 4. We start with the orbit-stabiliser theorem for sets, one of many results from group theory that can be adapted for approximate groups [75, Lemma 4.1]. Recall that if group G acts on a set X, then the stabiliser of an element x of X, is the subgroup $\mathrm{Stab}(x)$ of G consisting of all elements that fix x. The orbit $A(x)$ is the set $A(x) := \{g \cdot x : g \in A\}$.

Lemma 2.5 (Orbit-Stabiliser Theorem for sets)**.** *Suppose the group G acts on a set X, $x \in X$, and $A \subseteq G$ is finite. Then there exists a_0 in A such that*

$$|(a_0^{-1}A) \cap \mathrm{Stab}(x)| \geqslant \frac{|A|}{|A(x)|}, \tag{2.2}$$

and for all finite sets $B \subseteq G$,

$$|AB| \geqslant |\mathrm{Stab}(x) \cap B||A(x)|. \tag{2.3}$$

We often specialise Lemma 2.5 to the action of a group G on a subgroup H by left multiplication, so that the stabiliser of $H \in G/H$ is H itself and the orbit of H under a set A is AH/H, which is the number of distinct coset representatives in A. If $\psi : G \to G/H$ is the quotient map, then $A(H) = AH/H = \pi(A)$. The subgroup H does not need to be normal.

We use the following lemma when we wish to move from growth in a group to growth in a subgroup.

Lemma 2.6. *Suppose that H is a subgroup of G and that $A \subseteq G$ satisfies $|A^3| \leqslant K|A|$. If $B := A^{-1}A \cap H$, then*

$$|B^k| \leqslant |A_{(2k)} \cap H| \ll K^{6k-3}|B|.$$

Proof. By Lemma 2.5 Equation (2.2) where we have specialised to the action of a group G on a subgroup H by left multiplication, we have $|B| \geqslant \frac{|A|}{|A(H)|}$ (where $A(H)$ number of distinct cosets of H determined by elements of A). On the other hand, $B^k \subseteq (A^{-1}A)^k \cap H \subseteq A_{(2k)} \cap H$.

So we have that

$$|B^k| \leqslant |(A^{-1}A)^k \cap H| \leqslant |A_{(2k)} \cap H|.$$

By Lemma 2.5 Equation (2.3), taking A as A, and B as $A_{(2k)}$ we also have

$$|AA_{(2k)}| \geqslant |A_{(2k)} \cap H||A(H)|.$$

So, combining the above and using Lemma 2.2 to pass from $|A_{(2k)}|$ to $|A|$, we get that

$$|B^k| \leqslant |A_{(2k)} \cap H| \leqslant \frac{|AA_{(2k)}|}{|A(H)|} \ll K^{6k-3}\frac{|A|}{|A(H)|} \leqslant K^{6k-3}|B|.$$

□

2.5 Energy

Another way of considering how additively (and similar can be done for other operations) structured a set is to consider the number of solutions to

$$a + b = c + d. \tag{2.4}$$

We are going to have at least the trivial solutions $a = c$ and $b = d$. Considering our previous examples for additive structure of arithmetic progressions we are going to have a lot of solutions whilst in our example of a lack of such structure, geometric progressions, will only have the trivial solutions. This will lead to our intuition of lots of energy is equivalent to having some sort of structure.

We will define the additive energy of a set A as follows

$$E^+(A) := \left|\{(a,b,c,d) \in A^4 : a+b = c+d\}\right|.$$

We may also consider the energy between two sets

$$E^+(A,B) := \left|\{(a,b,c,d) \in A \times B \times A \times B : a+b = c+d\}\right|.$$

We can define other types of energy in the same way we change from sumsets to product sets. We also note that $E^+(A,B) = E^+(B,A) = E^+(A,-B)$.

There are many alternative ways of writing the above set definitions which can be more useful in calculations. These can be found in Lemma 2.9 of Tao and Vu [185] or proven via simple counting arguments. In particular we have

$$E^+(A,B) = \sum_x r^2_{A+B}(x) = \sum_x r^2_{A-B}(x). \tag{2.5}$$

We also have

$$E^+(A,B) = \sum_{x \in A+B} |A \cap (x-B)|^2 = \sum_{a \in A, b \in B} |A \cap (a+b-B)| = \sum_{a,a' \in A} |(B+a) \cap (B+a')|.$$

2.5.1 Trivial Bounds

Having now defined a quantity it is always useful to know how it behaves. We will start with some trivial bounds of

$$|A||B| \leqslant E^+(A,B) \leqslant \min\left(|A|^2|B|, |A||B|^2\right).$$

The first bound comes from we can choose a and b arbitrarily from A and B respectively and then choose $a = c$ and $b = d$. The second bound comes from noting that once a, b, and c is chosen, the $d = a + b - c$.

Having looked at what we can get trivially at the extremes we will next consider some examples.

Starting with our standard examples of arithmetic and geometric progressions, let A be the arithmetic progression $\{1,2,\dots,n\}$, from earlier in this section we have that $|A+A| = 2|A|-1$. the energy of A is maximal however so this provides an example with small sumset and large additive energy. We note this hold for other arithmetic professions just by scaling and possibly translating the set.

Moving on to the geometric progression, now take $A := \{r, r^2, r^3, \dots, r^n\}$ and recall we have $|A+A| = \binom{|A|}{2} \approx |A|^2$ but $E^+(A) = |A|^2$ which can be seen by considering the r-nary representation. This is thus an example of small additive energy and large sumset.

It can also be worthwhile knowing how easy the structure represented by low or high energy can be destroyed. As written above an arithmetic progression A has additive energy $|A|^3$. If we consider $B := (A \cup \{x\})$ where x an element which isn't part of the progression then $|B+B| = |A+A| + |A+x| + |x+A| = |\{x^2\} = 4|A|$ and so whilst not a minimal sumset, B still has small doubling however $E^+(B) \geqslant E^+(A) = |A|^3$, so we still have large additive energy.

Our next example will be a set with a large sumset and large additive energy, (the case of small sumset and small energy being unfeasible to the upcoming Lemma 2.7), consider the union of an arithmetic progression and a geometric progression. The sumset is then large, $|A+A| \geqslant \frac{1}{4}|A|^2$, as half the set is a geometric progression so contributes at least $\frac{1}{4}|A|^2$. The energy is also large as the arithmetic progression half contributes maximal energy. In particular, this shows that the implication that small energy implies large sumset holds but that the converse may not hold. A related but slightly more involved example due to Balog and Wooley [6] shows there exist sets where both the additive and multiplicative energies of any sufficiently large subset is non trivially large, unlike the prior example where we had two subsets each with one of additive and multiplicative energy large and the other nearly trivially small. To construct this example we start with $I = \{n^2, n^2+1, \dots, 2n^2-1\}$ and for our set we take

$$A = \bigcup_{i=0}^{n-1} 2^i I.$$

A has n^3 elements as it is the union of n distinct dilates of the interval I of size n^2. This example is a much smarter way of combining both an arithmetic progression, I, and a geometric progression (the dilates) than our previous example of a union. Starting with the multiplicative energy, consider $x \in AA$ then the number of ways of writing x as a product of two elements of A is $r_{AA}(x) \geqslant n$. This can be seen by seeing each element of A as the product of an element in I and a power of two so that $A \ni a = b2^k$, but then we have

$$x = aa' = bb'2^{k+k'} = bb'2^{(k+1)+(k'-1)} = \dots,$$

and the n choices should be clear from moving factors of two from a to a'. Hence

$$|AA| \leqslant n^5 = |A|^{5/3},$$

and by the forthcoming Lemma 2.7 (with multiplicative notation rather than the stated additive notation)

$$E^\times(A) \geqslant \frac{|A|^4}{|AA|} \geqslant |A|^{7/3}.$$

Note the same holds for any subset $A' \subseteq A$ of density r, that is $E^\times(A') \geqslant r^4|A|^{7/3}$. Moving now to the additive energy, we keep out notation for the subset A' and note that the additive energy of A' is at least the sum of n pieces of additive energies of A' intersecting $I, 2I, 4I, \dots$ and so (again using Lemma 2.7 for the second inequality)

$$E^+(A') \geqslant \sum_{i=0}^{n-1} E^+(A' \cap 2^i I) \geqslant \sum_{i=0}^{n-1} \frac{|A' \cap 2^i I|^4}{|A' \cap 2^i I + A' \cap 2^i I|}.$$

We are in the minimum case when the density of A' in each of $I, 2I, 4I, \dots$ is also r so

$$E^+(A') \geqslant \sum_{i=0}^{n-1} \frac{(rn^2)^4}{2n^2} \gg r^4 n^7 = r^4 |A|^{7/3}.$$

2.5.2 Trivial Results

We have the following bound on energy, the first two terms coming from the trivial bounds above. The third term comes from Cauchy-Schwartz which we shall show below.

$$E^+(A,B) \leqslant \min\left\{|A|^2|B|, |B|^2|A|, |A|^{3/2}|B|^{3/2}\right\}.$$

Without any loss of generality we may assume that one of the two sets is smaller, say for instance that $|B| \leqslant |A|$, then we have that $E^+(A,B) \leqslant |B|^2|A|$. Proving the third term,

$$\begin{aligned} E^+(A,B) &= \sum_{z\in(A-A)\cap(B-B)} |A\cap(z+A)||B\cap(z+B)| \\ &\leqslant \left(\sum_{z\in(A_A)\cap(B_B)} |A\cap(z+A)|^2\right)^{1/2}\left(\sum_{z\in(A_A)\cap(B_B)} |B\cap(z+B)|^2\right)^{1/2} \\ &= E(A)^{1/2}E(B)^{1/2} \leqslant |A|^{3/2}|B|^{3/2}. \end{aligned}$$

Another useful result is the connection between the cardinality of sumsets and energies once again shown by the Cauchy–Schwarz inequality to be the following lemma.

Lemma 2.7. *Let A and B be sets, then*

$$|A\pm B| \geqslant \frac{|A|^2|B|^2}{E(A,B)}.$$

Proof. Starting with the following definition of energy

$$E^+(A,B) = \sum_x r^2_{A+B}(x),$$

which we apply Cauchy-Schwartz to get

$$\sum_x r^2_{A+B}(x) \geqslant \frac{(\sum_x r_{A+B}(x))^2}{\sum_x \mathbb{1}^2_{A+B}} = \frac{|A|^2|B|^2}{|A+B|}.$$

Noting we could have instead started with

$$E^+(A,B) = \sum_x r^2_{A-B}(x)$$

completes the proof.

□

The above lemma can be found in multiple places (sometimes just with $A = B$) in the literature including [42, 162, 177, 185] as Remark 4.2 and Corollary 2.10 in the last two.

We may also wish to bound the energy of a union in terms of the energy of its pieces, in particular this is used in the proof of Theorem 4.9. The following lemma does that via two applications of Cauchy-Schwarz.

Lemma 2.8. *For $A = \bigcup_{i=1}^{X} A_i$, the union of X pieces, we have that the energy of A, is bounded by*

$$E\left(\bigcup_{i=1}^{X} A_i\right) \ll \left(\sum_{i=1}^{X} E(A_i)^{1/4}\right)^4.$$

Proof. Let X be the union of A_n's. The energy of X is defined as the number of solutions to the equation $a+b=c+d$ for $a,b,c,d \in X$. We consider

$$E_{\{i,j,k,l\}} = \left|\{a,b,c,d : a+b=c+d, a\in A_i, b\in A_j, c\in A_k, d\in A_l\}\right|.$$

As $X = \bigcup_n A_n$ then $E(X) \leqslant \sum_{i,j,k,l} E_{\{i,j,k,l\}}$ as for every solution to $a+b=c+d$ each of the variables have to be in at least one of the A_n's as they are in X however they could be double counted if the A_n's are not disjoint.

We next use Cauchy-Schwarz to bound $E_{\{i,j,k,l\}}$.

$$E_{\{i,j,k,l\}} = \sum_n r_{A_i+A_j}(n) r_{A_k+A_l}(n) \leqslant \left(\sum_n r^2_{A_i+A_j}(n)\right)^{1/2} \left(\sum_n r^2_{A_k+A_l}(n)\right)^{1/2}.$$

Now seeking to bound

$$\sum_n r^2_{A_k+A_l}(n)$$

which is the number of solutions of $a+b=a'+b'$ for $a,a'\in A_k$, $b,b'\in A_l$ or equivalently the number of solutions to $a-a'=b-b'$. Hence

$$\sum_n r^2_{A_k+A_l}(n) = \sum_n r_{A_k-A_k}(n) r_{A_l-A_l}(n)$$

and by another application of Cauchy-Schwarz

$$\leqslant E(A_k)^{1/2} E(A_l)^{1/2}.$$

Hence

$$E_{\{i,j,k,l\}} \leqslant E(A_i)^{1/4} E(A_j)^{1/4} E(A_k)^{1/4} E(A_l)^{1/4}.$$

We sum over everything to get back to X, giving

$$\sum_{i,j,k,l} E(A_i)^{1/4} \dots E(A_l)^{1/4} = \left(\sum_i E(A_i)^{1/4}\right)^4,$$

so $E(X) \ll \left(\sum_i E(A_i)^{1/4}\right)^4$. □

2.5.3 Higher Energies

So far we have introduced energy as the number of solutions of $a*b=c*d$ for some operation $*$. What is important about using four elements? Nothing particularly except it is the first non-trivial result after you note that the number on terms on each side of the equation must be balanced and the number of solutions to $a=b$ is trivial.

So-called higher energies increase the number of terms on each side. We take this definition from [152]

$$T_k(A) := \left|\left\{(a_1,\dots,a_k,a'_1,\dots,a'_k) \in A^{2k} : a_1+\dots+a_k = a'_1+\dots+a'_k\right\}\right|.$$

Such higher energies are introduced in depth by Schoen and Shkredov [146] and have been used and studied in several other papers including [151]. Particular examples of using third energies can be found in [147, 153], these are useful due to their association to collinear triples. We will also make use of higher energies later in this thesis, particularly in Chapter 5.

We note that the ideas used to provide trivial bounds for energy in Section 2.5.1 can be used again here. Explicitly they tell us that that $T_k(A) \geqslant |A|^k$ from taking $a_i = a'_i$ for all i and that these higher energies are bounded above by $|A|^{2k-1}$ which is seen by noting that by fixing all but the last term, this last term is now determined. It is also worth noting how higher energies of different k interact,

$$T_k(A) \leqslant |A|^{2(k-l)} T_l(A). \tag{2.6}$$

We can also gain a Lemma 2.7 equivalent that

$$|kA| \geqslant \frac{|A|^{2k}}{T_k(A)}.$$

There is also a second form of higher energies which come from considering how to generalise Equation (2.4). In the case of $T_k(A)$ we increased the number of variables on each side, however a second equation to consider is

$$a+b = c+d = e+f,$$

and the obvious continued generalisation of more equalities. This leads to the following higher energy definition.

$$E^+_k(A) := \left|\left\{(a_1,\dots,a_k,a'_1,\dots,a'_k) \in A^{2k} : a_1-a'_1 = a_2-a'_2 = \dots = a_k-a'_k\right\}\right|.$$

We note that another way of considering this is as the generalisation of Equation (2.5) as follows

$$E^+_k(A) = \sum (r_{(A+A)}(x))^k.$$

Using this definition also us to have non-integer k. We note that when $k=1$ this returns to just being a cardinality. During this thesis, we will make use of the $T_k(A)$ higher energies in Chapter 5 rather than this and so shall be brief. In [151], Remark 40 notes that a prior proof gives the following relation between these two higher energies as some sort of dual given that k is even

$$T_k(A)E_k(A) \geqslant \left(\frac{E_{3/2}(A)}{|A|}\right)^{2k}.$$

More dual relations can be found in [152]. An application of this other higher energy can be found in [158]. We also note that in [152] a further generalisation of these higher energies can be found, based on equations of the form

$$a_1+\dots+a_l = b_1+\dots+b_l = \dots = k_1+\dots+k_l,$$

that is k different sums of l elements all equal.

2.6 Expanders

In a similar way to moving to groups whose operation combine addition and multiplication, another potential avenue of attack to bound sets built from both operations such as the following two quantities

$$|A(A+A)| \quad \text{and} \quad |AA+A|.$$

Bounds for such sets are referred to as expander results, we prove several in Chapter 5 as corollaries.

Such bounds are also intricately linked with incidence geometry, our forthcoming Corollaries 5.2, 5.4, and 5.5 are expander style results that come from the new incidence results for hyperbola detailed in Chapter 5. Over the reals, the simple bounds come from the Szemerédi–Trotter theorem (Theorem 3.1 noted in Chapter 3). We make a note of the proof as it both provides the link between this chapter and the next as well as a framework that may be adapted for similar incidence results. To obtain this bound we consider a set of points $P = B \times (A+BC)$ and without loss of generality assume $|B| \leqslant |C|$, along with the set of lines L with slope in C and y intercept in A so $|L| = |A||C|$. We have the following incidence bound once we plug this example into the Szemerédi–Trotter theorem (Theorem 3.1).

$$\mathcal{I}(P,L) \ll |B|^{2/3}|A+BC|^{2/3}|A|^{2/3}|C|^{2/3} + |B||A+BC| + |A||C|.$$

And we note we have $\mathcal{I}(P,L) = |A||B||C|$ as each line has $|B|$ points on it. This gives us

$$|A||C||B| \ll |B|^{2/3}|A+BC|^{2/3}|A|^{2/3}|C|^{2/3}.$$

Thus we get the more general $|A||B||C| \ll |A+BC|^2$ and in particular $|A+AA| \gg |A|^{3/2}$.

Over finite fields we cannot just appeal to Szemerédi–Trotter, however Roche-Newton, Rudnev and Shkredov [128] show

$$|A+AA| \gg \min(|A|^{3/2}, p)$$

which matches Szemerédi–Trotter over $\mathbb{R}$. The p deals with cases of subfields, for example take $A = \mathbb{F}_p \subseteq \mathbb{F}_p$, then $A+AA = F_p$. This is one of a family of results of expanders over finite fields which are all bounded by $\gg \min(|A|^{3/2}, p)$, others include:

- $|AA+AA| \gg \min(|A|^{3/2}, p)$ (Rudnev [132]).
- $|(A-A)^2 - (A-A)^2| \gg \min(|A|^{3/2}, p)$ (Petridis [120], also see [26]).
- $|A(A+A)| \gg \min(|A|^{3/2}, p)$ (Aksoy-Yazici, Murphy, Rudnev, and Shkredov [200]).

The cutting edge result for $A+AA$ is due to Stevens and Warren [171] who proved, for $|A| \ |B|$, the slightly more general:

$$|A+AB| \gtrsim |A|^{3/2+3/170}.$$

This is an improvement on Roche-Newton and Warren [129].

Roche-Newton and Warren [129] also present a four-variable super quadratic expander

$$\left\{\frac{ac-db}{c-d} : a,b,c,d \in A\right\}.$$

Explicitly they prove this is $\gg |A|^{2+1/14}$ improving on the $\gg |A|^2$ of Murphy, Roche-Newton and Shkredov [112].Roche-Newton and Warren's methods stem from proving energy results for lines and noting that the line through points (a,b) and (c,d) has y-intercept $\frac{ac-db}{c-d}$. Continuing on the theme of expanders, Balog, Roche-Newton and Zhelezov [4] showed the existence of multiple six-variable super quadratic expanders. As may be expected, super quadratic growth is harder to manage with fewer variables, the two other known examples with four variables are due to Rudnev [131] and Shkredov [154].

Chapter 3

Incidence Results

This chapter seeks to ground the various incidence results which are used later in the thesis and provide the necessary background. In particular, as well as showing the connections with the previous chapter, we are building towards Chapter 5 where we present some new incidence results for hyperbolae. We will start with the basics of what is an incidence and what can be said simply. We will then explore results in this area which we will use as tools later and serve as intuition when we wish to move from the well-studied point and line incidences to the newer hyperbola incidences of Chapter 5. We will finish with the related distance problems, these will return in Chapter 5 where the connection between circles and hyperbolas (that is when -1 is a square they are equivalent) mean our results have implications for such distance problems.

3.1 Introduction to Incidence Geometry

Geometry and counting are among the oldest branches of mathematics[1] so it is unsurprising someone has tried to count geometric objects. One of the simplest ideas is how often does a collection of things touch. Taking lines as our example, it is clear that two straight lines (given standard Euclidean geometry) intersect in at most one point. How this bound increases given n lines is less obvious. It is a small generalisation to move to two different types of objects interacting for example points on lines. You can remove almost all the geometric underpinnings of this question, just keeping what it means for a point to be on a line. This remaining structure is called an incidence structure and finds itself in many areas of mathematics. For example, a hypergraph consists of a set of vertices (our points) and a set of hyperedges (our lines) consisting of subsets of the vertices, that is which points are on the given line. Other examples of incidence structures include block designs, projective and affine planes, and partial linear spaces. Due to the varied branches of mathematics in which these items turn up, there are multiple representations for incidence structures including; incidence matrices, Levi graphs, and pictorially (geometrically).

Incidence structures also exhibit a dual structure, that is given a structure of P points on a set of lines L we can consider each of our lines to be a point (so $l \in L$ corresponds to $p_l \in P'$) and

[1] The earliest records of geometry being from 5000 years ago in ancient Babylonia and the Indus Valley civilisation. The oldest example of tally marks for counting is between 25000 and 35000 years old.

take our lines as the sets of these p_l which correspond to lines which pass through a given point. This point-line duality can sometimes allow us to swap the role of points and lines when it is convenient.

3.1.1 What is an Incidence?

To understand incidence geometry it should be no surprise that you must first understand exactly what an incidence is. At first glance this may seem obvious, a point on a line is one, if a point is not on a line then it isn't. This view could then lead to a trivial understanding of at most the number of points and missing out all the potential multiplicities.

Formally we should see the set of incidences between a set of points P, and lines L, denoted $\mathcal{I}(P,L)$, as the number of pairs of a point and a line so that the point is on the line. That is

$$\mathcal{I}(P,L) := \left|\{(p,l) \in P \times L : p \in l\}\right|. \tag{3.1}$$

For those who prefer pictures or want intuition, Figures 3.1 and 3.2 should help.

(a) Zero incidences. (b) One incidence. (c) Two incidences.

Figure 3.1: Examples of 0,1 and 2 incidences involving a single point.

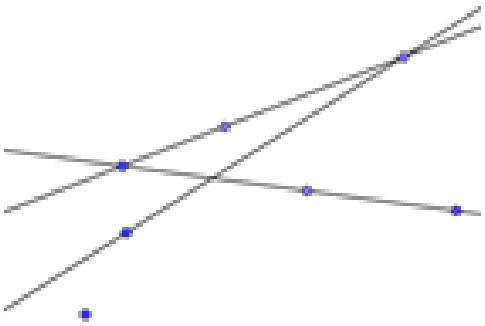
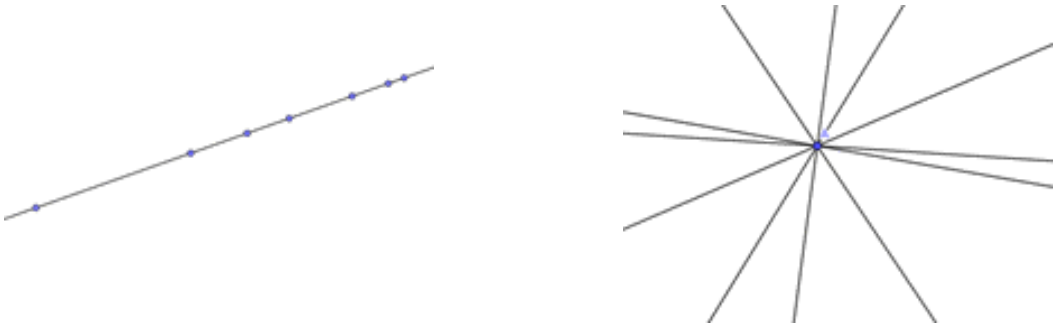


Figure 3.2: Eight incidences.

From this geometric intuition, the obvious extreme examples are lots of points on a single line and lots of lines through a single point. Note that these extremes are dual of each other.

We will consider the points distinct as the above definition (Equation (3.1)) tells us, although it is possible to consider the questions where the points or lines have weights assigned to them and can contribute more incidences which is the same as having multiple identical points or lines.



(a) A single line with all the points leading to $|P|$ incidences.

(b) A single point at the intersection of concurrent lines leading to $|L|$ incidences.

Figure 3.3: Extreme examples of incidences.

Although these pictures serve us well in the Euclidean setting where our standard understanding of points and lines allows our intuition to serve us well, it is the more formal Definition 3.1 that we will resort to in cases where our intuition generally fails such as in finite fields.

An alternate way of writing Definition 3.1 using representation functions, which can be more open to techniques such as Cauchy-Schwartz, is as follows:

$$\mathcal{I}(P,L) = \sum_{p\in P}\sum_{l\in L} \mathbb{1}_{p\in l}. \tag{3.2}$$

Another use of this way of considering incidences is that we can formulate a weighted version, that is we may want to decide certain incidences, points, or lines are more important and should count for more. To do this we set the functions $f(p)$ and $g(l)$ mapping points and lines respectively to their weights (a two-variable function could be used if individual incidence instead need to be assigned weights for an application) and then the weighted incidences can be calculated as follows

$$\sum_{p\in P}\sum_{l\in L} f(p)g(l)\mathbb{1}_{p\in l}.$$

As well as seeing this as a geometric object it can be seen in a few other ways. One such as a hyper-graph with the points being represented as the vertices and the lines as the hyper-edges. Taking Figure 3.2 as our example, we can represent this arrangement as the hyper-graph on the set of points $\{1,2,\ldots,7\}$ (numbered top-down although you could relabel as desired) with hyper-edges $\{1,2,3\}$, $\{3,4,5\}$ and $\{1,6\}$.

As well as different terminologies, there are several representations available. Geometric pictures such as Figure 3.2 may also be encoded in a $|P|\times|L|$ matrix with the element in the i^{th} row and j^{th} column if the i^{th} point is on the j^{th} line. Another representation is that of a Levi graph, a bipartite graph whose vertices are split into the points and the lines with edges connecting a point vertex and a line vertex if they are incident.

Before moving on we note the existence of a dual structure. That is given our arrangement of points and lines we can instead consider each line as a point and drawing a line connecting these

new points if they intersected with a point. This corresponds to the dual of the incidence matrix representation of incidences and swapping the labelling of the two subsets of vertices in a Levi graph. It is also important to be aware of the dimension of the problem you are working in, the plane has the principle of plane duality which states that moving to the dual of a theorem provides another valid theorem in the dual, however in higher dimensions the dual of a point line incidence theorem can be an incidence result between different objects such as hyperplanes. Explicitly in dimension n, the dual of objects of dimension r have dimension $n-1-r$ (so codimension $r+1$). This explains why, in higher dimensions, some bounds are not symmetric in $|P|$ and $|L|$, see for example [88].

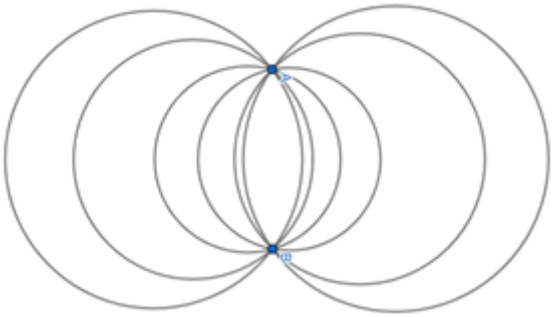
3.1.2 Trivial Results

Before seeking to prove the better results, it is a good idea to first consider what is trivial to say. This both gives a result to beat but can often highlight some of either the awkward cases or common pitfalls. For the following, it may be helpful to have a preferred example in mind, such as points and lines in the Euclidean plane which is the current example I will be taking, in part to allow simple diagrams and intuition to provide explanations.

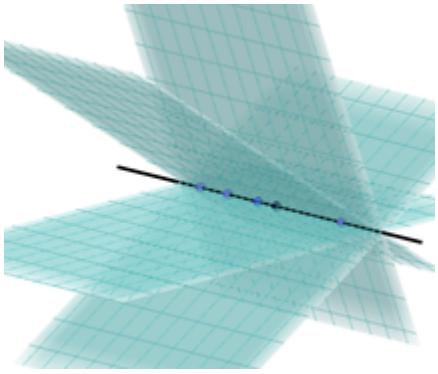
Starting with the most trivial of bounds, it is definitely the case that

$$\mathcal{I}(P,L) \leqslant |P||L|.$$

This is because every point can at the very most be on every line. However, with just a bit of thought, it can also be seen that this bound is only obtained in degenerative cases of a single point or a single line such as in Figures 3.3a and 3.3b. This is due to a line being defined by two points. We will have to deal with such trivial cases in future theorems by adding $+|P|+|L|$. Also of note, in some other examples, these degenerate cases happen slightly more often, such as between points and circles there can be up to two points as three points determine a circle (see Figure 3.4a and note we have similar for hyperbola) or between points and planes if all the points are on a line the bound can be met for any number of points and planes, (see Figure 3.4b).



(a) Many circles can intersect in two points.



(b) Many planes can intersect a line of points.

Figure 3.4: Examples of potential difficulties with other geometric objects.

Returning to the Euclidean plane and incidences between points and lines we have a lot of cases (indeed most) where the really trivial bound of $|P||L|$ is not tight. We can however come across another slightly less trivial bound stemming from the Cauchy-Schwartz inequality.

Lemma 3.1 (Trivial Incidence Bound)**.** *Given a set of points, P, and a set of lines, L, the number of incidences is bounded by*

$$\mathcal{I}(P,L) \leqslant \min\Big(|L||P|^{1/2} + |P|, |P||L|^{1/2} + |L|\Big).$$

Proof. For ease of notation, set $\mathcal{I} = \mathcal{I}(P,L)$ and by applying Cauchy-Schwartz to Equation (3.2) we have

$$\begin{aligned} \mathcal{I}(P,L)^2 = \mathcal{I}^2 &= \left(\sum_{p\in P}\sum_{l\in L} \mathbb{1}_{p\in l}\right)^2 \\ &\leqslant \sum_{p\in P} 1 \cdot \sum_{p\in P}\left(\sum_{l\in L}\mathbb{1}_{p\in l}\right)^2 \\ &= |P| \sum_{p\in P}\sum_{l\in L}\sum_{l'\in L} \mathbb{1}_{p\in l}\mathbb{1}_{p\in l'}. \end{aligned} \tag{3.3}$$

We split the sum up into two cases, that when $l = l'$ and when $l \neq l'$. The first case is just the number of incidences, $\mathcal{I}$ again. For the second case, we rearrange the order of summation and consider

$$\sum_{l\in L}\sum_{l'\in L}\sum_{p\in P} \mathbb{1}_{p\in l}\mathbb{1}_{p\in l'}.$$

As two distinct lines meet at a single point

$$\sum_{p\in P} \mathbb{1}_{p\in l}\mathbb{1}_{p\in l'} \leqslant 1,$$

and thus

$$\sum_{l\in L}\sum_{l'\in L}\sum_{p\in P} \mathbb{1}_{p\in l}\mathbb{1}_{p\in l'} \leqslant |L|^2.$$

Plugging this back into Equation (3.3) we get

$$\mathcal{I}^2 \leqslant |P|(I + |L|^2), \tag{3.4}$$

so $\mathcal{I} \ll |P|$ if the first term dominates or $\mathcal{I} \ll \sqrt{|P||L|^2}$ if the second term dominates. We note that whilst this does not keep our inequalities strict and instead falls back to asymptotic notation for ease it illustrates the method. we shall provide the full calculation below Finally, we note that we are done as we could have instead applied the Cauchy-Schwartz the other way around and two points define a line or indeed appeal to duality. Having now completed the proof we provide the

full calculation to continue from Equation (3.4), this essentially involves completing the square.

$$\begin{aligned}
\mathcal{I}^2 - |P|\mathcal{I} &\leqslant |P||L|^2 \\
\left(\mathcal{I} - \frac{|P|}{2}\right)^2 &\leqslant |P||L|^2 + \frac{|P|^2}{4} \\
\mathcal{I} &\leqslant \frac{|P|}{2} + \sqrt{|P||L|^2 + \frac{|P|^2}{4}} \\
\mathcal{I} &\leqslant |P| + \sqrt{|P||L|^2}
\end{aligned}$$

This completes the proof without the requirement of asymptotic notation.

□

Of note is that when $|P| = |L| = n$ this is a bound of $n^{3/2}$ and it is this 3/2 which would like to be improved. This raises the question of just how far can we improve it? An example (taken from Tao's blog [181], originally due to Elekes) shows that we cannot do better than 4/3 (in a field of characteristic 0, in characteristic p we must also restrict the size of N so that $N < \left(\frac{p}{2}\right)^{3/2}$ else we manage to fill up a whole field and thus don't grow as expected). This matches the exponent of the Szemerédi-Trotter Theorem which is stated below as Theorem 3.1. We construct the example below as Example 3.1 and provide a second example later.

3.1.3 Survey of Non-Trivial Results

Whereas our trivial bound of Lemma 3.1 works for any field (and indeed any dimension) being purely combinatorial in nature, most bounds are affected by the field. Over the real numbers, the Szemerédi-Trotter theorem [176] (proven in 1983), up to constants, is the best we can expect. However, all of its proofs (we will touch on some of these briefly below) rely on the topology of the reals. We do not have this in other fields and so while progress has been made in the case of finite fields, the question remains open to reaching the expected best possible bounds. In the complex numbers Tóth [191] and Zahl [202] have succeeded in reaching the best possible bound (that is the exponents match Szemerédi-Trotter), generalising the Szemerédi-Trotter theorem. As such we will consider the different cases depending on which field separately when discussing the historical progress towards these bounds.

We also note we have been vague on what exactly we want as a line, we may in fact consider other objects such as planes, curves or circles among others. We provide a few results of incidence results between objects other than just points and lines in the next section, some of which will be used later in the proofs of our own work. In Chapter 5 we shall prove results where our lines are curves, in particular, hyperbolae.

A final comment, although most of the results are about what happens in the plane, the same question can be asked about higher dimensions, although there then arises the problem of ensuring you are truly in a higher dimension case and do not have, for example, a plane embedded in three dimensions and as such are bounded by the same obstructions as the two-dimensional

problem. Some papers that consider higher dimensions include [2, 37, 88, 163] and we will comment further on this topic below.

3.1.3.1 Reals

Starting first with the real case as it is the most intuitive and easiest to visualise, we collect some of the important results in incidence geometry. Starting chronologically, Gallai solved (although Melchoir had proved an equivalent a few years earlier) a problem of Sylvester which lead to the Sylvester-Gallai theorem which states that in the Euclidean plane, for every finite point set it is possible to find a line that passes through exactly two points or passes through all of the points (harking back to one of our problem cases in our trivial incidence results). Erdős noted this tells us that a set of n points (not all on the same line) determines at least n lines. This is then a specific case of the De Bruijn–Erdős theorem, which provides a lower bound for the number of lines determined by n points in a projective plane. Green and Tao [63] provide a more robust version.

For us, perhaps the most prominent incidence result, and indeed a result that many others are compared to in this area of research is the Szemerédi-Trotter theorem. It provides the best possible (up to constants) answer in the reals to the initial question of this Chapter; how many incidences are there between given sets of points and planes? Starting with the statement of the result we have the following.

Theorem 3.1. *[Szemerédi-Trotter][176] Let $P \subset \mathbb{R}^2$ be a set of points and L a set of lines in the real plane, then*

$$\mathcal{I}(P,L) \ll |P|^{2/3}|L|^{2/3} + |P| + |L|.$$

This result has been proved in various ways by multiple people. The first proof due to its namesakes, Szemerédi and Trotter [176] used cell decomposition, a technique also used by [31] where the plane is split up into cells with the understanding that outside of edge cases only a fraction of lines passes through a given cell. László Székely [174] instead proved this theorem making use of the crossing number of graphs in a simpler proof. All of these proofs rely in some manner on the topology of the reals to work, that is if we instead try to prove the same bound as in Theorem 3.1 for finite fields rather than the reals we find it is false.

It is worth noting that in general the Szemerédi-Trotter theorem cannot be improved upon (bar the constant). This is clear from the following example due to Elekes.

Example 3.1. *Consider a set of points*

$$P = \left\{(a,b) \in \mathbb{Z}^2 \,:\, 1 \leqslant a \leqslant N; 1 \leqslant b \leqslant 2N^2\right\},$$

and a set of lines

$$L = \left\{(x, mx+b) \,:\, m,b \in \mathbb{Z}; 1 \leqslant m \leqslant N; 1 \leqslant b \leqslant N^2\right\}.$$

Note that $|P| = 2N^3$, $|L| = N^3$ and each line passes through N points (from taking $x \in \{1,\dots,N\}$) so $N|L| = N^4 \leqslant \mathcal{I}(P,L) \ll N^4$.

A second example due to Erdős, and rediscovered by Edelsbrunner and Welzl [38], exists. We shall detail it as Example 3.2, once we have introduced k-rich lines and the corresponding restatement of Theorem 3.1, in Subsection 3.1.4.

The secondary terms in Theorem 3.1 correspond to the degenerative examples from above, where all the lines intersect at a single point and where every point is on a single line.

Although the exponents are sharp the constant is not. Improvements in the constant for the crossing lemma lead to corresponding improvements in these constants. Pach, Radoičić, Tardos, and Tóth [116] proved a constant of 2.5 for Szemerédi-Trotter (the original showed that the constant 10^{60} worked) whilst Pach and Tóth [118] showed the constant must be greater than 0.42 using a variant of Example 3.2. The best known constant in the crossing number inequality is currently due to Ackerman [1]. The Szemerédi-Trotter theorem has been generalised to several other cases, in particular over the complex numbers by Toth [191] and Zahl [202] who also proved results between two-dimensional algebraic surfaces and a Szemerédi-Trotter type theorem for complex unit circles in $\mathbb{C}^2$. A further generalisation of Szemerédi-Trotter to higher dimensions, $\mathbb{R}^d$ was found by Agarwal and Aronov [2]. They state that for n points and m hyperplanes which are spanned by the pointset, then the number of incidences between the points and hyperplanes is bounded by

$$O\left(m^{\frac{2}{3}}n^{\frac{d}{3}} + n^{d-1}\right).$$

Edelsbrunner [37] has a construction that lets us see that this bound is optimal asymptotically. Solymosi and Tao [163] provide a bound for incidences between points and k-dimensional varieties of bounded degree in $\mathbb{R}^d$.

A consequence of Theorem 3.1, that is a famous result in its own right, is Beck's theorem, although the result was released at the same time as Szemerédi and Trotters (indeed earlier in the same journal). Beck's theorem settled a conjecture of Erdős in the positive and can also be seen as a variation of an older result due to Kelly and Moser [84]. Beck's theorem [7] states given a finite number of points in the plane you are in one of two extremes, that a large fraction of the points are on one line, or many lines are needed to connect the points. Stating this explicitly gives the following.

Theorem 3.2. *[Beck's Theorem][7] Given n points in the plane, there exist positive constants, c, k such that either there exists a line with $\frac{n}{c}$ of the points or at least $\frac{n^2}{k}$ lines each containing at least 2 points.*

It has been noted by Elekes and Toth [43] that it is not a simple matter to extend this result to higher dimensions. Indeed consider the example in $\mathbb{R}^3$ shown in Figure 3.5 of 2 skew lines (that is non-intersecting and non-parallel) with n points on each. These $2n$ points only span $2n$ planes and so you cannot simply replace the plane with space and lines by planes. If instead, you remain interested in points and lines in higher dimensions then you can project to a random plane. Higher dimensional versions have been proven for example by Thao [34].

We note that we prove a Beck's theorem style result (Corollary 5.9) in Chapter 5 for Möbius transforms rather than lines.

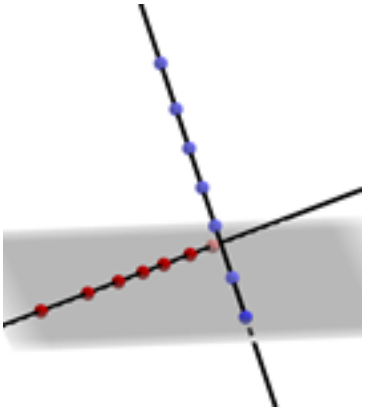


Figure 3.5: Two skew lines taking $n = 7$.

3.1.3.2 Finite Fields

Having looked at fields of characteristic zero in the previous subsection, we will now move to consider finite fields. These bring in a new difficulty due to similar issues that the existence of finite subfields did in Chapter 2. Recalling from Subsection 3.1.2 that given n points and n lines we have a trivial bound of $n^{3/2}$ and it is this exponent we wish to beat in the manner of the Szemerédi-Trotter theorem (Theorem 3.1) which obtains $n^{4/3} = n^{3/2-1/6}$, we will now provide examples where we cannot expect to do any better.

For our first example, we consider the plane over $\mathbb{F}_p$ and take our point set, P as the entire plane, that is p^2 points. Similarly, we take all p^2 lines as our set of lines L. We note that each line is incident to p points so $\mathcal{I}(P,L) = p^3 = |P|^{3/2} = |L|^{3/2}$, which is the trivial estimate and so in this instance we cannot do better. We note as a second example we can embed the previous example as a subplane into a larger plane over $\mathbb{F}_q = \mathbb{F}_{p^n}$. This reinforces the dependence is on p rather than $q = p^n$. As such, these and similar cases must be ruled out for progress to be made. We note that one way to deal with these examples is to consider two separate regimes, where $|P| = |L| \leqslant p^c$ and when $|P|$ or $|L|$ are large with respect to p. Here we will focus on the first, covering the second in more detail in Section 3.1.6.

Having identified some of the difficulties stopping us from simply generalising Theorem 3.1 to finite fields, we survey some of the historical progress made towards this goal. Bourgain, Katz and Tao [14] made progress using their sum-product estimate to gain an epsilon on the trivial 3/2. A quantitative bound was given by Helfgot and Rudnev [76] in prime fields showing that you can take epsilon equal to 1/10678, derived from a Beck's theorem (Theorem 3.2) equivalent that states that for a point set $P = A \times A$ that has n elements, $n < p$, taken in the plane over $\mathbb{F}_p$, P determines $\geqslant cn^{1+1/267}$ lines. Jones [80], at a slight worsening of the epsilon, showed that over a general finite field $\mathbb{F}_q$, $\mathcal{I}(P,L) \leqslant cn^{3/2-1/12838}$, given the required restriction of "antifield", that is ensuring that under projections the sets are not too close to being in a subfield. Further work includes results from Grosu [64], who show that for sufficiently small subsets of $\mathbb{F}_p$ you can treat the finite fields like the reals with respect to the Szemerédi Trotter theorem and thus get the same optimal exponent of four thirds. Kollár [88] considers dimension three analogues of the

finite plane providing a bound on the incidences between m lines and n points of $mn^{2/5}$ (note that in space rather than the plane we cannot swap the role of the points and the lines). He also proves that, over the real or complex numbers, the same holds but with a bound of $mn^{1/3}$ instead.

As stated work has also been done in the "large set" case (with respect to p), some such papers include [122, 188, 194, 195]. These will be considered in more detail later.

The following incidence bound on points and planes is due to Rudnev [134] and is both used here as well as in our later proofs.

Theorem 3.3. *[134] Let P be a finite set of points in $\mathbb{F}_p^3$ and Π be a finite set of planes in $\mathbb{F}_p^3$, with $|P| \leqslant |\Pi|$. Let k be the maximum number of collinear points in P. Then*

$$\mathcal{I}(P,\Pi) \ll |P|^{1/2}|\Pi| + k|\Pi| + \frac{|P||\Pi|}{p}.$$

Over general fields the result holds without the last term given an extra constraint of $|P| < p^2$ if the field has a positive characteristic p.

Which, whilst not truly fitting here as a point line incidence theorem is used to prove, along with Cauchy-Schwarz, the following point line incidence theorem due to Stevens and de Zeeuw [172]. We will also make use of these incidence theorems during Chapter 5. Theorem 3.3 has also lead to many other advances in this area of mathematics. It is worth noting that this point plane theorem is optimal, this can be seen by several examples including the three dimensional integer lattice where $P = \{1,\dots,n^2\} \times \{1,\dots,n\} \times \{1,\dots,n\}$ and Π is the set of n^2 planes with fixed x coordinate taken from $\{1,\dots,n^2\}$. Each plane then contains n^2 points and there are at most n^2 collinear points in P so $\mathcal{I}(P,\Pi) = n^4 = |P|^{1/2}|\Pi| = k|\Pi|$. Another example comes from considering a unit sphere in $\mathbb{F}_p^3$. Also of interest, if we had a Beck's theorem for $\mathbb{F}_p^2$, Theorem 3.3 would follow using Cauchy-Schwarz. This is detailed in the appendix of [121].

We now state Stevens and de Zeeuw's point line incidence theorem. Note that it holds for arbitrary fields (with a constraint on the characteristic of the field if it is positive) but we state it for $\mathbb{F}_p$ as this is how we use it.

Theorem 3.4. *[172] Let $A,B \subset \mathbb{F}_p$ with $|A| \leqslant |B|$ and let L be a set of m lines in $\mathbb{F}_p^2$, assume that $|A|m \leqslant p^2$. Then the number of incidences between the point set $P = A \times B$ and L satisfies the bound*

$$|\mathcal{I}(P,L)| = |\{(u,l) \in P \times L : u \in l\}| \ll |A|^{3/4}|B|^{1/2}m^{3/4} + |A||B| + m.$$

3.1.4 k-Rich Objects

In applications, we also find sometimes that we care more about special objects, in the case of incidences between points and lines this is often the case that we care about either points which have lots of lines or lines containing lots of points. We define a k-rich line as a line containing at least k points and a k-rich point in a like manner.

3.1.4.1 k-Rich lines

Again I will write about lines although if it is your preference these 'lines' could instead be planes or curves.

We can swap between an incidence bound a bound on the number of k-rich lines by

$$k|L_k| \leqslant \mathcal{I}(P, L_k).$$

We will make use of this alternative way of seeing incidence bounds in later chapters, specifically, we will use the upcoming incidences results for lines over finite fields (Lemma 3.2 and Corollaries 3.1 and 3.2 in particular) in Chapter 5 as tools to prove our new hyperbola incidences. As an example, we can rephrase Szemerédi-Trotter as the number of k-rich lines is at least

$$O\left(\frac{n^2}{k^3} + \frac{n}{k}\right). \tag{3.5}$$

This is done in Szemerédi and Trotters original paper [176].

Having now defined what we mean by k-rich lines and stated the Szemerédi-Trotter theorem in this regime we are in the right position to detail the second example (due to Erdős) as to the optimally of this theorem which we referred to in Subsection 3.1.3.1.

Example 3.2. *Consider a $\sqrt{n} \times \sqrt{n}$ lattice of points, $P := \{(a,b) : 0 \leqslant a, b < \sqrt{n}\}$. We will consider all possible k-rich lines, and denote the number of these lines by $m(k)$. Note that we can simplify this to considering k-rich lines which pass through the origin (or indeed any point), denoting the number of such lines $m_O(k)$, as $m(k) = \frac{|P| m_O(k)}{k}$. We bound*

$$m_O(k) \gg \left(\frac{\sqrt{n}}{k}\right)^2$$

by noting that a k-rich line (in the $\sqrt{n} \times \sqrt{n}$ lattice) passes through a point in the $\frac{\sqrt{n}}{k} \times \frac{\sqrt{n}}{k}$ lattice and we are double counting though points whose coordinates aren't coprime (the line through $(2,2)$ has already been counted when it passed through $(1,1)$), this leaves us to count the number of pairs (a,b) such that a is coprime to b and both $0 \leqslant a, b < \frac{\sqrt{n}}{k}$. This is the sum of the Euler function,

$$\sum_{k=1}^{n} \varphi(k) = \frac{3}{\pi^2} n^2 + O\left(n (\log n)^{\frac{2}{3}} (\log\log n)^{\frac{4}{3}}\right),$$

where the bound comes from [196]. This gives $m_k \gg \frac{n^2}{k^3}$ which when compared to Bound (3.5) shows Theorem 3.1 is optimal up to constant.

We note we could also provide a bound for the number of k-rich points if desired. Similarly Agarwal and Aronov's result [2] for incidences between points and hyperplanes can be rephrased as a bound on k-rich hyperplanes (those containing k or more points) of

$$O\left(\frac{n^d}{k^3} + \frac{n^{d-1}}{k}\right).$$

We can do similar with other incidence results, when we consider Stevens and de Zeeuw's incidence result, Theorem 3.4 (restricting to $P = A \times A$ for ease with $|A| = n$) we get a bound for the number of k-rich lines (in the finite fields setting) of

$$O\left(\frac{n^5}{k^4} + \frac{n}{k}\right).$$

When specialised to $\mathbb{F}_p$ the Stevens - de Zeeuw theorem's corollary from [111] provides the slightly refined Lemma.

Lemma 3.2. *Let $A \subset \mathbb{F}_p$ and let $2|A|^2/p \leqslant k \leqslant |A|$ be an integer that is greater than 1. The number, l_k, of k-rich lines satisfies*

$$l_k \ll \min\left(\frac{p|A|^2}{k^2}, \frac{|A|^5}{k^4}\right).$$

This Lemma will be used later in Chapter 5. We will also use the following different corollary of Theorem 3.4 due to Stevens and de Zeeuw [172] in a different part of Chapter 5.

Corollary 3.1. *Let P be a set of points in $\mathbb{F}_p^2$ such that $|P| \leq p^{15/13}$, and let L_k be the set of lines which pass through at least $k \geq 2$ points of P. Then we have*

$$|L_k| \ll \frac{|P|^{\frac{11}{4}}}{k^{\frac{15}{4}}} + \frac{|P|}{k}.$$

Another result due to Stevens and de Zeeuw we will use later in Chapter 5 is the following asymmetric incidence result. This statement (and the proof of the result) is in Stevens thesis [170][Theorem 4.8], the chapter it is from based on [172].

Lemma 3.3. *Let $A \times B \subseteq \mathbb{F}^2$ be a finite set of points with $|A| \leqslant |B|$ and let $|A|^{1/2} \leqslant k \leqslant |A|^{1/2}|B|^{1/2} + |A|^{2/3}|B|^{1/3}$ be an integer. Let L_k be the set of k-rich lines with respect to $A \times B$. In positive characteristic p, assume that $\frac{|A|^3|B|^2}{k^2} \leqslant p^2$. Then*

$$|L_k| \ll \frac{|A|^3|B|^2}{k^4}.$$

In our application, we will take our field as $\mathbb{F}_p$ and so we trivially meet the $\frac{|A|^3|B|^2}{k^2} \leqslant p^2$ criterion. Outside of the acceptable range for k we will use Cauchy Schwarz, precisely we will use the form found in Bound (3.5):

$$O\left(\frac{|A|^2|B|^2}{k^3} + \frac{|A||B|}{k}\right).$$

In particular when $k > |A|^{1/2}|B|^{1/2}$ the second term of the Cauchy-Schwartz bound dominates as $\frac{|A|^2|B|^2}{k^3} < \frac{|A||B|}{k}$ so we will use this in this case. In the small k case note that the number of k-rich lines must be less than the number of lines defined by the point set which is itself smaller than $(|A||B|)^2 \ll \frac{|A|^3|B|^2}{k^4}$ for small k and so this is subsumed into the first term. More thoroughly, starting with Theorem 3.4 we have

$$k|L_k| \leqslant \mathcal{I}(A \times B, L_k) \ll |A|^{3/4}|B|^{1/2}|L_k|^{3/4} + |L_k| + |A||B|.$$

So as long the k is greater than the implied constant we may rearrange to get the bound we want, and when k is smaller we apply the argument from above for the small k case. Combining these extra cases into the above Theorem 3.3 provides the corollary we will use later in Chapter 5.

Corollary 3.2. *Let $A \times B \subseteq \mathbb{F}_p^2$ be a finite set of points with $|A| \leqslant |B|$ and $k \geqslant 2$, then the number of k-rich lines is bounded by*

$$|L_k| \ll \frac{|A|^3|B|^2}{k^4} + \frac{|A||B|}{k}.$$

3.1.5 Non-Point-Line Incidence Results

As well as considering different fields, we can consider incidences between different objects. One direction is to consider curves rather than lines, another is to consider higher-dimensional objects such as planes. Chapter 5 will consider hyperbolae in more depth. The motivation for moving to quadratic curves, and particularly the two-dimensional families like translates of a given curve (taking the curve as the hyperbola $xy = 1$ provides the setting for much of Chapter 5), is that they naturally come next. That is they still intersect another curve by a finite number of points (two rather than one) and three points define a curve. The case of quadratic curves splits into circles, hyperbolae and parabolae and these subcases are distinct and different from each other. Starting with the translates of parabola, $(x-a)^2 = y-b$, these are in some sense the same as point-line incidences as after a change of variables $(a,b) \mapsto (a, b+a^2) = (a', b')$ and $(x,y) \mapsto (x, y-x^2) = (x', y')$, the equation becomes $y' = -2ax' + b'$. As such Example 3.1 also gives the optimal lower bound for the number of incidences between points in the rectangle and the translates of the parabola $y = x^2$. For translates of the unit circle and hyperbola however we suspect much stronger bounds should exist, these bounds are connected to the Erdős unit distance conjecture which we shall discuss in Section 3.3.1.

As to be expected there are differences and new difficulties compared to the point-line examples else we would just port over the results. Perhaps one of the most obvious is that our line equivalents may no longer be limited in meeting in just one point, for example, two different circles can meet in two places and planes meet in a line. Another way to consider this is that lines are defined by two points, whereas over objects require more (circles and hyperbola need three points on them, a plane needs three not all on a line and so on). As well as difficulties though, we also get new applications and the tools associated with the new applications. Many of these objects can be associated with matrix groups and the growth of these groups. Circles (and thus hyperbolae in the case where $i = \sqrt{-1} \in \mathbb{F}$ as then $x^2 + y^2 = x^2 - (iy)^2$) have links to the Erdős distance problem (see Subsection 3.3).

Conducting a short survey of results of this nature include Pach and Sharir's [117] work on incidences between points and curves. They show that both the number of incidences between m points and n anchored unit circles in $\mathbb{R}^3$, and the number of tangencies between m directed points and n arbitrary circles in the plane, is $O(m^{3/5}n^{3/5} + m + n)$. A similar bound is also derived for curves of *almost two degrees of freedom*. That is a set of curves C such that any pair of curves in C has $O(1)$ intersections with each other, similarly any pair of points only has $O(1)$ curves passing through them and finally if a curve passed through a pair of points p, q there is a polynomial of

constant degree, $F(x,y)$, associated with the problem such that the points p and q have a curve passing through them if and only if $F(p,q)=0$. This links into what we were saying above that lines meet in on point and so only 1 line for a given set of two lines. Another example is the family of unit circles in $\mathbb{R}^3$ that pass through a given fixed point. Pach and Sharir bound for the incidences between m points and n of these more general curves is again $O(m^{3/5}n^{3/5}+m+n)$ with some added extra terms depending on how many curves or dual curves lie on an infinitely-ruled surface.

Cilleruelo, Iosevich, Lund, Roche-Newton, and Rudnev [30] prove an incidence result for points and spheres in general finite fields. For the point set P and a set of spheres, S in $\mathbb{F}_q^d$, the bound is

$$\frac{|P||S|}{q} - |P|^{1/2}|S|^{1/2}q^{d/2} < \mathcal{I}(P,S) < \frac{|P||S|}{q} + |P|^{1/2}|S|^{1/2}q^{d/2}.$$

An explicit application is a Beck's theorem (Theorem 3.2) equivalent which says given a point set $P \subseteq \mathbb{F}_q^2$ with $|P| \geqslant 5q$, then the set P determines a positive proportion of all circles. This is optimal up to multiplicative constants.

Continuing the theme of incidence results about circles, this time over $\mathbb{C}^2$, Zahl [202] shows that for the number of point-circle incidences between a pointset P and a set of unit circles S is bounded by $O(m^{2/3}n^{2/3}+m+n)$ in the flavour of Theorem 3.1. This is a special case of an incidence result that bounds the incidences of m points and n two-dimensional algebraic surfaces in $\mathbb{R}^4$ by at most $O(m^{\frac{k}{2k-1}}n^{\frac{2k-2}{2k-1}}+m+n)$, given that the algebraic surfaces behave like pseudoflats with k degrees of freedom, and that $m \leqslant n^{\frac{2k+2}{3k}}$. Other special cases Zahl proves include the number of incidences between a pointset $P \subset \mathbb{R}^4$ of size m and n planes in $\mathbb{R}^4$, with the restrictions that any two of the planes meet in at most one point and $m \leqslant n$, is $O(m^{2/3}n^{2/3}+m+n)$, and another proof of the Szemerédi-Trotter theorem over complex numbers originally due to Toth [191]. Clarkson, Edelsbrunner, Guibas, Sharir, and Welzl [31] proved several bounds for curves, circles and spheres. We also note the point plane theorem due to Rudnev [134] (Theorem 3.3) already stated above fits into this section.

Finally, we draw attention to point hyperbola incidences, stating the following results of Shkredov and Bourgain which serves as some of the inspiration for Chapter 5 and will be considered in more detail there as well as compared to each other. Bourgain [10] proved the following qualitative theorem.

Theorem 3.5. *For all $\varepsilon > 0$, there is $\delta > 0$, as follows. Let $A \subset \mathbb{F}_p$, $H \subset SL_2(p)$ satisfy the conditions: $1 \ll |A| < p^{1-\varepsilon}$, $|H| > |A|^{1+\varepsilon}$, and $|H \cap gS| < |H|^{1-\varepsilon}$ for any proper subgroup $S \subset SL_2(\mathbb{F}_p)$ and $g \in SL_2(\mathbb{F}_p)$. For $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{F}_p)$, let $h \in H$ be identified with the curve $cxy - ax + dy - b = 0$. Then the number of incidences*

$$\sigma(A,H) < |A|^{1-\delta}|H|.$$

Shkredov [157] proved a quantitative theorem.

Theorem 3.6. *[Theorem 1][157] Let $A,B,C,D \subseteq \mathbb{F}_p$ be sets. Then for any $\lambda \neq 0$, one has*

$$|\{(a+b)(c+d) = \lambda : a \in A, b \in B, c \in C, d \in D\}| - \frac{|A||B||C||D|}{p} \lesssim$$

$$\lesssim |A|^{1/4}|B||C||D|^{1/2} + |A|^{3/4}(|B||C|)^{41/48}|D|^{1/2}.$$

These theorems serve as inspirations to the author and Rudnev's [140] Theorem 5.3, another hyperbola incidence which will be covered in greater detail in Chapter 5.

3.1.6 Statistical Terms

So far we have been considering these incidence results asymptotically, the case of "p large enough". This has allowed us to ignore annoying cases where we have too many things, similar to in Chapter 2 where growth falls apart when we are too close to a subfield. When we have almost every line and point we are going to get close to the maximum number of incidences.

In this subsection, we will comment on what can be done in these cases. This style of result has been an ongoing theme of research for Vinh and collaborators in papers including [188, 194, 195]. Thang and Vinh [188] obtained a Szemerédi–Trotter type theorem and thus also a sum-product estimate via graph-theoretic methods (explicitly by studying a variant of the Erdős–Rényi graph). The main result we will concern ourselves with here is Vinh's Szemerédi-Trotter type theorem [194]. We will make use of this result later in this thesis specifically Chapter 5.

Theorem 3.7 (Theorem 3)**.** *[194] Let P be a collection of points and L be a collection of lines in $\mathbb{F}_q^2$. Then we have*

$$|\{(p,l) \in P \times L : p \in l\}| \leqslant \frac{|P||L|}{q} + q^{1/2}\sqrt{|P||L|}.$$

We note the short proof of this statement from [107] which combines the triangle inequality, Cauchy-Schwartz and the following identity on the second moment of the incidence function which is part of their Lemma 1, which states

$$\sum_l \left(|l \cap P| - \frac{|P|}{q}\right)^2 \leqslant q|P|. \tag{3.6}$$

Where the sum is over all lines in $\mathbb{F}_q^2$. We will then proceed to prove this bound which comes from calculating the variance.

Proof.

$$\begin{aligned}\left|\mathcal{I}(P,L)-\frac{|P||L|}{q}\right| &= \left|\sum_{l\in L}\left(|l\cap P|-\frac{|P|}{q}\right)\right| \\ &\leqslant \sum_{l\in L}\left||l\cap P|-\frac{|P|}{q}\right| \\ &\leqslant \sqrt{|L|\sum_{l\in L}\left(|l\cap P|-\frac{|P|}{q}\right)^2} \\ &\leqslant \sqrt{q|L||P|}\end{aligned}$$

□

We shall now prove Bound (3.6), the proof again coming from [107]. To do this we start by expanding out the bracket and then considering and consolidating the second two terms noting that every point is passed through by $q+1$ points and there are $q(q+1)$ lines (used to pass from the first to the second line).

$$\begin{aligned}\sum_l\left(|l\cap P|-\frac{|P|}{q}\right)^2 &= \sum_l\left(|l\cap P|^2-\frac{2|P|}{q}|l\cap P|+\frac{|P|^2}{q^2}\right) \\ &= \sum_l|l\cap P|^2-\frac{2|P|}{q}(q+1)|P|+q(q+1)\frac{|P|^2}{q^2} \\ &= \sum_l|l\cap P|^2-(q+1)\frac{|P|^2}{q} \\ &= \sum_l|l\cap P|^2-|P|^2-\frac{|P|^2}{q} \\ &\leqslant \sum_l|l\cap P|^2-|P|^2.\end{aligned}$$

To finish we just need a formula for $\sum_l|l\cap P|^2$ so we calculate it.

$$\begin{aligned}\sum_l|l\cap P|^2 &= \sum_l\left(\sum_{p\in P}\mathbb{1}_{p\in l}\right)^2=\sum_l\sum_{p,p'\in P}\mathbb{1}_{p\in l}\mathbb{1}_{p'\in l} \\ &= \sum_{p\in P}\sum_l\mathbb{1}_{p\in l}+\sum_{p\neq p'\in P}\sum_l\mathbb{1}_{p\in l}\mathbb{1}_{p'\in l}\end{aligned}$$

Here we need to use that $q+1$ lines are incident to a point and that a line is defined by two distinct points thus

$$\sum_{p\in P}\sum_l\mathbb{1}_{p\in l}=|P|(q+1)$$

and

$$\sum_{p\neq p'\in P}\sum_l\mathbb{1}_{p\in l}\mathbb{1}_{p'\in l}=\sum_{p\neq p'\in P}1=|P|(|P|-1).$$

Combining these gives

$$\sum_l|l\cap P|^2=|P|^2+q|P|,$$

and so

$$\sum_l \left(|l \cap P| - \frac{|P|}{q} \right)^2 \leqslant |P|^2 + q|P| - |P|^2 = q|P|$$

as desired.

We start by noting that the result of Theorem 3.7 is the optimal exponents under certain conditions before comparing this to our earlier established intuition due to the trivial results of Subsection 3.1.2 as well as the expected value of such incidences. To this end consider $|P| = |L| = n = q^{3/2}$, Theorem 3.7 then gives

$$|\{(p,l) \in P \times L : p \in l\}| \leqslant q^2 + q^2 = 2n^{4/3}.$$

Recalling Example 3.1, which we note has no restrictions on the size of n allowed, we see that this 4/3 exponent is the best we can hope for. It is also worth noting that the first term dominates for $|P||L| \gg q^3$, and it is in this range that Vinh's result is stronger than other incidence results considered in Section 3.1.3.2.

Turning to why we have named this section statistical terms we consider the expected value of the number of incidences. On each line, there are q points which may be in our point set P, as the chance of a point being in P is $|P|/q^2$. So the expected number of incidences is $|L||P|/q$ which we note is our first term, this means that another way of viewing Theorem 3.7 is that for sufficiently large sets (such as $|P||L| \gg q^3$) the number of incidences is about the expected number of incidences.

Finally, we will compare Theorem 3.7 to the trivial observations in Subsection 3.1.2 to complete our intuition. If we presume we are in the case where $|P||L| > q^3$ and thus it is the first term that dominates, it is a definite improvement on $|P||L|$ the most trivial bound, if we do not have $|P||L| > q^3$ then the second term is still an improvement on the trivial bound as long as $q < |P||L|$. Comparing instead to Lemma 3.1 where we have $|L||P|^{1/2}$ or $|L|^{1/2}|P|$ (we can throw away the second term as it is at most q^2) the improvement on the first term is when $|P| \leqslant q^2$ or $|L| \leqslant q^2$ respectively which is always the case. The second term is an improvement when $q \leqslant |L|$ or $q \leqslant |P$ respectively.

As Theorem 3.7 does better in the large set case, in applications later in this thesis we will split into cases, using Vinh's result for the case $|P||L| > q^3$ and Stevens and de Zeeuw's result otherwise allowing us to use the best bound for the circumstances.

We also note the character sum estimate by Iosevich et al. [69] which fits a similar theme. This is best possible for $|A|^2|H| > p^3$ and trivial when $|H| < p$.

$$\sigma(A,H) \leqslant \frac{|A|^2|H|}{p} + 2|A|\sqrt{p|H|}.$$

3.2 Relation to Sum Products

Incidence geometry is related to sum-product style results with Elekes' proof of Theorem 2.2 as shown in the previous chapter being the obvious example where the Szemedéri-Trotter

theorem (Theorem 3.1) is used to prove his result. This comes from considering the pointset $P = (A+A) \times (AA)$ and a corresponding set of lines of the form $y = a(x-b)$ with $a, b \in A$. An incidence bound on these sets them becomes a bound on the size of $|P| = |A+A||AA|$ given these sets have $|A|^3$ incidences as each of the $|A|^2$ lines passes through $|A|$ points of the form $(b+c, ac)$. Similarly, Section 2.6 shows another application of incidence results to sum-product style questions, an application that will be revisited in Chapter 5.

Another connection is between the number of incidences and sum-product style questions is that of lines and the affine group. The affine group (which will be looked at in more detail in Chapter 4) has been studied by Elekes in [40–42]. Rudnev and Shkredov [137] look at the connections between growth in the affine group and incidence results of the corresponding affine lines. They provide an incidence result between points and lines which depends explicitly on the energy of the affine transformations of the lines. We provide a nonlinear analogue of their Theorem 8 in Subsection 5.5.2. This work also provides bounds for sets such as $A(A+A)$ and $AA+A$. Roche-Newton and Warren [129] further built on these results. Murphy [106] proved upper and lower bounds for the number of lines in general position that are rich in a Cartesian product point set. This allows a geometric proof of Bourgain's asymmetric sum-product estimate. Petridis, Roche-Newton, Rudnev, and Warren [121] use incidence results to prove a nontrivial energy bound for a finite set of affine transformations over a general field. This leads to bounds for the growth of subsets of the affine group. In Chapter 4 we prove some generalisations (Theorems 4.9 and 4.10) of some of their result, Theorem 4.7 in this thesis.

Other matrix groups also have a geometric interpretation. Of particular interest is SL_2 and its relation to hyperbolae. Rudnev and Shkredov [137] study the relation of geometric incidence results and growth in SL_2. As an example from later in this thesis, Chapter 5 has several results relating to incidence results hyperbola which leads to sum-product style results as corollaries (see as example Corollary 5.2). The reason why SL_2 is of such interest, apart from being the group related to hyperbola and Möbius transformations which we can say something about (see Chapter 5), is SL_2 is what other matrix groups are "locally" like. Alternatively, the special linear lie algebra, the lie algebra for the special linear group is often used as the basis of study of other lie algebras. Further details can be found in Tao's blog posts [183][182].

3.3 Distance Problems

Also of note in discrete geometry and of use later in this thesis is distance problems. These concern questions about either the number of distances that can be found for a given number of points, possibly with restriction on their arrangement, or given a certain type of distances how can the points be arranged.

This section with primarily be to set up some background understanding of some of the results and methods used in Chapter 5 where we use concepts including pinned distances and convert some ideas from the standard Euclidean distances to Minkowski distances as well as prove Corollary 5.1, a result on unit distances. We note that this corollary is not as strong as we would have liked due to the structure required in our Theorem 5.3. We will comment on these

restrictions and provide a "bad" example.

Further reading for open questions involving distances can be found in Chapter five of Research Problems in Discrete Geometry [16].

3.3.1 Unit Distances

In 1946 Erdös conjectured in [45] that given n points in the plane, the maximum number of pairs at distance say 1 is $n^{1+\frac{c}{\log\log n}}$. He proved this as a lower bound and proved an upper bound of $n^{3/2}$. Spencer, Szemerédi and Trotter improved this upper bound to $n^{4/3}$ [169] which remains the current best known bound. Both of these results were over the reals. Schade [145] determined the extremal sets up to isomorphism for $n \leqslant 14$. In particular:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Unit distances	0	1	3	5	7	9	12	14	18	20	23	27	30	30

Table 3.1: The number of unit distances for small n due to Schade [145].

Given some further constraints (that is the set in question is well distributed and the unit distance is much smaller than the diameter of the set), Iosevich [78] improves the bound of 4/3.

The question of bounding unit distances has also been asked in three dimensions with Erdős [48] showing lower and upper bounds (up to constants) of $n^{4/3}$ and $n^{5/3}$ respectively. Clarkson, Edelsbrunner, Guibas, Sharir and Welzl [31] Kaplan, Matoušek, Safernová, and Sharir [81] and Zahl [201] made improvements lowering the upper bound to $n^{3/2}$. The current best result is due to Zahl [203] (by using techniques of Sharir and Zahl [150] applied to spheres) to get $n^{3/2-1/392+\varepsilon}$.

For $d \geqslant 4$ an example due to Lenz and reported in [48] shows the number of unit distances can be about n^2, without additional constraints.

This question is related to the number of intersections of points and circles (spheres in dimension three and so on for higher dimensions). This can be seen by considering the number of intersections of all circles of radius one with centres in P and the points P themselves. You could then easily adapt this question for other variants of distance (in particular we will consider this with Minkowski distances in Corollary 5.1, and detail Minkowski distances themselves in Section 5.2) by instead considering their sets of equal distance, that is the analogue of circles.

Another common statement of this question is to ask how dense a unit distance graph is (a graph where the points are connected by an edge if they are at distance one apart). The hypercube graph then is an example providing a trivial lower bound of $n \log n$ which is beaten by the lower bound of Erdős stated above.

This question can be similarly asked over other fields although I am not aware of a nontrivial, that is better than the $O(|P|^{3/2})$, bound (with $|P| < p$) on the number of realisations of a nonzero distance between pairs of a point set $P \subset \mathbb{F}_2$ in positive characteristic. This bound comes from Cauchy-Schwarz being applied to the set of all unit circles, C, with centres in P (so $|C| = |P|$).

Note we can use other conic sections, spheres and the like to allow us to deal with Minkowski distances or higher dimensions and so on. Explicitly

$$\sum_{c\in C}\sum_{p\in P}\mathbb{1}_{p\in c}\leqslant\sqrt{|P|\sum_{c\in C}\left(\sum_{p\in P}\mathbb{1}_{p\in c}\right)^2}\ll|P|^{3/2}.$$

Zahl [204] proved the following theorem which provides a lower bound of $n^{3/2}$ for P being a set in *three*, rather than *two* dimensions (when -1 is not a square in $\mathbb{F}$ and $|P|<p^2$).

Theorem 3.8. *Let $\mathbb{F}$ be a field in which -1 is not a square. Let $r\in\mathbb{F}\setminus\{0\}$ and let $P\subset\mathbb{F}^3$ be a set of n points in $\mathbb{F}^3$, with $n\leqslant(char(\mathbb{F}))^2$ (if $\mathbb{F}$ has characteristic zero then we impose no constraints on n). Then there are $O(n^{3/2})$ pairs $(x_1,y_1,z_1),(x_2,y_2,z_2)\in P$ satisfying*

$$(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2=r^2.$$

3.3.2 Distinct Distances

As a sort of opposite to the previous problem, another problem regarding distances due to Erdős is his distinct distances problem which asks about the number of distinct distances that can be made by arranging n points in the real plane? Equivalently can we find bounds in terms of $|P|$ of the following set

$$\Delta(P)=\{|x-y|:x,y\in P\}.$$

Trivially we have at most $\binom{|P|}{2}$ distances along with a zero distance. We can also expect this in the random case as if any pair of points happens to share a distance with another pair a small perturbation will disrupt this. As such the more interesting question is what is the minimum number of distinct distances possible given n points. Erdős [45] proved that for $|P|=n$ and c a constant

$$\sqrt{n-\frac{3}{4}}-\frac{1}{2}\leqslant\min\Delta(P)\leqslant\frac{cn}{\sqrt{\log n}}.$$

Erdős conjectured the upper bound (the example being the square grid) to be the true value and almost reached by Guth and Katz [67].

Progress was made in incremental steps as in Table 3.2.

As well as the plane this question can be considered in higher dimensions. Taking $D(n)$ to be the number of such distances and the dimension to be d Erdős proved the following

$$n^{1/d}\ll D(n)\ll n^{2/d},$$

conjecturing the upper bound to be the true value (the example being the $n^{1/d}\times\cdots\times n^{1/d}$ integer lattice). In the specific case of three dimensions, Clarkson, Edelsbrunner, Gubias, Sharir and Welzl [31] proved a lower bound of $n^{1/2}$ which was improved to $n^{77/141-\varepsilon}$ by Aronov, Pach, Sharir and Tardos [3] who also showed for the general case $n^{1/(d-90/77)-\varepsilon}$ Solymosi and Vu [166] improved the lower bound to $n^{2/d-2/d(d+1)}$. More details can be found in [59].

Lower Bound	Year	Author
$n^{2/3} \ll \min \Delta(P)$	52	Moser [104]
$n^{5/7} \ll \min \Delta(P)$	84	Chung [28]
$\frac{n^{4/5}}{\log n} \ll \min \Delta(P)$	92	Chung, Szemerédi & Trotter [29]
$n^{4/5} \ll \min \Delta(P)$	93	Székely [175]
$n^{6/7} \ll \min \Delta(P)$	01	Solymosi & Tóth [165]
$n^{\frac{4e}{5e-1}-\varepsilon} \ll \min \Delta(P)$	03	Tardos [186]
$n^{\frac{48-14e}{55-16e}-\varepsilon} \ll \min \Delta(P)$	04	Katz & Tardos [83]
$\frac{n}{\log n} \ll \min \Delta(P)$	15	Guth & Katz [67]

Table 3.2: Improvements for the Erdős Distinct Distances problem.

A continuous analogue of the Erdős distinct distances problem is the Falconer conjecture [52] which states a set of points that has a large Hausdorff dimension must determine a set of distances that is also large in measure. Partial results include [44][66][99]. Under non-Euclidean distance functions, this conjecture may be false [53][91].

We also note that when considering a specific different distance function, the Minkowski distance, which we will introduce and define in Section 5.2 and make use of ourselves in Chapter 5, Rudnev and Roche-Newton [127] proved the following result.

Theorem 3.9. *Let P be a point set in the real plane and so that the Minkowski distances between pairs of points, $\Delta_{1,1}(P)$, are not all zero then*

$$\Delta_{1,1}(P) \gg \frac{|P|}{\log |P|}.$$

3.3.2.1 Finite Fields

This question, like many others, can be considered also over the finite fields. However, issues arise from the lack of order and thus the more limited tools available than in the reals. Erdős' arguments from [45] works regardless of the field to show that a subset A of $\mathbb{F}^2$ determines at least $|A|^{1/2}$ non-zero distances. One of the early results to improve in this direction over $\mathbb{F}_p$ is due to Bourgain, Katz, and Tao in 2003 [14] who provide a result when $q \equiv 3 \mod 4$ is a prime, which says the set of distances determined by $|A|$ has a non-quantitative non-trivial bound ass follows

$$|\Delta(A)| \gg |A|^{1/2+\varepsilon}.$$

Iosevich and Rudnev [79] showed that the above cannot be true in general without $q \equiv 3 \mod 4$, this is related to the issue of isotropic distances which will arise in Chapter 5. They instead reformulated the question to ask how large a set in a general finite field $\mathbb{F}_q$ should be to have $\gg q$ distinct distances. This variation is often called the Erdős-Falconer distance problem. Iosevich and Rudnev showed that as long as $|A| \geqslant 4q^{(d+1)/2}$, then A had all possible distances. Hart, Iosevich, Koh, and Rudnev [68] constructed the natural obstructions, providing examples to show that in some cases you cannot improve the exponent.

Restricting ourselves to the case $d = 2$, Chapman, Erdogan, Hart, Iosevich and Koh [26] proved that the exponent 3/2 can be decreased to 4/3 and still get a positive proportion of all distances which is in line with Wolff's result [199] on the Falconer distance problem over the reals.

Murphy and Petridis [109] show the existence of an infinite family of subsets of $\mathbb{F}_q^2$ with $q^{4/3}$ elements that do not contain every distance. For other recent developments see [87].

This question is also connected to the finite field Kakeya Conjecture proposed by Wolff [198] and proved by Dvir [35].

3.3.3 Pinned Distances

Pinned distances were once again introduced by Erdős and are a related concept to the distinct distances problem. The idea behind pinned distances is how many distances are there from a specific special (pinned) point, that is for an $x \in P$ we define

$$\Delta_x(P) := \{|x - y| : y \in P\}.$$

This allows us to define pinned distances as

$$|\Delta_{pin}(P)| := \max_x |\Delta_x(P)|.$$

Erdős conjectured (in same paper [45]) that

$$|\Delta_{pin}(P)| \gg \frac{|P|}{\sqrt{\log |P|}}.$$

That is, asymptotically, pinned distances act the same as the standard distance problem. The number of pinned distances should be less than or equal to the number of distinct distances.

The best bound over the reals is due to Katz and Tardos [83] who prove

$$|\Delta_{pin}(P)| \gg |A|^{0.8641}.$$

In the continuous falconer setting the best bound is due to Guth, Iosevich, Ou and Wang [66].

Over the finite fields, the best bound is due to Murphy, Petridis, Pham, Rudnev and Stevens [110], who proves that in $\mathbb{F}_p$, if $|A| \geqslant p^{5/4}$ then the pinned distances of A contain a positive proportion of all the possible pinned distances. They also provide a bound for $4p < |A| < p^{5/4}$. They also prove the following theorem for general finite fields

Theorem 3.10. *Let $A \subset \mathbb{F}^2$ be a set of points. If $\mathbb{F}$ has positive characteristic $p > 0$, assume in addition that $|A| \leqslant p^{4/3}$. Then either A is contained in a single isotropic line, so all pair-wise distances are zero, or*

$$|\Delta_{pin}(A)| \gg |A|^{2/3}.$$

We will use ideas from this paper later in Chapter 5, in particular the concept of Bisector energy and Hinges.

Chapter 4

Growth in Three Dimensional Lie Groups

This chapter is based on my paper [113] which is joint work with Brendan Murphy. Section 4.5.1 continues from where the paper left off towards the general case, predominately looking at the 3×3 case which is joint work before commenting on Murphy, Pyber, Szabo and Eberhard's work [36] towards the generalisation and completion of mine and Murphy's two-dimensional results.

4.1 Background

Whilst Chapter 2 touched upon some of the results to do with growth in groups we will reconsider them here and go into more detail. We will also focus on a couple that serves in some way as the inspiration and starting point for the new theorems of myself and Murphy which will be introduced, detailed and proved later in this chapter.

Growth in groups started with (ignoring the case where you consider subsets that do not grow at all, that is the subsets which are subgroups and you are in the domain of group theory in general) Freiman [56] and his eponymous Theorem 2.8. Development continued with Ruzsa who supplied a new proof of Theorem 2.8 in [142]. Another early result is due Bourgain, Katz and Tao who in [14] provide a sum-product type result over $\mathbb{F}_p$ which can be seen as saying that a subset of $\mathbb{F}_p$ cannot have both small sumset and product set unless it is very small or almost everything, that is close to one of the two subfields of $\mathbb{F}_p$ (the trivial subfield and the whole field).

A result that is more obviously about general groups rather than just variations on the sum-product phenomena is Helfgot's [73] result concerning growth in $G = SL_2(\mathbb{F}_p)$. This lead to a string of results for similar groups which we will consider below. We state a rephrased form of Helfgott's theorem (noting that Helfgott proved this for fields of prime order and Dinai [33] generalised this to arbitrary finite fields) from Tao's book [184][Theorem 1.5.20].

Theorem 4.1. *Let A be a K-approximate group in $G := SL_2(\mathbb{F}_p)$ that generates G for some $K \geqslant 2$. Then one of the following holds:*

- *(Close to trivial) One has* $|A| \ll K^{O(1)}$
- *(Close to G) One has* $|A| \geqslant K^{-O(1)}|G|$.

I also note that the way of seeing this as a question of approximate groups and this notation is due to Tao's foundational work [177]. We describe approximate groups in Section 2.4.

Many of the later results can be considered in one of two frameworks - the first following that of Freiman Theorem 2.8 (that is sets which do not grow are like a generalised progression of some sort) and the second following instead Helfgott's result, Theorem 4.1. Providing a brief list of these results below.

- Helfgott [74] generalises his result to $SL_3(\mathbb{F}_p)$, saying that with the same setup and notation as in Theorem 4.1 bar G now being $SL_3(\mathbb{F}_p)$, if $|A| < G^{1-\varepsilon}$ then $|A^3| \gg |A|^{1+\delta}$ with $\delta > 0$ and the implied constant depend only on ε.
- Growth in $SL_n(\mathbb{F}_p)$ was proved independently by Pyber and Szabó [125], and Breuillard, Green, and Tao [21] showing that any approximate subgroup of $SL_n(\mathbb{F}_q)$ which generates the group must be small or nearly the entirety of $SL_n(\mathbb{F}_q)$. I note the argument in Breuillard, Green, and Tao also works for all Chevalley groups.
- Tao [179] proves a Freiman theorem-esque result for soluble groups which uses coset nilprogressions rather than coset progressions which were used by Green and Ruzsa [62] in their own generalisation to arbitrary abelian groups of the arithmetic progressions of Freiman's theorem. Explicitly Tao shows that any subset, with small doubling, of a solvable group is controlled by a set whose iterated products grow polynomially which are themselves contained in a virtually nilpotent group.
- Free groups were considered by both Razborov [126] and Safin [144]. Respectively, Razborov showed that given a finite subset of a free group (with at least two non-commuting elements) you have the following bound
$$|A^3| \geqslant \frac{|A|^2}{(\log|A|)^{O(1)}},$$
whilst Safin proves the similar (still a subset A with at least two non commuting elements) bound for
$$|A^n| \gg |A|^{\frac{n+1}{2}}.$$
- For torsion-free nilpotent groups (say Γ), Breuillard and Green [19] show that if A is a k-approximate subgroup of Γ then A is controlled by a given nilpotent progression.

  This result is generalised to arbitrary nilpotent groups in Tointons book [190][Chapter 6].
- A series of results for various linear groups were proven including Breuillard and Green's papers [18, 20]] and Gill and Helfgott's paper [60]) which will be looked at in greater detail below as they are similar in flavour to my new results presented in this chapter, in particular Theorems 4.2 and 4.4.

For the rest of this chapter, the following results form the foundations of and provide (by means of wishing to generalise them) inspiration for mine and Murphy's Theorem 4.6.

The first is the following result of Gill and Helfgott, in many ways the inspiration of the upcoming Theorem 4.6 of mine and Murphy's was to try to generalise this result from $\mathbb{F}_p$ to $\mathbb{F}_q$.

Theorem 4.2 (Gill and Helfgott, Theorem 1 [60])**.** *Let A be a subset of $GL_n(\mathbb{F}_p)$ such that $\langle A\rangle$ is solvable. Then, for every $K \geqslant 1$, either*

- $|A^3| \geqslant K|A|$*, or else*
- *there is a unipotent subgroup U_R, a solvable group S, and an integer k depending only on n such that*
  - *U_R and S are both normal in $\langle A\rangle$, and S/U_R is nilpotent,*
  - *A^k contains U_R, and*
  - *$|A^k \cap S| \geqslant K^{-O_n(1)}|A|$.*

We use the notation $T_n(\mathbb{F})$ to denote the group of invertible $n \times n$ upper triangular matrices over a field $\mathbb{F}$. Theorem 4.2 is proved by reducing to the case where $A \subseteq T_n(\mathbb{F}_p)$. Theorem 4.2 was strengthened in the joint work of Gill, Helfgott, Pyber, and Szabó to the following, removing the requirement that $\langle A\rangle$ is solvable.

Theorem 4.3 (Gill, Helfgott, Pyber, and Szabó Theorem 2 [60])**.** *Let A be a subset of $GL_n(\mathbb{F}_p)$. Then, for every $C \geqslant 1$, either*

- $|A_{(3)}| \geqslant C|A|$*, or else*
- *there are two subgroups $H_1 \leqslant H_2 \leqslant GL_n(\mathbb{F}_p)$ and an integer $k \ll_n 1$ such that*
  - *H_1 and H_2 are both normal in $\langle A\rangle$, and H_1/H_2 is nilpotent,*
  - *$A_{(k)}$ contains H_1, and*
  - *$|A_{(k)} \cap H_2| \geqslant C^{-O_n(1)}|A|$.*

When combined with work of Tointon [189] this result can be stated in the manner of a K-approximate group of $GL_n(\mathbb{F}_p)$ being controlled by a coset nilprogression of given rank and step.

Commenting briefly on the proof, this result is achieved by reducing to the case of a connected solvable linear algebraic group, splitting this into a unipotent and a maximal torus before applying Corollary 3.2 from [74] to the action of conjugation of the torus on the unipotent. We also note that one of the major obstructions to the proof working over arbitrary finite fields is that it relies on the fact that a subgroup chain of unipotent subgroups

$$U_1 > U_2 > \dots$$

is of length at most n^2 when n is the dimension of the matrices we are considering. Gill and Helfgott conjectured that this should not be a major problem and as such their result should hold for any finite field $\mathbb{F}_q$, where $q = p^r$ is any prime power. You can get such a result as a corollary of Theorem 4.2 by embedding $GL_n(\mathbb{F}_{p^r})$ into $GL_{rn}(\mathbb{F}_p)$, but in this case, the constants would tend rapidly to infinity with r which we would like to avoid.

Breuillard and Green [20] proved a similar result over the complex numbers.

Theorem 4.4 (Breuillard and Green, Theorem 1.4' [20])**.** *Let $K > 1$. Suppose that $A \subseteq T_n(\mathbb{C})$ is a set with $|A^3| \leqslant K|A|$. Then there is some set $A' \subseteq A$ with $|A'| > K^{-C}|A|$ that is contained in a left coset of a nilpotent subgroup of $T_n(\mathbb{C})$ of step at most $n-1$.*

To grasp what this result tells us we need to understand what being in a coset of a nilpotent subgroup of step at most $n-1$ means. Starting with the definition

Definition 4.1. *G is a nilpotent subgroup if it has a lower central series terminating in the trivial subgroup after a finite number of steps. That is there exist a series*

$$G = G_0 \rhd G_1 \rhd \cdots \rhd G_n = \{1\}$$

with $G_i := [G_{i-1}, G] = G_{i-1}^{-1} G^{-1} G_{i-1} G$. It is of step s if it terminates in s steps. We note we can replace the lower central series with the upper central series or Central series.

A nilpotent subgroup of step at most 1 is an abelian subgroup which will provide the comparison to Theorem 4.6 which only deals with 2×2 matrices. It helps to keep in mind that a nilpotent subgroup is almost abelian with the step saying how close.

The proof follows an induction argument, making use of the maps

$$\pi, \pi' : T_n(\mathbb{C}) \to T_{n-1}(\mathbb{C})$$

where π maps to the $n-1 \times n-1$ submatrix consisting of the top left corner, that is removing the bottom row and rightmost column and π' maps instead to the bottom right $n-1 \times n-1$ submatrix. Once this is done they show that an iterated commutator lives in the intersection of the kernels of π and π' (that is the matrices with ones on the main diagonal, zeros everywhere else except the upper right entry) and make use of a lemma that states that a group generated by $s+1$ nested commutators is s step nilpotent to find the required coset of a nilpotent subgroup. This provides a method for generalising Theorem 4.6 to higher dimensions but complications arise from the presence of nontrivial finite subfields of $\mathbb{F}_q$ which neither $\mathbb{F}_p$ or $\mathbb{C}$ have to deal with.

In Section 4.3 I present mine and Murphy's continuation in this direction, generalising Theorem 4.2 from $\mathbb{F}_p$ to $\mathbb{F}_q$ in the very specific case that $n = 2$. I will discuss work towards the general dimensional result in Sections 4.5.1 and 4.5.2.

As a final note, I present the following theorem of Murphy.

Theorem 4.5 (Theorem 35 (Product theorem for Aff$(1,\mathbb{F}_p)$))**.** *[106] If A is a subset of Aff$(1,\mathbb{F}_p)$ such that $|A^3| \leqslant K|A|$, then either $\geqslant \frac{|A|}{3}$ elements of A are contained in a torus, or, defining U as the subgroup of unipotent matrices,*

$$K^{10}|A| \gg |A/U|^{1/2}|A|,$$

or

$$K^{10}|A| \gg |A/U|p.$$

This can be seen as the affine version of the upcoming Theorem 4.6 and note the same paper has a similar result for replacing the finite field with the complex numbers. Theorem 4.5 states that if a subset of the group of affine matrices does not grow (it has small tripling) then we are in a case where the subset is either in a torus or covered by a few cosets of the unipotent matrices. A stronger version appears in [137] as their Theorem 5 which also splits up the cases similarly with the subset with small tripling being in the stabiliser of an element x which is a torus or covered by a few cosets of the unipotent matrices. It is however of interest as a comparison and sort of ancestor of the Theorem 4.6, in that Theorem 4.6 also splits subsets of a matrix group (now the upper triangular 2×2 matrices rather than affine matrices) into cases of being nearly contained in a coset of an abelian subgroup such as a torus.

4.2 Preliminary Results

This section will note a couple of results that we will use in the proof of Theorem 4.6 in the next section, we will also use some results from Chapter 2.

We provide the definition of a right inverse of a function to allow us to state the following lemma which is part (iv) of Lemma 2.12 (Tao's splitting lemma) in Tointon's paper [189], which itself follows Tao [177]. The definition is as follows, for a function $f : A \to B$, we say that $\phi : B \to A$ is a *right inverse* if $f \circ \phi(x) = x$ for all $x \in B$. So moving on to the lemma that we will use later.

Lemma 4.1. *Let N be a normal subgroup of G, let $\pi : G \to G/N$ be the quotient map, and let A be a finite subset of G. Let k be a positive integer and let $\phi : \pi(A^k) \to A^k$ be a right inverse. Then for all a in A^k, we have*

$$a \in \phi(\pi(a))(A^{-k}A^k \cap N).$$

Hence

$$A^k \subseteq \phi(\pi(A^k))(A^{-k}A^k \cap N).$$

Whilst we do not state the proof here, its essence is seeing that the identity is contained in $A^{-k}A^k \cap N$ and $a \in \phi(\pi(a))$.

Another result used in the proof of Theorem 4.6 is the following sum-product result [108, Theorem D]. This result requires us to define $\mathrm{Span}_F(X) := \sum_i f_i x_i$ where $f_i \in F$ and $x_i \in X$.

Proposition 4.1. *Let X be a finite subset of an $\mathbb{F}_q$-vector space, let $D \subseteq \mathbb{F}_q$ be a set of scalars, and let $F = \langle D \rangle$ be the subfield generated by D. If $|X + DX| \leqslant K|X|$ for some $K \geqslant 1$, then either $K \geqslant |D|^{1/10}$ or*

$$|X| \geqslant \frac{1}{2K^4}|\mathrm{Span}_F(X)|,$$

and

$$\mathrm{Span}_F(X) \subseteq 4DX - 4DX.$$

4.3 Product Theorem in T_2

We will start this section by stating the first of mine and Murphy's theorems we will prove in this chapter. We will continue by discussing the theorem before moving on to prove the result.

Theorem 4.6. *If $A \subseteq T_2(\mathbb{F}_q)$ satisfies $e \in A, A = A^{-1}$, and $|A^3| \leqslant K|A|$, then either*

1. *There is an $\mathbb{F}^*$-potent group H such that $|A^2 \cap H| \gg K^{O(1)}|A|$, or*
2. *There is a subgroup $U \leqslant U_2(\mathbb{F}_q)$ such that $U \subseteq A^{O(1)}$ and $\langle A \rangle / U$ is abelian.*

This result can be seen as a classification of approximate subgroups of the 2×2 upper triangular matrices, that is given a subset A of $T_n(\mathbb{F}_q)$ which does not grow, then we are in one of two cases, either A is nearly contained in a coset of an abelian subgroup (that is either the first case in the theorem, or a subcase of the second where we are contained in the coset of a torus), or the upper right-hand entries of the elements of A essentially form a vector space over the subfield generated by ratios of the diagonal of elements of A. In particular, in this case, the unipotent subgroup U has the form

$$U = \left\{ \begin{pmatrix} 1 & w \\ 0 & 1 \end{pmatrix} : w \in W \right\},$$

where W is a vector space over a subfield $F \leqslant \mathbb{F}_q$ and is thus a subgroup of the full unipotent subgroup. It is this case that does not appear in the previous literature as in the case of $\mathbb{F}_p$ and $\mathbb{C}$, the only trivial finite subfields we have are the trivial subfield or in the finite field case the whole group.

As I stated earlier Theorem 4.6 is related to Gill and Helfgott's Theorem 4.2. As such for ease of comparison it is worth considering what our conclusions look like in the style of Theorem 4.2. Taking our cases separately, in the first $U_R = \{\mathrm{id}\}$ is the trivial subgroup and $S = \langle A^2 \cap H \rangle$, is the subgroup generated by $A^2 \cap H$. In the second case, U_R becomes our U and $S = \langle A \rangle$. Considering Theorem 4.6 and its proof over the complex numbers leads to U being trivial as we have no other finite subfields to choose from and Theorem 4.6 shares the same conclusion as the 2×2 case of Breuillard and Green's Theorem 4.4.

Having discussed the result and its connection to prior results we turn to the proof which is based on the proof from mine and Murphy's paper [113].

Proof of Theorem 4.6. First, note that as the diagonal elements of products of matrices in $T_2(\mathbb{F}_q)$ are just the product of the elements, that is for $A,B \in T_2(\mathbb{F}_q)$ the top left entry of AB is equal to the top right entry of A times that of B. In particular, commutators, being of the form $A^{-1}B^{-1}AB$, are unitriangular matrices and so we have the following

$$[A,A] = \{[a,a']\colon a,a' \in A\} \subseteq A^4 \cap U_2(\mathbb{F}).$$

Let $u\colon \mathbb{F}_q \to U_2(\mathbb{F}_q)$ be the isomorphism defined by

$$u\colon x \mapsto \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$$

and let $X := u^{-1}(A^4 \cap U_2(\mathbb{F}_q))$, that is X is the set of top right entries of unitriangular matrices which are also the product of four matrices in A and thus contains the top right entries of all commutators of matrices in A.

Let $\chi\colon T_2(\mathbb{F}_q) \to \mathbb{F}_q^*$ denote the homomorphism defined by

$$\chi\colon \begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix} \mapsto \frac{a_{11}}{a_{22}},$$

and let $D = \chi(A)$, the set of diagonal ratios. Conjugation of $U_2(\mathbb{F}_q)$ by an element g of $T_2(\mathbb{F}_q)$ corresponds to multiplication by $\chi(g)$, that is we have $tu(v)t^{-1} = u(\chi(t)v)$ for any $t \in T_2(\mathbb{F}_q)$ and $v \in F_q$. Using this fact we have $u(X + DX) \subseteq A^{10} \cap U_2(\mathbb{F}_q)$.

By Lemma 2.6,

$$|X + DX| = |u(X + DX)| \leqslant |A^{10} \cap U_2(\mathbb{F}_q)| \ll K^{30}|A^2 \cap U_2(\mathbb{F}_q)| \leqslant K^{30}|X|,$$

hence by Proposition 4.1, either $|D| \ll K^C$ or there is a vector space W over the field F generated by D such that $X \subseteq W \subseteq 4DX - 4DX$.

In the first case, there is a $\mathbb{F}^*$-potent subset S of A^2 such that $|S| \gg K^{-C}|A|$.

In the second case, set $U = u(W)$. Then U is contained in $A^C \cap U_2(\mathbb{F}_q)$ and $A^2 \cap U_2(\mathbb{F}_q) \subseteq U$, so for any set $\Lambda \subseteq A$ of left coset representatives of A modulo $U_2(\mathbb{F}_q)$, we have

$$A \subseteq \Lambda \cdot U$$

by Lemma 4.1. Since the ratios of the diagonal terms of elements of Λ are contained in D and W is a vector space over the field F generated by D, the subgroup U is normalised by Λ, hence $\langle A \rangle = \langle \Lambda \rangle U$.

It remains to show that $\langle A \rangle / U$ is abelian. Let $\phi\colon \langle A \rangle \to \langle A \rangle / U$ denote the quotient map. Since $\Lambda \subseteq A$, we have $[\Lambda,\Lambda] \subseteq U$, so

$$[\phi(\Lambda), \phi(\Lambda)] = \phi([\Lambda,\Lambda]) = \{e\},$$

so $\phi(\Lambda)$ generates an abelian subgroup of $\langle A \rangle / U$. But

$$\big\langle \phi(\Lambda) \big\rangle \cong \langle \Lambda \rangle U / U = \langle A \rangle / U, \tag{4.1}$$

so $\langle A \rangle / U$ is abelian, as claimed. □

4.4 Energy Results in Matrices

Having proven Theorem 4.6, we now consider a couple of related energy bounds for matrices, Theorem 4.9 will consider the same case of 2×2 upper triangular matrices as Theorem 4.6. Theorem 4.10 deals instead with the Heisenberg group. After some shared setup we will deal with each in their own subsections. We note that we make use of further energy results in Chapter 5, specifically, we prove Lemmas 5.1 and 5.2 which are energy bounds (one a higher energy bound) again for matrices, but of the form

$$\begin{pmatrix} -a & ab+1 \\ -1 & b \end{pmatrix}$$

rather than the upper triangular and Heisenberg matrices here.

The ideas behind the proof of these theorems come from the following result for the affine group,

$$\mathrm{Aff}(\mathbb{F}) := \left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} : a,b \in \mathbb{F}, a \neq 0 \right\},$$

of Petridis, Roche-Newton, Rudnev and Warren [121] which states:

Theorem 4.7 (Corollary 6 [121])**.** *Let A be a subset of the affine matrices over $\mathbb{F}_p$ have no more than M elements in a coset of a torus and no more than m elements in a coset of the unipotent subgroup. Suppose $m|A| \leqslant p^2$. Then*

$$\min\{|AA|, |A^{-1}A|\} \gg m^{-1/2}|A|^{3/2} + M^{-1}|A|^2.$$

In particular, if $|AA| = K|A|$ and $m|A| \leq p^2$, then A has $\gg \frac{|A|}{K^2}$ elements in a coset of the unipotent subgroup or $\gg \frac{|A|}{K}$ elements in a coset of a torus.

This result deals with two-dimensional groups rather than the three-dimensional groups of Theorems 4.9 and 4.10. It is also of note that the affine group is a subgroup of the upper triangular matrices, in fact, we could see an element of $T_2(\mathbb{F})$ as some constant times an element of the affine group in some manner, we will look at this in more detail in Section 4.4.1.

The most general idea of the proof is to split the energy up in similarly rich pieces before identifying the set of three equations from considering the three non zero entries in the matrices with a point plane incidence and making use of Rudnev's point-plane incidence bound [133]. This is a variation (that is we restrict to the case where we have almost all points and thus do not need the last term) in the statement of Theorem 3.3.

Theorem 4.8. *Let $\mathbb{F}_q$ be a field, and let P and Π be finite sets of points and planes respectively in $\mathbb{P}^3$. Suppose that $|P| \leqslant |\Pi|$, and that $|P| \ll p^2$. Let k be the maximum number of collinear points in P. Then the number of incidences satisfies*

$$\mathcal{I}(P,\Pi) \ll |\Pi||P|^{1/2} + k|\Pi|.$$

Although this is a result over $\mathbb{F}_q$ note we are still restricted in terms of p and need small sets (about at most p). As we will explain later, this is what leads to some differences between Theorem 4.6 and Theorem 4.9.

4.4.1 Energy in 2×2 Triangular Matrices

As stated above we will deal with the two theorems separately; we start with the energy result for 2×2 upper triangular matrices. Before we get to the statement and then proof of the result we will consider some background on the group which will be of use.

We start with some comments on the structure of $T_2(\mathbb{F}_q)$, first, it is the direct product of its centre $\Lambda \leqslant D_2(\mathbb{F}_q)$, a subgroup of diagonal matrices with equal elements on the main diagonal equivalently seen as a constant multiplied by the identity matrix, and a subgroup Γ that is isomorphic to the affine subgroup. This formalises what was said earlier about $T_2(\mathbb{F}_q)$ being a constant times the affine group in a way. we define the projection $\rho\colon T_2(\mathbb{F}_q) \to \Gamma$ as the homomorphism

$$\rho\colon \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \mapsto \begin{pmatrix} a/c & b/c \\ 0 & 1 \end{pmatrix}.$$

Continuing the theme of identifying subgroups, note that Γ contains the normal unipotent subgroup $U_2(\mathbb{F}_q)$ (when $a/c = 1$). We also note the following definition, for $g \in A$, we call a multiple of g by an element of Λ a *dilate* of g.

For the trivial bounds for this energy term, recall Section 2.5. However, we also cannot expect to have a non-trivial upper bound on $E(A)$ unconditionally, since A can lie in a coset of an abelian subgroup. To this end, it is worthwhile considering what the maximal abelian groups of $T_2(\mathbb{F}_q)$ are. They arise as Λ (itself an abelian group) times a maximal abelian subgroup of Γ. Explicitly writing out the form of elements of such groups we have two cases

$$\begin{pmatrix} a & b \\ 0 & a \end{pmatrix} \in \Lambda U_2 \qquad \text{and} \qquad \begin{pmatrix} a & (c-a)x \\ 0 & c \end{pmatrix} \in \Lambda T,$$

where T is a maximal torus in Γ, that is isomorphic to the diagonal subgroup $D_2(\mathbb{F}_q)$.

Recall we have seen these groups before, in particular, the first is the $\mathbb{F}^*$-potent subgroup called out in case one of Theorem 4.6 whereas the second, the tori, is noted as a subcase of the second case of Theorem 4.6. We also note as both are abelian groups they are 1-step nilpotent and thus are the subgroups that would satisfy Breuillard and Green's Theorem 4.4 in the 2×2 case.

Given this background, we now state the result before moving straight into the proof[1].

Theorem 4.9. *Let $A \subseteq T_2(\mathbb{F}_q)$ and M_1 be the maximum number of elements of A in a coset of ΛT, M_2 be the maximum number of elements of A in a coset of ΛU_2, M_3 in a coset of U_2, suppose $|A|M_3 \leqslant p^2$.*

Then we have the following energy estimate

$$E(A) \lesssim |A|^{5/2} M_2^{1/2} + |A|^2 M_1,$$

and hence

$$|A^{-1}A|, |AA| \gtrsim \frac{|A|^2}{M_1 + \sqrt{|A|M_2}}.$$

[1] Which again from mine and Murphy's paper [113] with minor additions.

Proof. We begin by splitting A into dyadic pieces, then we estimate the energy of each dyadic piece by partitioning them further and controlling these with the point-plane theorem.

To this end, define the subset $A_m \subseteq A$ by

$$A_m := \{g \in A : m \leq |A \cap g\Lambda| < 2m\}.$$

Note that A is the union

$$A = \bigcup_{j=0}^{M} A_{2^j},$$

where $M \ll \log|A|$.

Thus by two applications of Cauchy-Schwarz, we have

$$E(A) = E\left(\bigcup_{j=0}^{M} A_{2^j}\right) \lesssim \sup_j(E(A_{2^j})). \tag{4.2}$$

This is where we lose our log factors. This can be seen by Lemma 2.8, where we have the first line of the following, we then proceeded by bounding each $E(A_{2^j})$ by the supremum of these energies.

$$\begin{aligned} E\left(\bigcup_{j=0}^{M} A_{2^j}\right) &\ll \left(\sum_{j=0}^{M} \left(E\left(A_{2^j}\right)\right)^4\right)^{1/4} \\ &\ll \left(M \sup_j \left(E\left(A_{2^j}\right)\right)^4\right)^{1/4} \\ &\ll \log^{1/4}|A| \sup_j \left(E\left(A_{2^j}\right)\right) \\ &\lesssim \sup_j \left(E\left(A_{2^j}\right)\right) \end{aligned}$$

We will show that for any $m \geq 1$, we have

$$E(A_m) \ll |A_m|^{5/2} M_2^{1/2} + M_1 |A_m|^2, \tag{4.3}$$

provided that $M_3|A_m| \ll p^2$. Since $M_3|A| \ll p^2$ by assumption, we have $E(A_{2^j}) \ll |A|^{5/2} M_2^{1/2} + M_1|A|^2$ for all j, so we may bound the right-hand side of Equation (4.2) by taking a supremum:

$$E(A) \lesssim |A|^{5/2} M_2^{1/2} + M_1 |A|^2.$$

Thus the proof is complete, pending the proof of Equation (4.3).

We write

$$Q(A_m) := \left\{(g,h,u,v) \in A_m^4 : g^{-1}h = u^{-1}v\right\}, \tag{4.4}$$

so that $E(A_m) = |Q(A_m)|$. We will write elements of $T_2(\mathbb{F}_q)$ as

$$g = \begin{pmatrix} g_1 & g_2 \\ 0 & g_3 \end{pmatrix},$$

adopting the convention that if $h \in T_2(\mathbb{F}_q)$, then h_1, h_2, h_3 bear the same relationship to h as g_1, g_2, g_3 do to g.

To proceed, we partition the set $Q(A_m)$ into pieces, which we will then be able to control via the point-plane incidence bound (Theorem 4.8).

We can split the energy of A up as follows

$$E(A_m) = \sum_{C_1, C_3} \left|\{(g,h,u,v) \in Q(A_m) \colon g_1 v_1 = C_1 = h_1 u_1, g_3 v_3 = C_3 = h_3 u_3\}\right|.$$

To see that this is correct, suppose that $g^{-1}h = u^{-1}v$. Then, by comparing the three entries in the matrices, we have a set of three equations:

$$g_1 v_1 = h_1 u_1, \qquad g_3 v_3 = h_3 u_3, \qquad u_1 h_2 - u_1 h_3 \frac{g_2}{g_3} = g_1 v_2 - g_1 v_3 \frac{u_2}{u_3}. \tag{4.5}$$

We write Q_{C_1,C_3} for the set of solutions corresponding to the fixed pair of values $C = (C_1, C_3)$ in the decomposition above:

$$Q_C = Q_{C_1,C_3} := \left|\{(g,h,u,v) \in Q(A_m) : g_1 v_1 = C_1 = h_1 u_1,\ g_3 v_3 = C_3 = h_3 u_3\}\right|.$$

We define additional sets, also indexed by pairs of values $C = (C_1, C_3)$ from the above decomposition:

$$\mathcal{P}_C = \mathcal{P}_{C_1,C_3} := \{(g,v) \in A_m \times A_m : g_1 v_1 = C_1,\ g_3 v_3 = C_3\}.$$

Next, we seek to bound the quantity Q_C using Theorem 4.8. The last of the conditions in Equation (4.5) is the condition that represents point-plane incidences. The planes are given by projective co-vectors

$$\left(-u_1 h_2 : u_1 h_3 : 1 : -\frac{u_2}{u_3}\right),$$

and points by projective vectors

$$\left(1 : \frac{g_2}{g_3} : g_1 v_2 : g_1 v_3\right) = \left(1 : \frac{g_2}{g_3} : g_1 v_2 : C_3 \frac{g_1}{g_3}\right).$$

The points and planes as above are multisets, since the ratios $\frac{g_2}{g_3}$ and $\frac{g_1}{g_3}$ are defined module Λ. By the definition of A_m, each point and plane occurs with multiplicity (or "weight") $\sim m$.

Thus, to account for the weights, we consider the worst possible case. This is when the number of points/planes is equal to $|\mathcal{P}_C|/m$ and each incidence is counted m^2 times

$$Q_C \ll |\mathcal{P}_C|^{3/2}\sqrt{m} + (km)|\mathcal{P}_C|, \tag{4.6}$$

where k is the maximum number of collinear points and the latter estimate is valid provided that we have the following which is due to the p-constraint on the application of the point-plane.

$$|\mathcal{P}_C| \leqslant mp^2 \tag{4.7}$$

We remark that if points in $\mathcal{P}_C$ are collinear, then their projection on the coordinates $(1 : \frac{g_2}{g_3} : \frac{g_1}{g_3})$ are collinear, so we get a line in Γ, recalling that Γ that is a subgroup of $T_2(\mathbb{F}_q)$ isomorphic to the

affine subgroup. A line in Γ is a coset of a torus T or the unipotent group $U_2(\mathbb{F}_q)$. If the line (coset in Γ) has k elements and each element has $\sim m$ dilates in $T_2(\mathbb{F}_q)$, the quantity km is bounded by the maximum number of elements of A in a coset of ΛT or ΛU_2, which we bound by $M_1 + M_2$.

To continue, we will sum Estimate (4.6) over all values of (C_1, C_3) using the fact that

$$\sum_{(C_1,C_3)} |\mathcal{P}_{(C_1,C_3)}| = |A_m|^2,$$

and a supremum estimate for $|\mathcal{P}_C|$. Namely, for a fixed $C = (C_1, C_3)$, if we consider the set $\mathcal{P}_C$, if we know $g \in A_m$ then we know v_1 and v_3. The maximum number of v_2, knowing v_1 and v_3 is the maximum number n of elements of A_m in a coset of U_2. Since each element has $\sim m$ dilates in A_m, then $n \ll M_2/m$.

Hence by summing Estimate (4.6) over all values of (C_1, C_3) and using the above facts we have that

$$E(A_m) \ll |A|^{5/2} M_2^{1/2} + |A|^2 M_1, \tag{4.8}$$

and this is valid, by Estimate (4.7) as long as

$$|\mathcal{P}_C| \leqslant |A_m| n \leqslant m p^2,$$

this constraint appears with $m = 1$ and M_3 replacing n in the statement of the theorem and thus we are done. □

Having now proved both this Theorem 4.9 and Theorem 4.6, both results about growth in $T_2(\mathbb{F}_q)$ it would be remiss not to compare them. Both results show that a possible obstruction to growth is being close to a coset of

$$\begin{pmatrix} a & b \\ 0 & a \end{pmatrix} \in \Lambda U_2 \text{ or } \begin{pmatrix} a & (c-a)x \\ 0 & c \end{pmatrix} \in \Lambda T.$$

Theorem 4.6 also had another obstruction where the upper right entries form a subfield, however, this does not arise in Theorem 4.9. This is because of the restriction that $|A|M_3 \leqslant p^2$ (which itself stems from the use of Rudnev's point plane incidence Theorem 4.8) meaning we do not have enough elements to fill a unipotent subgroup and so case two of Theorem 4.6 only occurs in Theorem 4.9 when U is not trivial which is the torus, ΛT, case above. This shows a weakness of this second approach although, on the other hand, it is a stronger result quantitatively.

4.4.2 Energy in the Heisenberg Group

This subsection will follow the general layout as the previous one, exchanging $T_2(\mathbb{F}_q)$ for the Heisenberg group and Theorem 4.9 for 4.10 which generalises results of Shkredov [156, Theorem 2].

Starting with the definition and basic information behind the Heisenberg group we have

$$H = H(\mathbb{F}_q) := \left\{ \begin{pmatrix} 1 & g_1 & g_3 \\ 0 & 1 & g_2 \\ 0 & 0 & 1 \end{pmatrix} : g_1, g_2, g_3 \in \mathbb{F}_q \right\},$$

where we write

$$g = (g_1, g_2, g_3) \qquad \text{or} \qquad g = \begin{pmatrix} 1 & g_1 & g_3 \\ 0 & 1 & g_2 \\ 0 & 0 & 1 \end{pmatrix},$$

for a specific element of H. Having stated that this is a group I should be able to give you the group operation, inverses and the identiy. To this end we have

$$gh = (g_1 + h_1, g_2 + h_2, g_3 + h_3 + g_1 h_2)$$

$$g^{-1} = (-g_1, -g_2, -g_3 + g_1 g_2),$$

$$g^{-1} \cdot h = (h_1 - g_1, h_2 - g_2, (h_3 - g_3) + g_1(g_2 - h_2)),$$

and $(0,0,0)$ is the identity element.

We will use LZ for the two-dimensional abelian subgroups of the Heisenberg group. Explicitly we take L as the subgroup defined by $g_3 = 0$ and $\alpha g_1 + \beta g_2 = 0$ depending on some α, β. Z is the centre of H, that is the subgroup formed of elements of the form $(0,0,g_3)$. These play a similar role to the centre, Λ of $T_2(\mathbb{F}_q)$ and the abelian subgroups: the unipotent subgroup and the tori.

We now state Theorem 4.10 before proving following similar ideas to the proof of Theorem 4.9 with the required adjustments.

Theorem 4.10. *Let $A \subset H(\mathbb{F}_q)$, let m be the maximum number of elements of A in a coset of Z, and let M be the maximum number of elements in a coset of LZ. If $|A|m \leqslant p^2$ and $m \leqslant \sqrt{|A|}$, then we have the energy estimate*

$$E(A) \lesssim |A|^{5/2} m + |A|^2 M,$$

and hence,

$$|A^{-1}A|, |AA| \geqslant \frac{|A|^2}{M + m\sqrt{|A|}}.$$

We note that unlike Theorem 4.9 we do not lose any log factors in the bound for the Heisenberg group.

Proof. We will reuse, from the proof of Theorem 4.9, the notation

$$Q(A) := \{(g,h,u,v) \in A^4 : g^{-1} \cdot h = u^{-1} \cdot v\}. \tag{4.9}$$

As in the proof of Theorem 4.9 we start by comparing the three equations. We will then extract a weighted point-plane equation and turn to Theorem 4.8 to provide bounds.

Equating $g^{-1}\cdot h = u^{-1}\cdot v$, and fixing

$$h_1+u_1 = C_1 = g_1+v_1, \qquad h_2+u_2 = C_2 = g_2+v_2,$$

then the third equation (in the manner from Theorem 4.9's set of Equations (4.5)) is

$$[-(g_3+v_3)+g_1g_2-C_2g_1]+g_1u_2-g_2u_1+[h_3+u_3-u_1u_2+C_2u_1] = 0. \tag{4.10}$$

As before this is a weighted point-plane equation. Following the same general method, let Q_{C_1,C_2} denote the set of solutions corresponding to the fixed pair of values $C=(C_1,C_2)$ in the decomposition above, that is,

$$Q_C = Q_{C_1,C_2} := \left|\left\{(g,h,u,v)\in Q(A): \begin{array}{l} g_1+v_1 = C_1 = h_1+u_1 \\ g_2+v_2 = C_2 = h_2+u_2 \end{array}\right\}\right|.$$

Still following the previous subsections notation, only now $C=(C_1,C_2)$, we define the set

$$\mathcal{P}_C = P_{C_1,C_2} := \{(g,v)\in A\times A : g_1+v_1 = C_1,\ g_2+v_2 = C_2\}.$$

As in the previous proof, we will bound the quantity Q_C using Theorem 4.8. Equation (4.10) represents point-plane incidences, where points are given by projective vectors and planes by projective co-vector

$$\big(-(g_3+v_3)+g_1g_2-C_2g_1 : g_1 : g_2 : 1\big), \qquad \big(1 : u_2 : -u_1 : h_3+u_3-u_1u_2+C_2u_1\big). \tag{4.11}$$

The points and planes come with multiplicity as they did in the last case. This multiplicity, in the case of points, is the number of realisations of the sum g_3+v_3. Observe that given C_1,C_2 and knowing g_1,g_2 we then known v_1,v_2. Thus the maximum number of realisations of the sum g_3+v_3 (since g_1,g_2, as well v_1,v_2 are fixed for a given point) is bounded by the maximum number of elements of A in a coset of U, which we denote as m.

We continue by applying the point-plane bound, as in the previous section, only now $C=(C_1,C_2)$. Again, to account for weights, we consider the worst possible case; when the number of points/planes is equal to $\frac{|\mathcal{P}_C|}{m}$, each incidence is counted m^2 times. Summing over C we obtain, once again, the Estimate (4.6),

$$Q_C \ll |\mathcal{P}_C|^{3/2}\sqrt{m} + (km)|\mathcal{P}_C|.$$

For the p-constraint on the application of the point-plane theorem we take the most ample $|\mathcal{P}_C|\leqslant p^2$, i.e. the case $m=1$.

Observe that geometrically, for a given value of (g_3+v_3), Equations (4.11) are a quadric over the (g_1,g_2) plane, and hence k is bounded by the number of collinear points in the (g_1,g_2) plane, and the km term can be interpreted as the maximum number of elements of A in a coset of a subgroup $H=LZ$.

Summing over C, since $|\mathcal{P}|\leqslant |A|m$, we get the analogue of (4.8) as follows:

$$E(A) \ll |A|^{5/2}m + |A|^2 M, \tag{4.12}$$

where M is the maximum number of elements in a coset of LZ. Hence, we have established the theorem. □

Having concluded the proof we will mention a few comments. First, the Heisenberg group consists of 3×3 matrices and so, whilst still being in some sense three dimensional (when defining a specific Heisenberg matrix we need to specify three entries, the g_1, g_2, and g_3), it is a step towards generalising Theorem 4.6 to $T_3(\mathbb{F}_q)$. Secondly, it is interesting due to containing an example of a set that does not grow and is not in a coset of an abelian subgroup but rather a 2-step nilpotent group and as such serving as an illustration as to why Theorem 4.4 concludes that sets in $T_n(\mathbb{C})$ with small tripling must have a large overlap with a coset of a nilpotent group of step $n-1$ (with n being from $n\times n$ matrices). This example also shows that Theorem 4.10 is sharp in some sense. Explicitly we may have that $m \gg \sqrt{|A|}$ and so we can have $E(A) \gg |A|^3$. For example, consider

$$A := \left\{ \begin{pmatrix} 1 & g_1 & g_3 \\ 0 & 1 & g_2 \\ 0 & 0 & 1 \end{pmatrix} : g_1, g_2 \in [1,\dots,n], g_3 \in [1,\dots n^2] \right\}. \tag{4.13}$$

Hence $A = \{(g_1,g_2,g_3)\} \cong [1,\dots n]\times[1,\dots n]\times[1,\dots n^2]$, so if $g\cdot h = s$ then

$$s_1 = g_1 + h_1, \qquad s_2 = g_2 + h_2, \qquad s_3 = g_3 + h_3 + g_1 h_2 .$$

Thus a typical $s \in AA$ will have $\sim n\times n\times n^2 = |A|$ representations as a product of two elements, and $|AA| \sim |A|$. Moreover, further multiplication by elements of A will not cause growth either.

Remarking now on similar results from the literature, Hegyvári and Hennecart [72] obtain lower bounds for size of $(A,B,C)^2$ (over $\mathbb{R}$ or $\mathbb{F}_P$) where (A,B,C) is the special Cartesian product subset of the Heisenberg where $g = (g_1,g_2,g_3) \in (A,B,C)$ if $g_1 \in A$, $g_2 \in B$, and $g_3 \in C$. Hegyvári and Hennecart also show that the fourfold product of a subset, $A = U \rtimes Z = \{(x,y,z) : (x,y) \in U, z \in Z\}$, of the Heisenberg group over $\mathbb{F}_p$ contains at least $|U|\left(1 - \frac{p^4}{\sqrt{2|A|^3}}\right)$ cosets of the type $(x,y,\mathbb{F}_p)$ in [71] and higher dimensional results of a similar flavour are found in [70].

We also remark on a stronger (for smaller values of m) bound for $E(A)$ due to Shkredov [156, Theorem 13] who estimated $E(A)$ in the special Cartesian product case, namely when each component of $g = (g_1,g_2,g_3)$ lies, independently, in some scalar set. This result would lead to the bound $|AA| \gg |A|^{7/4}$ if $m = 1$. Its proof follows similar methods but benefits from the Cartesian product setting enabling the application of n incidence bound twice, rather than once. Specifically note that in the case $m = 1$, then A is a graph over a set B in the (g_1,g_2)-variables, and $\sum_C |\mathcal{P}_C|^2$ is equal to the additive energy of B.

4.5 Higher dimension matrices

Having covered the contents of Murphy and my paper [113] we turn to the obvious question; what about higher dimension? Whilst I started to answer this question, predominately focusing upon the 3×3 case and the obstructions brought with it, whilst trying to follow similar methods as to the proof of Theorem 4.4, I did not fully answer this question. This section will seek to cover what I did do as well as consider the work of Eberhard, Murphy, Pyber, and Szabó [36].

4.5.1 3×3 Matrices

As stated above we will first consider the 3×3 matrices case, that is when does a subset $A \subseteq T_3(\mathbb{F}_q)$ (satisfying $e \in A$, and $A = A^{-1}$ for ease), have small tripling $|A^3| \leqslant K|A|$? To this end we recall the maps

$$\pi, \pi' : T_3(\mathbb{F}_q) \to T_2(\mathbb{F}_q)$$

defined in Section 4.1 where π maps to the 2×2 submatrix consisting of the top left corner, that is removing the bottom row and rightmost column and π' maps instead to the bottom right 2×2 submatrix. That is given the following matrix in $T_3(\mathbb{F}_q)$ then π corresponds to red submatrix and π' the blue one.

$$\begin{pmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{pmatrix}$$

The general idea is then to use Theorem 4.6 to restrict growth in the blue and red submatrices leaving only the elements in the c location to be dealt with. The aim then is to categorise when these can grow.

In Breuillard and Green's Theorem 4.4 this is done by induction on the dimension of the matrices (base case being trivial as 1×1 matrices as a group is abelian), here we will discuss the required steps for the 3×3 matrices. If A is our subset, then by the inductive hypothesis both $\pi(A)$ and $\pi'(A)$ have large subsets that are contained in cosets of an abelian (nilpotent of step 1) group and we can apply say π' to the large subset of $\pi(A)$ which also has a large subset, say A', contained in a coset of an abelian (nilpotent of step 1) group. They then proceed (after defining suitably symmetric sets) to use a lemma that this set is either generates a $(n-1)$-nilpotent group or resort to a sum-product bound to limit the number of ratios on the diagonal and find the required nilpotent group this way. The use of nilpotent groups is not enough in our case because one of our cases in Theorem 4.6 (the second) includes the case where we are filling a subfield that may not be abelian.

Having outlined our intent we now turn to some thought-provoking examples, most of the examples have a dual gained by reflecting in the line from the bottom left to the top right of the matrix, effectively swapping the roles of π and π'. For the examples, I just give a typical element

rather than the group for simplicity. First, note the group consisting of elements of the form

$$\begin{pmatrix} 1 & 0 & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

is an abelian group, so if we can deal with $\pi(A)$ and $\pi'(A)$ without them affecting the top right entry we are done, however, this is not a common case as will be seen in our future examples.

As a second example, the following has both the image of π and π' as a subset of the diagonal group, isomorphic to a subset of the multiplicative group of $\mathbb{F}_q$ (that is matrix multiplication corresponds to multiplication of entries in the "a" entry for π and the "f" term respectively for π').

$$\begin{pmatrix} a & 0 & c \\ 0 & 1 & 0 \\ 0 & 0 & f \end{pmatrix}$$

However, we may embed the whole T_2 example by just considering the corner entries. Whilst we know how to deal with the T_2 using Theorem 4.6, note that just using the homomorphisms π and π' would not catch this example.

Similarly to the above, we can have two copies of the T_2 embedded in our three-dimensional matrices by considering

$$\begin{pmatrix} a & 0 & c \\ 0 & d & e \\ 0 & 0 & f \end{pmatrix}.$$

In particular, we have one of our homomorphisms leading to the abelian torus case (from the a and d), the other (the d, e, and f) could be any of the cases from Theorem 4.6 but so too could the a, c, and f submatrix be considered as the 2×2 case in its own right. This manages to have the two 2×2 cases mostly independent (just sharing f rather than having major interaction under multiplication) unlike the more obvious example with the two 2×2 cases are $\pi(A)$ and $\pi'(A)$ who both interact with the c entry.

A rough idea of an approach would require you to consider three general cases. The simplest of our three cases is when we have few ratios of diagonal elements (we shall call these roots so the set of roots is $\left\{\frac{a_{ii}}{a_{jj}} : 1 \leqslant i < j \leqslant n\right\}$, this can be visualised as being close to the case

$$\begin{pmatrix} a & * & * \\ 0 & a & * \\ 0 & 0 & a \end{pmatrix},$$

and the step up from an $\mathbb{F}^*$-potent case from earlier. The second case is those matrices where we have a non-trivial action on the "c" entry, this is further split into when we have no small roots so the diagonal elements are all different and when we have some small roots (so diagonals consisting of (a,a,b) or (a,b,b)). The final third case is where our diagonal looks something like (a,b,a), so we have small roots but conjugation acts trivially on the upper right entry. The general

approach would then be to deal with the first case on its own using the pigeonhole principle to find a large subset with equal entries on the diagonal which is contained in a nilpotent subgroup and thus we are done. We may need some pigeonholing to get a suitable large set where the small roots are as desired (for example trimming away some of the small root cases). Having dealt with the first case, next you would identify the unipotent group before showing it has a generating set contained in a power of A. These apply to both case two and case three, they differ in that case two can then be finished by pivoting whilst the third case requires different methods.

Having identified the rough outline next we will break some of these steps down into more detail. In particular, we will consider what the unipotent subgroup is. The subgroup we want will be the last in the lower central series, in this three dimensional setting this is the subgroup $[G,[G,G]]$ where G is the group generated by our subset A. It is clear to see this can be the identity, (if G is for example $\mathbb{F}^*$-potent, going further the lower central series will not terminate at the identity only if there is an element whose diagonal entries are not all the same) which is reminiscent of the first case of Theorem 4.6. Sean Eberhard points out that this unipotent subgroup we are looking for can also be seen as $[T,U]$ where U is the unipotent subgroup of G, so $G \cap U_3(\mathbb{F}_q)$, and T is the image of G modulo U, or if we define a homomorphism $\phi : T_3(\mathbb{F}_q) \to D_3(\mathbb{F}_q)$ then $T = \phi(G)$. For the next steps, we shall name this unipotent subgroup we have found H. We shall consider some examples of what this H can be, starting simply it could be trivial as stated above, as in torus example in Theorem 4.6. At the other extreme, if H is the entirety of $U_3(\mathbb{F}_q)$, we are in the case where A is almost everything. Indeed we can see U, the unipotent bit of G as $U = C_U(T) * H$ where $C_U(T)$ is similar to the centraliser in that it is the elements of U fixed by T.

For the next point, that of showing H has a generating set contained in a power of A, the approach would be to use a pivoting argument to cover $H/[H,H]$ by some powers of A, case two can be finished by more pivoting whilst the third case is more involved. The last case deals with $[H,H]$ which can be seen as the set of unitriangular matrices(so ones on the main diagonal) with zeros on the first off-diagonal and something, possibly non-zero in the upper right. Note that $H/[H,H]$ can be seen as an abelian group whilst modulo $[H,H]$ as we are dealing with matrices of the form

$$\begin{pmatrix} 1 & a & * \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix},$$

which can be associated with just the entries (a,b). Then multiplying two of these elements together gives us

$$(a,b) * (c,d) = (a+c, b+d).$$

From this, we see that $H/[H,H]$ acts like a subgroup of $\mathbb{F}^2$ under vector addition and is thus abelian. Note that we really need to be working in the quotient group as most subsets B which project to some subset of $H/[H,H]$ are not of this form, that is the coset representatives are not necessarily (indeed usually are not) an abelian group although the quotient is. The subgroup $[H,H]$ is generated by commutators $[x,y]$ with $x,y \in H$. These commutators are determined by the cosets of x and y modulo $[H,H]$, that is in some way we can ignore the upper right entries, for now, so taking x and y as elements of $H/[H,H]$ which you would aim to show is an abelian

group covered by some power of A, say A^k, as above and thus A^{4k} contains a generating set of $[H,H]$ and the intersection of these two sets has small tripling. This means you would have a subset of an abelian group whose sumset is small and generated by elements of the form $ab-dc$ which should have some multiplicative structure and so the goal could be reached by leveraging this structure through expander graphs or Fourier analysis depending on preference.

4.5.2 $n \times n$ Matrices

Similar steps should work for even higher dimensions, however at each step, any extra difficulties and conditions will need to be propagated up, this, of course, being the obvious obstacle with the non-abelian option in dimension two. Whilst we do not provide further details here we make note of the work of Eberhard, Murphy, Pyber, and Szabó [36] which answer this question.

The obvious part will be the continuation of one case being mostly in a coset of a nilpotent subgroup of suitable step. How the second part generalises is a bit harder, although if we instead seek a result of the form of Theorem 4.2 the way ahead may be clearer.

Considering some of the steps from the previous subsection in this higher dimension case, the corresponding unipotent which is the last term of the lower central series is now more complicated to calculate than the prior $[G,[G,G]]$, however, the other version of considering $[T,U]$ where U is now $G \cap U_n(\mathbb{F}_q)$, and T is the image of G modulo U for our now higher dimension G and U, still works, as does considering it as the last term in the lower central series, the series is now longer so we are instead dealing with some similar longer term $[G,[\dots,[G,G]\dots]]$.

We finish by quoting the new result of Eberhard, Murphy, Pyber, and Szabó [36].

Theorem 4.11. *Let $\mathbb{F}_q$ be an arbitrary field. Let $A \subseteq GL_n(\mathbb{F}_q)$ be a finite symmetric subset such that $|A^3| \leqslant K|A|$. Then there are subgroups $H \trianglelefteq \Gamma \leqslant \langle A \rangle$ such that*

- *A is covered by $K^{O_n(1)}$ cosets of Γ,*
- *Γ/H is nilpotent of step at most $n-1$,*
- *H is contained in $A^{O_n(1)}$.*

As a final note, H may be taken as the last term of the lower central series of Γ.

Chapter 5

Incidence Bounds with Möbius Hyperbolae in Positive Characteristic

This chapter is based on the paper *Incidence bounds with Möbius hyperbolae in positive characteristic* [140] which is joint work with Misha Rudnev. We will start by setting up the chapter by providing the background of research the new results build upon as well as introducing the required terminology and ideas before turning to the proofs from the paper. We will also consider some applications (which we shall group at the end with the other applications) and directions further research could go. After considering [140], we will conclude the chapter by considering the joint work of myself and Audie Warren [197] which builds upon the earlier parts of the chapter, specifically generalising Theorem 5.5 to Theorem 5.8 and considering more consequences and corollaries from this now general incidence bound for Möbius transforms and points.

5.1 Background

Building on the prior general statements on incidences from Chapter 3, we will now look at incidences between points and hyperbola. Whilst a selection of results were covered earlier we will now focus specifically on hyperbola incidence results and the lay of the land leading to the paper which is this chapter's namesake.

Most incidence results concern themselves with lines and perhaps planes or circles. For hyperbola or other curves much less work has been done. First considering what is known over the real (and complex) numbers, Spencer, Szemerédi, and Trotter [169] note that in the proof of the Szemerédi-Trotter theorem (Theorem 3.1), affine lines can be replaced by pseudo lines which include circles (which allowed them to further work on the unit distance problem) and hyperbola. This thus gives a bound on the number of incidences between a point set $A \times A$ and a set of hyperbola H of

$$\mathcal{I}(A \times A, H) \ll \left(|A|^2|H|\right)^{2/3} + |H| + |A|^2.$$

Over the reals, the best known bound, for incidences between points and Möbius hyperbolae, is due to Solomon and Sharir [149].

Theorem 5.1. *[149] Let $A \subset \mathbb{R}$ be a set of n real numbers, and consider the set of Möbius transformations on $\mathbb{R}$, the number of k-rich transformations is bounded by*

$$m_k \ll \frac{|A|^4}{k^3} + \frac{|A|^6}{k^{11/2}} \log k \, .$$

We remark that over $\mathbb{C}$ the best known to my knowledge bound is

$$m_k \ll \frac{|A|^6}{k^5} \, ,$$

due to Solymosi and Tardos [164].

Turning to the finite fields case, as stated in Chapter 2, our best Szemerédi-Trotter equivalent is due to Stevens and de Zeeuw [172] (Theorem 5.5) however its proof is not as amenable to have lines replaced by psuedolines and in particular translates of the hyperbola $y = \frac{1}{x}$. The previous generation of incidences theorems in positive characteristic came from combining concepts from line geometry in the projective three-space (for example see [132, 135, 204]) with Guth and Katz's algebraic theorem [67]. These incidence results are also not particularly able to have the lines changed for hyperbola, with the difficultly heuristically being that they can only be as non-linear as the Guth-Katz theorem allows. As we cannot easily change the current line incidence results in finite fields to hyperbola results as in the reals, they have instead been studied as their own thing.

The first non-trivial incidence result for hyperbolas in finite fields is due to Bourgain [10] and is a qualitatively nontrivial bound which we stated as Theorem 3.5. This theorem is for a three-parameter set of hyperbola H rather than the two parameter Cartesian product of the forthcoming Theorem 5.2 or the two parameter general set of hyperbola in mine and Rudnev's results (Theorems 5.3 and 5.4) stemming from translates of $y = \frac{-1}{x}$. What Theorem 3.5 tells us is that there is a nontrivial incidence estimate for $\sigma(A,H)$ in our finite field $\mathbb{F}_p$ case if; A is not all of $\mathbb{F}_p$ (that is $|A| < p^{1-\varepsilon}$), and the number of Möbius hyperbolae $|H|$ is essentially greater than $|A|$. There is also the additional assumption that much of H cannot lie in a coset of a proper subgroup of $SL_2(p)$ ($|H \cap gS| < |H|^{1-\varepsilon}$) can be weakened to the subgroup being abelian, owing to the recent energy bounds in the affine group by Petridis et al [121].

A quantitative result followed[1] due to Shkedrov [157] which can be regarded as a special case of the above Theorem 3.5.

Theorem 5.2. *[157, Shkredov, Theorem 16] Let $A \subseteq \mathbb{F}_p$ and $H = B \times B \subseteq \mathbb{F}_p^2$ be a set of translates*

[1]We quote a symmetric variant rather than the original formulation of Shkredov's which has two Cartesian products involving four scalar sets. The original can be found in Chapter 3 as Theorem 3.6.

of the hyperbola $xy=-1$. Then

$$\left|\sigma(A,H)-\frac{|A|^2|H|}{p}\right| \lesssim \min\Big(|A|^{1/2}|H|+|A|^{3/2}|H|^{3/4}, \tag{5.1}$$
$$|A|^{3/4}|H|+|A|^{5/4}|H|^{41/48}\Big).$$

The right-hand side of (5.1) is split into two lines, the first line is nontrivial when $|H|>|A|^2$ and it is the second term which should be seen as the main one. The second line provides a non-trivial estimate for $|H|\gtrsim|A|^{12/7}$ and it is this second line from which the logarithmic factor is hidden (there are no logarithms hidden in the first line) by the $\lesssim$ symbol.

These are the only two preceding results for hyperbola incidences before our results (Theorem 5.3 and 5.4) although we also note Helfgott's [73] result for growth in subsets of $SL_2(\mathbb{Z}/p\mathbb{Z})$ which has obvious connections once we associate hyperbola with elements of $SL_2(\mathbb{Z}/p\mathbb{Z})$ and is used by Bourgain's Theorem 3.5.

We will briefly outline the proofs here so when we consider the proofs of Theorem 5.3 and 5.4 later in this section we can compare them. Bourgain's Theorem 3.5's proof is based upon repeated applications of the Cauchy-Schwarz inequality where after each application the input H, viewed as a set of SL_2-transformations, is replaced by $H'=H\circ H^{-1}$ (we just write HH^{-1} and do not make a distinction between SL_2 and PSL_2). After each application of the Cauchy-Schwarz inequality, you can split (again by Cauchy-Schwarz) the count into an estimate for the energy of the set of Möbius transformations

$$E(H):=\left|\left\{(h_1,h_2,h_1',h_2')\in H^4 : h_1h_2^{-1}=h_1'h_2'^{-1}\right\}\right|, \tag{5.2}$$

and the quantity $\sigma(H',A)$. For this later quantity both Bourgain and Shkredov use a trivial bound. It is estimating the quantity $E(H)$ and its further iterates,

$$T_k(H):=\left|\left\{(h_1,\dots h_k,h_1',\dots h_k')\in H^{2k} : h_1h_2^{-1}h_3\cdots=h_1'h_2'^{-1}h_3'\dots\right\}\right|,$$

that provide nontrivial savings. Bourgain makes this saving from repeated applications of the L^2-smoothing lemma of Bourgain and Gamburd [11], each application gaining a very small saving. Note another proof can be found in the appendix of [155]. The small saving comes from a combination of Helfgott's theorem on growth and expansion in $SL_2(p)$ [73] and the (non-commutative) Balog-Szemerédi-Gowers theorem in $SL_2(p)$. After taking a sufficient (large) number of iterations the claim is proven however extracting a quantitative lower bound on δ in Theorem 3.5 seems prohibitively difficult.

The proof of Shkredov's Theorem 5.2 instead uses two applications of Cauchy-Schwarz (making use of bounds on T_2 and T_3 energies rather than much larger energies as Bourgain did) to gain the quantitative bound for $\sigma(A,H)$ of the theorem. This seems to rely on H being a two-parameter family. Similar non-commutative energy estimates can be found in [121] and [113], the latter of which is covered in-depth in Chapter 4. Shkredov still uses a trivial estimate for

$\sigma(H', A)$ however has a much stronger explicit sum-product type L^2-estimate (Lemma 5.3) for $H' = HH^{-1}HH^{-1}$. It is this second part that is currently reliant on the set of translates H being two-dimensional as we do not know a quantitative estimate for $E(H)$, where H is a (sufficiently small relative to p) general set of $SL_2(p)$ transformations. Should such an estimate be found it would imply directly a variant of Helfgott's theorem on growth and expansion in $SL_2(p)$ [73] and consequently make much of that proofs machinery redundant.

5.2 Minkowski Distances

During this chapter, we will make use of and find connections to *Minkowski* or *pseudo-Euclidean* distances. This section aims to introduce this distance, provide background and build the foundation of the intuition as to why these distances are cropping up and why they are of use to us. We will begin with the following definition.

Definition 5.1. *A space with a quadratic form is called a* pseudo-Euclidean space.

This is not the most enlightening. A *quadratic form* $\mathbf{q} : \mathbb{F}^n \to \mathbb{F}$ where $x \mapsto \mathbf{q}(x) := x^T Q x$ for a symmetric map Q. $\mathbf{q}$ is the analogue of the square of the norm, so for our standard intuition in Euclidean space, $x^2 + y^2$. Luckily for us, these can be classified by pairs $(k, n-k)$ so in some basis of $\mathbb{F}^n$ we can define

$$\mathbf{q}(x) := (x_1^2 + \ldots x_k^2) - (x_{k+1}^2 + \ldots x_n^2).$$

Note that $(n, 0)$ would be standard Euclidean space, and if -1 is a square in the field, $\mathbb{F}$, then working over the pseudo-Euclidean space is the same as the Euclidean space. Due to its origins in physics, the first k coordinates are referred to as space-like and the rest as time-like.

For our results in this chapter, we will work in the specific pseudo-Euclidean space $(1, 1)$ over finite fields. As such, for us, the Minkowski distances will be defined as follows:

Definition 5.2. *The Minkowski distance between points* $q = (x, y), q' = (x', y') \in \mathbb{F}^2$ *is*

$$D_M(q, q') := (x - x')^2 - (y - y')^2.$$

We will also introduce the notion of the isotopic cone, that is all those points that are distance zero from the origin, $\mathbf{q}^{-1}(0)$ and similarly isotropic distances as those distances between distinct points which are zero.

We will next take a short detour into the history and origins of this quantity before returning to consider its properties. Named for Hermann Minkowski, who introduced the distance in the four-dimensional case (technically that is $(3, 1)$ in the above definitions) as a way to understand his student's[2] theory of special relativity. Minkowski [102] built on Poincare's [124] work who considered rotations of the four dimensional Euclidean sphere $x^2 + y^2 + z^2 + (ict)^2 = C$. Minkowski used real coordinates for time rather than imaginary in [101] leading to the concept of spacetime.

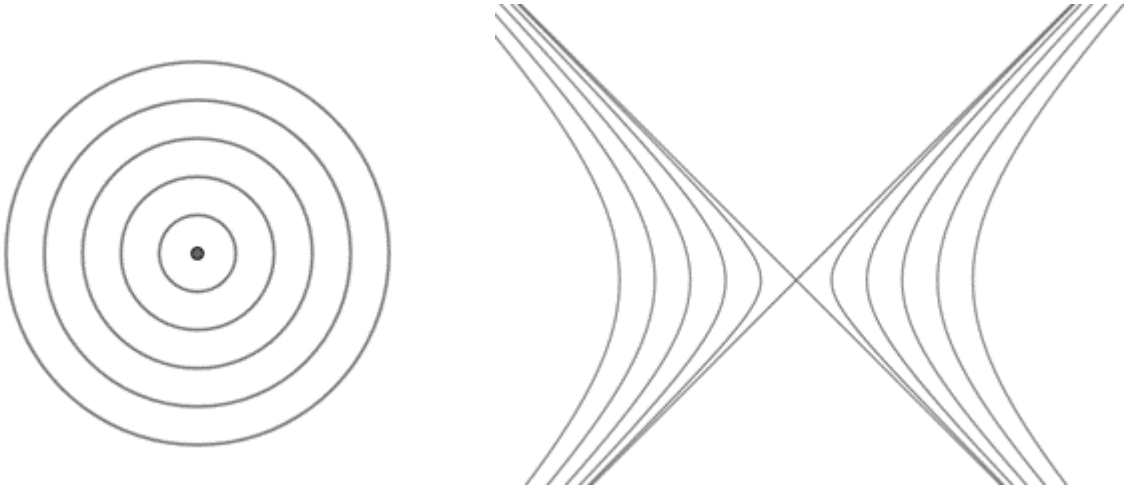
[2] Albert Einstein

Einstein took a further step when he moved to using Riemannian geometry to include gravitation. Minkowski space is one of the most studied pseudo-Euclidean spaces. The geometry associated was studied by Poincaré [124], this involved looking at its rotations, the Lorentz group, and the Poincaré group when translations are also included. The Poincaré group is the analogue of Euclidean groups in ordinary Euclidean spaces.

Although for physicists it is the four dimensional space that is of use, mathematicians have long baulked at fixing dimensions when they can instead generalise to an arbitrary dimension. Our other main example and the one we will be interested in for the proofs of our results later in this chapter is the *Artinian plane*, *pseudo-Euclidean plane*, or *hyperbolic plane*, explicitly it is pseudo-Euclidean space corresponding to (1,1). In this example, our quadratic form is $x^2 - y^2$. This can also be considered as the plane of split-complex numbers. This is also the only example where the isotopic cone (the cross formed by the asymptotes of $x^2 + y^2 = c$) splits the space into four open sets.

We now turn to some properties. To start with we note that this is not a metric and nor are our examples above metric spaces. This is obvious as there exist points that are distance zero apart (the isotropic distances) but distinct points, for example, the origin and a point (x,x). Instead, it is a pseudo-metric. This leads to some major differences in comparison with standard Euclidean space.

The equidistant sets of the pseudo-Euclidean and standard Euclidean space are enlightening, Figure 5.1 has five equidistant sets for standard Euclidean space which are circles and the same five equidistant sets but in our (1,1)-pseudo-Euclidean example. That they are hyperbola should give a good clue as to why they may be of interest later in this chapter. We note that as we work over finite fields -1 may be a square and thus the two images are identical.



(a) Sets of equidistant in the Euclidean plane.

(b) Sets of equidistant in the pseudo-Euclidean plane.

Figure 5.1: The differences in equidistant sets.

We also note that with a transformation of $x = x' + y'$, $y = x' - y'$ we rotate Figure 5.1b so that the equidistant sets are in the same orientation as $y = \frac{1}{x}$, that is they share their asymptotes.

This changes the Minkowski distance to being the area of the rectangles formed with the points, (x,y), (x',y') as two opposite corners, $(x-x')(y-y')$. Roche-Newton and Rudnev consider both Minkowski distances and these rectangular areas in [127], in particular Theorem 3.9 stated earlier.

5.3 Incidence Result for Möbius Hyperbola

In this section we will talk about trivial observations for bounds on incidences between points and hyperbola before stating our theorems and comparing the methods of their proofs with previous research, we will then turn to the proofs themselves.

Before we get started we will first define our notation for the rest of the chapter. We define our point set $P = A \times A \subset \mathbb{F}^2$ and H a set of translates of the hyperbola $xy = -1$ in the form

$$y = a + \frac{1}{b-x} : (a,b) = h \in H.$$

We identify H with the set of Möbius transformations

$$h(x) = a + \frac{1}{b-x}. \tag{5.3}$$

We will also represent the hyperbola $y = a + \frac{1}{b-x}$ by the SL_2 matrix $h = \begin{pmatrix} -a & ab+1 \\ -1 & b \end{pmatrix}$.

We define

$$\sigma(A,H) := \sum_{h \in H} \sum_{x \in A} \mathbb{1}_A(h(x))$$

as the number of incidences between points in $A \times A \subset \mathbb{F}^2$ and hyperbolae in H. The analysis below extends trivially to the case $xy = \lambda$ where the -1 has been replaced by any other nonzero $\lambda \in \mathbb{F}$. Indeed, if $-\lambda$ is a square, this is seen by scaling the whole problem, that is $xy = \lambda$ becomes $\frac{x}{\sqrt{-\lambda}} \frac{y}{\sqrt{-\lambda}} = x'y' = -1$. Besides, the forthcoming analysis applies to the translates of the hyperbola $xy = 1$ as well, simply by replacing the $a + \ldots$ in (5.3) by $a - \ldots$. Also note that these results apply to the case of two sets A, A' of the same cardinality, in which case the value $\lambda = -1$ is set by scaling one of the sets.

Turning now to the simple observations, each hyperbola can contain at most $|A|$ points in $A \times A \subset \mathbb{F}$ leading to a trivial bound of $\sigma(A,H) \sim |A||H|$. We now provide a trivial example which exhibits $\sigma(A,H) \ll |A||H|$. Consider A_1 an arithmetic progression, $A_2 = \frac{1}{A_1}$ and $A = A_1 \cup A_2$. For H we take the horizontal translates of $xy = -1$ by elements of A_1. None of these hyperbolae intersect (this is the trivial aspect) and all support $\sim |A|$ points so the trivial estimate is reached. This tells us that when $|H| \ll |A|$ we cannot do better than the trivial estimate as we can in some sense fit all the hyperbola in without intersections so each can act in the trivial example similar to when all the points are on a line from Chapter 3. As such we will predominately deal with the case $|H| > |A|$.

Finally, before turning to the results we shall quickly consider what incidences between the translates of $y = \frac{1}{x}$ look like. We note that we can have up to two incidences such as between the

red and green or blue and green hyperbolas in Figure 5.2 or just a single incidence as between blue and red. We could also have no incidences such as two hyperbolae which are horizontal (equivalently vertical) translates or like a slightly perturbed blue and red example.

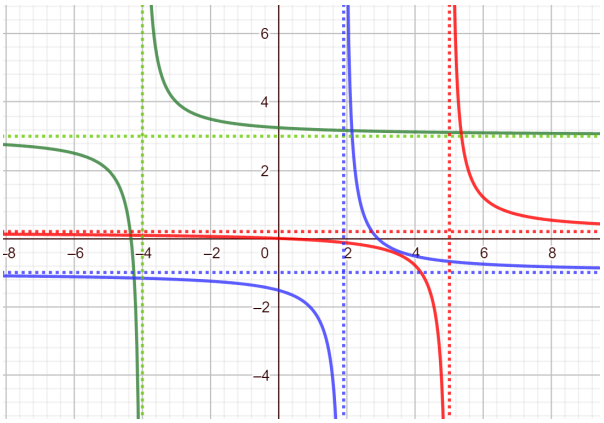


Figure 5.2: Some example incidence between translates of the hyperbola $y = \frac{1}{x}$.

5.3.1 Our Results

Our main result (Theorem 5.3) strengthens Shkredov's Theorem 5.2. We will state it here, comment briefly on the key comparisons of the proof with those of Theorem 5.2 before proving an intermediate theorem in the next subsection and the full proof in the following. We will assume our energy estimates and cover them separately in Section 5.4.

Theorem 5.3. *Let $A \subset \mathbb{F}$, with $|A| < \sqrt{p}$ if $\mathbb{F}$ has positive characteristic p. For a set of translates H of the hyperbola $y = \frac{-1}{x}$, with $|H| > |A|$, suppose at most M translates $(a,b) \in H$ have the same abscissa or ordinate. Then with*

$$M_1 = \begin{cases} M, & \text{if } |H| \leqslant |A|^{3/2}, \\ |H|^{2/11}|A|^{8/11} & \text{otherwise} \end{cases},$$

one has the estimate

$$\sigma(A,H) \ll |A|^{1/2}|H| + |A|^{\frac{6}{5}}|H|^{\frac{4}{5}}M_1^{\frac{1}{10}}. \tag{5.4}$$

Furthermore, with

$$M_2 = \begin{cases} M, & \text{if } |H| \leqslant |A|^{4/3}, \\ |H|^{3/22}|A|^{9/11} & \text{otherwise} \end{cases},$$

one has the estimate

$$\sigma(A,H) \ll |A|^{3/4}|H| + |A|^{11/10}|H|^{17/20}(M_2^{1/10} + |H|^{1/15}) \tag{5.5}$$

where $(M_2^{1/10} + |H|^{1/15})$ can be replaced by $|H|^{1/16}$ if $H = B \times B$.

Estimate (5.4) is using T_2 energy bounds, whilst Estimate (5.5) uses T_3 energy bounds and they are generalisations of the first and second line of Shkredov's Estimate (5.1) respectively.

Considering when they are non-trivial, Estimate 5.4 is nontrivial for $H > |A|^{3/2}$ and if we assume $M \leqslant |H|^{1/2}$ (as it is in Theorem 5.2) it is non-trivial for $|H| > |A|^{4/3}$. Estimate (5.5) is always nontrivial for $|H| > |A|^{4/3}$, and assuming $M \leqslant |H|^{2/3}$, is non-trivial for $|H| > |A|^{6/5}$. Estimate (5.5) yields a better bound for rich hyperbolae, and we note if we continue the method (assuming we could get the corresponding energy bounds) each step of Cauchy-Swartz we would do would give better results only for more and more rich hyperbolae. For most applications it is Estimate (5.4) we will use as it is superior in the regime when $|H| \sim |A|^2$ which is of most interest. However, the proof behind Estimate (5.5) reveals interesting connections with the Erdős distinct distance problem via Minkowski distances, further building on their obvious natural connections to hyperbolae (recall Figure 5.1).

Our second theorem (Theorem 5.4) is an adaptation of Theorem 5.3 to the case $\mathbb{F} = \mathbb{F}_p$, where we provide additional estimates which cover the case $|A| \geqslant \sqrt{p}$. As we will be considering the case where $|A|$ and $|H|$ are large we note the character sum estimate by Iosevich et al. [69], which is best possible for $|A|^2|H| > p^3$, and trivial when $|H| < p$:

$$\sigma(A,H) \leqslant \frac{|A|^2|H|}{p} + 2|A|\sqrt{p|H|}.$$

We note the similarities to such results considered in Section 3.1.6.

Theorem 5.4 also briefly deals with the real case, $\mathbb{F} = \mathbb{R}$. this is done via a recalculation, using instead the state of the art incidence theorem (Theorem 5.1) for modular hyperbolae due to Sharir and Solomon [149]. This is done as when we have the assumption that the number of translates $(a,b) \in H$ with the same abscissa or ordinate is $O(|H|^{1/2})$, then for $|A| \ll |H| \lesssim |A|^{19/13}$, Estimate (5.6) is stronger than the Szemerédi-Trotter estimate $\sigma(A,H) \ll |A|^{4/3}|H|^{2/3}$. If $|H| \lesssim |A|^{16/13}$, this is true even without any assumptions on H. This does not help with the preferred range $|H| \sim |A|^2$ which is where we would truly like to be able to beat the Szemerédi-Trotter estimate.

Theorem 5.4. *Assume the notations of Theorem 5.3.*

Let $\mathbb{F} = \mathbb{F}_p$. If $|A||H|^2 \leqslant p^3$, one can remove the constraint $|A| < \sqrt{p}$ as to (5.4) *and have it with the extra term $\dfrac{|A|^{5/4}|H|}{p^{1/4}}$ in the right-hand side.*

Furthermore,[3] *if $|A||H|^4 \leqslant p^5$, one can remove the constraint $|A| < \sqrt{p}$ as to Estimate* (5.5) *and have it with the extra term $\dfrac{|A|^{9/8}|H|}{p^{1/8}}$ in the right-hand side.*

If $A \subset \mathbb{R}$, then

$$\sigma(A,H) \lesssim |A|^{1/2}|H| + |A|^{7/6}|H|^{2/3}M_1^{1/6} + |A|^{\frac{23}{22}}|H|^{\frac{9}{11}}M_1^{\frac{1}{11}}. \tag{5.6}$$

Having now stated both of these results we now give a brief overview of the proof (which will be presented in Subsection 5.3.3) and compare it to the prior literature. Our proof of Theorem 5.3

[3] One can add more intermediate range estimates by fetching more cases from the forthcoming Lemma 5.4 but they do not appear to be sufficiently enlightening.

follows Shkredov's proof of Theorem 5.2 with two main points of deviation. These being a new "intermediate" incidence bound in Theorem 5.5 and noting and leveraging the connection of the quantity $T_3(H)$ with the L^2-moment of the Minkowski distance in the set H, this is contained in Lemma 5.2 and will be further discussed in Section 5.4. The new "intermediate" incidence bound of Theorem 5.5 is a corollary of the Stevens - de Zeeuw incidence bound (Theorem 3.4) for lines and points. After one application of Cauchy-Schwarz, this allows us to make the improvements on Theorem 5.2 (in particular the first term in the right-hand side of Estimate (5.1)) resulting in Estimate (5.4). Lemma 5.2 generalises Shkredov's sum-product type bound, Lemma 5.3 which is the particular case of H being a Cartesian product (and in this case slightly stronger being able to leverage more out of this fact). Whilst we use the connections to the Minkowski distance, Shkredov made use of a sharp Euclidean bound by Guth-Katz [67] of the Erdős distinct distance problem [45]; it was adapted to the Minkowski or pseudo-Euclidean distance in [127].

Contrasting with Bourgain's proof of Theorem 3.5, further iterating Cauchy-Schwarz does not create new savings for us, this is due to the fact that for $n > 3$ we do not have a way of getting strong quantitative estimates for $T_n(H)$ except using the L^2-smoothing lemma. On top of this, the iteration allows us to get non-trivial bounds for richer hyperbola, explicitly on the kth step of the iteration one can only get a nontrivial bound on the number of hyperbolae in H, which are $|A|^{1-2^{-k}}$-rich. This can be seen by the very first terms on the right-hand side of Estimates (5.4) and (5.5). Also, the efficiency of using a nontrivial incidence bound for $\sigma(A,H')$ (our intermediate theorem) decreases with each step of the iteration.

We will further generalise these results as joint work with Audie Warren later in this chapter, specifically Section 5.5.

5.3.2 Intermediate Incidence Bounds for Möbius Hyperbolae

In this section, we will state and prove our intermediate Theorem 5.5. A Möbius Hyperbola is identified with a SL_2 (or PSL_2) transformation, this is the same as in the statement of Theorem 3.5. We will also use the definition that a hyperbola (transformation) h is k-rich if it supports $\geqslant k$ points of $A \times A$, namely $|A \cap h(A)| \geqslant k$. Further details can be found in Section 3.1.4.

Our intermediate theorem is a bound for incidences between points and Möbius hyperbolae. Over the reals, we previously stated the best known such bound due to Solomon and Sharir [149] (Theorem 5.1) and similarly Solymosi and Tardos [164] for complex numbers. Theorem 5.5 is a weaker analogue in positive characteristic.

Theorem 5.5. *Let $A \subset \mathbb{F} = \mathbb{F}_p$ and H be a set of $m > |A|$ Möbius hyperbolae in $\mathbb{F}^2$. Then the number of incidences between $P = A \times A$ and H satisfies*

$$\sigma(A,H) \ll \frac{|H||A|^2}{p} + |A|^{1/2}|H| + \min\left(|A|^{7/5}|H|^{4/5}, p^{1/3}|A|^{4/3}|H|^{2/3}\right). \tag{5.7}$$

Moreover, if $k > \sqrt{|A|}$ and $|A| < \sqrt{p}$, then for any $\mathbb{F}$ of positive characteristic p, the maximum number of k-rich Möbius hyperbolae is $O\left(\frac{|A|^7}{k^5}\right)$.

The proof of this result will require the following two incidence statements: the Stevens - de Zeeuw theorem [172] (this is stated slightly more generally in Chapter 3 as Theorem 3.4) and its corollary, Lemma 3.2 from [111] in the specific case of $\mathbb{F} = \mathbb{F}_p$.

Theorem 5.6. *The number of incidences between the point set $P = A \times A$ and a set L of affine lines in $\mathbb{F}^2$, with $|A||L| < p^2$ is*

$$\mathcal{I}(P,L) \ll |A|^{5/4}|L|^{3/4} + |L| + |A|^2.$$

Lemma 3.2. *Let $A \subset \mathbb{F}_p$ and let $2|A|^2/p \leqslant k \leqslant |A|$ be an integer that is greater than 1. The number, l_k, of k-rich lines satisfies*

$$l_k \ll \min\left(\frac{p|A|^2}{k^2}, \frac{|A|^5}{k^4}\right).$$

Before we begin the proof we note the following key fact which underlies the proof that (non-horizontal and non-vertical) lines in $A \times A$ can be viewed as affine (a particular case of Möbius) transformations.

The proof is the same as in mine and Rudnev's paper [140], a more detailed proof of a generalised version can be found later in this thesis (in particular this is the proof of Theorem 5.8 in Section 5.5.1) after that proof details will be made clear about where this varies from the following proof.

Proof of Theorem 5.5. Let $q = (a,a') \in A \times A = P$ be a point in our point set. Let $H_q \subseteq H$ be the subset of hyperbolae incident to the point q. Similarly to above, for $k \geqslant 1$, refer to a hyperbola of H_q as k-rich if it supports at least k points of $A \times A$ different from $q = (a,a')$. Next, we identify H_q with the set of (projective) Möbius transformations, M_q (that is liner-fractional maps $f(z) = \frac{az+b}{cz+d}$ with $ad - bc \neq 0$) mapping a to a'. Also for a $g \in M_p$ we have that $\frac{1}{a'-z} \circ g \circ (a - \frac{1}{z})$ maps infinity to infinity. The use of projective mappings allows us to deal with the points at infinity and deal with the issues which arise from division by zero.

This shows that the number of incidences between H_q and points in P other than q is equal to the number of incidences between $m = |H_q|$ affine lines and the point set $B \times C := \frac{1}{a-A} \times (a' - A^{-1})$. Note $|B| = |C| = |A|$. This is because after applying the required transformations to get our curves to be lines we apply the same transformations to our pointset. Examples of the explicit calculations proving this statement can be seen in the proof of Theorem 5.8.

Next, we use Lemma 3.2 to bound the number of k-rich lines when

$$\max\left(\frac{2|A|^2}{p}, 2\right) \leqslant k \leqslant |A|.$$

Note that in the more general case of $\mathbb{F} \neq \mathbb{F}_p$, (where we are dealing with the second part of Theorem 5.5 and providing a bound for the number of k-rich Möbius hyperbolae) we have taken $|A| < \sqrt{p}$, so $\frac{2|A|^2}{p} \leq 2$. Over $\mathbb{F}_p$, when $2 \leqslant k < \frac{2|A|^2}{p}$ one cannot have any nontrivial incidence bound, which accounts for the first term in Estimate (5.7). We also add the trivial Cauchy-Schwarz bound $l_k \leq |A|^4/k^2$.

Combining this with Lemma 3.2 yields the following bound on the number, m_k, of k-rich transformations in H_q by

$$m_k \ll \min\left(\frac{|A|^5}{k^4}, \frac{p|A|^2}{k^2}, \frac{|A|^4}{k^2}\right).$$

The third term (the one coming from simple Cauchy-Schwarz) in the latter estimate is smaller than the first one if $k \leqslant \sqrt{|A|}$. We proceed, assuming that $k > \sqrt{|A|}$, accounting for the case to the contrary by including the second term in the estimate of the theorem.

To continue, we sum over $q \in A \times A = P$ observing that we count each k-rich hyperbola in H at least k times. This bounds m_k, the number of k-rich hyperbolae in H as follows:

$$m_k \ll \min\left(\frac{|A|^7}{k^5}, \frac{p|A|^4}{k^3}\right). \tag{5.8}$$

The proof is concluded by the standard conversion of the latter estimate into an incidence bound. Assuming $m_k \ll |A|^7/k^5$ and we take some $k = k_*$ and optimise between the estimate $\ll \frac{|A|^7}{k_*^5}$ for the number of incidences, supported on k_*-rich hyperbolae and $\ll mk_*$ for the rest of the hyperbolae. Choosing $k_* = |A|^{7/5}|H|^{-1/5}$ accounts for the first term under the minimum in the theorem's claim. Doing the same thing assuming $m_k \ll \frac{p|A|^2}{k^2}$ accounts for the remaining term and completes the proof in the case $\mathbb{F} = \mathbb{F}_p$.

In the case of a general field $\mathbb{F}_q$, we note that once we are only interested in $k \geqslant \sqrt{|A|}$, the constraint $|A| < p$ and the trivial estimate $|A|^4/k^2$ on the number of k-rich lines guarantee that the condition $|A||L| < p^2$ of Theorem 5.6 is satisfied as to the set L of k-rich lines, and hence one has $|L| \ll |A|^5/k^4$ as was used above.

□

5.3.3 Proof of Theorems 5.3 and 5.4

We will now provide the proofs of Theorems 5.3 and 5.4. To do this we will assume some energy lemma (Lemma 5.1 and Lemma 5.2) which we will prove and discuss in Section 5.4. The general outline of the proof is we will use Cauchy-Schwarz to split our bound into an energy part dealt with by our energy Lemmata and a part dealing with incidences and Möbius hyperbolae which we deal with using our intermediate Theorem 5.5. We prove Theorem 5.3 and Theorem 5.4 together, presenting the proof of Theorem 5.3 and add additional remarks in the special case $\mathbb{F} = \mathbb{F}_p$ pertaining to Theorem 5.4 when we allow $\frac{|A|^2}{p} \gg 1$. The proof is the same as in mine and Rudnev's paper [140] with added detail.

Proof. To prove the Bound (5.4) we start with pruning away the set of translates of the hyperbola that lie on the union of a small number of very rich vertical or horizontal lines. This is done only if $M > |A|$, otherwise at this stage we do nothing. Let H_1 denote the translates, lying on at most $\frac{|H|}{x|A|}$, say vertical lines with at least $x|A|$ translates per line, for some $x \geqslant 1$. They contribute,

trivially, at most $\frac{|A||H|}{x}$ to the quantity $\sigma(A,H)$. Assuming $M = x|A|$ we determine x by setting

$$\frac{|A||H|}{x} = |A|^{6/5}|H|^{4/5}(x|A|)^{1/10},$$

the right-hand side being the bound we will prove in the immediate sequel. This means $x = |H|^{2/11}/|A|^{3/11}$. This is exceeds 1 only if $|H| > |A|^{3/2}$, in which case we have a saving that determines the choice of M_1 apropos of Estimate (5.4), hence $x|A| = |H|^{2/11}|A|^{8/11}$. This determines the choice of M_1.

We do the same thing concerning the Bound (5.5), where we interpolate

$$\frac{|A||H|}{x} = |A|^{11/10}|H|^{17/20}(x|A|)^{1/10},$$

then $x = \frac{|H|^{3/22}}{|A|^{2/11}} > 1$ if $|H| > |A|^{4/3}$, this determines the choice of M_2.

We now move on to the of proving (5.4). Retaining the notation H for the remaining set of translates, apply Cauchy-Schwarz to the summation over A, with a shortcut $\sigma = \sigma(A,H)$ we get the following:

$$\sigma^2 = \left(\sum_{h\in H, a\in A} \mathbb{1}_A(ha)\right)^2 \leqslant |A| \sum_{u\in HH^{-1}} r_{HH^{-1}}(u) \sum_{a\in A} \mathbb{1}_A(ua) \tag{5.9}$$

We will use this argument several times later in this thesis (for example in the proofs of Theorems 5.9 and 5.10 and has been used in other work including [137] and [157]), as this is an important argument and step we will go into depth as to exactly what is happening here whilst later we will just claim that by Cauchy-Schwartz we arrive at Inequality (5.9) analogue for that specific argument. To this end of fully explaining the steps behind Inequality (5.9) we start with the following lines of inequalities before explaining each line below.

$$\begin{aligned}
\sigma^2 &= \left(\sum_{h\in H, a\in A} \mathbb{1}_A(ha)\right)^2 \\
&\leqslant_{C.S.} \sum_{a\in A} 1 \cdot \sum_{a\in A} \left[\sum_{h\in H} \mathbb{1}_A(h(a))\right]^2 \\
&\leqslant |A| \sum_{a\in A} \sum_{u\in HH^{-1}} \mathbb{1}_A(ua) r_{HH^{-1}}(u) \\
&\leqslant |A| \sum_{u\in HH^{-1}} r_{HH^{-1}}(u) \sum_{a\in A} \mathbb{1}_A(ua)
\end{aligned} \tag{5.10}$$

Starting at the top, the first line is just the definition of the incidences, that is the number of incidences is equal to the sum of the number over all hyperbola of the number of points each hyperbola intersects, we have squared both sides in anticipation of the upcoming Cauchy-Schwartz step. The second line is this Cauchy-Schwartz step, specifically, we are using Cauchy-Schwartz in the summation over $a \in A$ with our a_is as one. It is the next step that has been hidden, we have stated this is just by Cauchy-Schwartz and the moving from line two to line

three is done by reevaluating what the sum squared looks like. We will now consider the object

$$\left[\sum_{h\in H}\mathbb{1}_A(h(a))\right]^2$$

in greater detail. Our first, trivial observation is that the terms $\mathbb{1}_A(h(a))$ are either one or zero and if they are zero they obviously cannot contribute to the sum. The second is that once we have expanded out the square each term is the product of two such $\mathbb{1}_A(h(a))$, say $\mathbb{1}_A(h(a))\mathbb{1}_A(h'(a))$, noting as we sum over the $a\in A$ later, a is currently acting as a constant. Note that $\mathbb{1}_A(h(a))\mathbb{1}_A(h'(a))$ can only be non-zero if both $\mathbb{1}_A(h(a))$ and $\mathbb{1}_A(h'(a))$ are non-zero, otherwise the term can be thrown away. We are thus wishing to count the number of pairs oh hyperbola h, h' such that both transform a to another point in A. That is if $h(a)=b\in A$ and $h'(a)=b'\in A$ then hh'^{-1} maps an element in A (specifically b') to another (to b) and so both $\mathbb{1}_A(h(a))$ and $\mathbb{1}_A(h'(a))$ being non-zero is the same as $\mathbb{1}_A(hh'^{-1}(b'))$ being non zero, this looks fine bar that our a is now b' but as we are summing over all $a\in A$ which includes b' all this means is we are counting them in a different order but importantly are still counting them. Another way to consider this is the number of maps consisting of a composition of a hyperbola taking a to A and an inverse hyperbola mapping back into A, that is a $u\in HH^{-1}$ which maps a into A, however, we must then be careful as some of these u's will be more popular and need to be counted multiple times as there are multiple ways to make them, that is we can replace

$$\left[\sum_{h\in H}\mathbb{1}_A(h(a))\right]^2$$

with

$$\sum_{u\in HH^{-1}}\mathbb{1}_A(ua)r_{HH^{-1}}(u).$$

The final line comes from reordering summation to make future Cauchy-Schwartz more clear with the $r_{HH^{-1}}(u)$ going to become our energy part and the $\mathbb{1}_A(ua)$ going to be bounded via our intermediate Theorem 5.5. Note that in the higher energy case we effectively add another step of Cauchy-Schwartz so we are dealing with $HH^{-1}HH^{-1}$ rather than HH^{-1}.

Having now established the reasoning behind Inequality (5.9) we aim to move to a specific rich case of hyperbola, to do this w set

$$\Delta := \frac{\sigma^2}{3|A||H|^2}. \tag{5.11}$$

By the pigeonhole principle, since $\sum_{u\in HH^{-1}}r_{HH^{-1}}(u)=|H|^2$, a positive proportion of the set of incidences is supported on the set Ω of Möbius hyperbola, such that for $u\in\Omega$ we have

$$\forall u\in\Omega,\ \sum_{a\in A}\mathbb{1}_A(ua)\geqslant\Delta.$$

Henceforth we assume $\Delta\gg 1$, for otherwise

$$\sigma\ll|A|^{1/2}|H|,$$

which accounts for the first term in Estimate (5.4).

Applying Cauchy-Schwarz to the summation in u, restricted to Ω in (5.9), yields

$$\sigma^4 \ll |A|^2 E(H) \sum_{u\in\Omega} \left(\sum_{a\in A} \mathbb{1}_A(ua)\right)^2, \tag{5.12}$$

where $E(H)$ is the energy, defined by (5.2).

Using Formula (5.8), we conclude that if $\Delta \gg \frac{|A|^2}{p}$, that is unless

$$\sigma \ll \frac{|A|^{3/2}|H|}{p^{1/2}}, \tag{5.13}$$

one has (after dyadic summation in $k \geqslant \Delta$ in Formula (5.8))

$$\sum_{u\in\Omega} \left(\sum_{a\in A} \mathbb{1}_A(ua)\right)^2 \ll \min\left(\frac{|A|^7}{\Delta^3}, \frac{p|A|^4}{\Delta}\right).$$

It is not immediately obvious where this bound comes from so we shall go into detail here. We will reuse this argument later (for example in the proof of Theorem 5.10). As stated we will seek to bound the quantity on the left using dyadic summation and the bound on the number of k-rich hyperbola from (5.8) which states that the number of k-rich hyperbolae in H, m_k, is bounded as follows:

$$m_k \ll \min\left(\frac{|A|^7}{k^5}, \frac{p|A|^4}{k^3}\right). \tag{5.14}$$

Starting with the dyadic summation, we introduce some more notation in the fashion of Δ and Ω with $\Delta_k := 2^k\Delta$ and Ω_k the corresponding set of Δ_k-rich hyperbola. Precisely Ω_k is the subset of H such that for every $u \in \Omega_k$ we have that, upto a factor of two, u supports Δ_k points, that is, for $u \in \Omega_k$, $\sum_{a\in A} \mathbb{1}_A(u(a)) \sim \Delta_k$. Having split up our popular set Ω into dyadic pieces we can sum over each of these separately so we have

$$\sum_{u\in\Omega} \left(\sum_{a\in A} \mathbb{1}_A(ua)\right)^2 = \sum_k \sum_{u\in\Omega_k} \left(\sum_{a\in A} \mathbb{1}_A(ua)\right)^2.$$

Next we use the fact from the definition of Ω_k that for $u \in \Omega_k$, $\sum_{a\in A} \mathbb{1}_A(u(a)) \sim \Delta_k$ so

$$\sum_{u\in\Omega} \left(\sum_{a\in A} \mathbb{1}_A(ua)\right)^2 \sim \sum_k \sum_{u\in\Omega_k} \Delta_k \sum_{a\in A} \mathbb{1}_A(ua) = \sum_k \Delta_k \sum_{u\in\Omega_k} \sum_{a\in A} \mathbb{1}_A(ua). \tag{5.15}$$

We now consider what the sum $\sum_{u\in\Omega_k} \sum_{a\in A} \mathbb{1}_A(ua)$ means, we add a one every time that the point a lies on $u \in \Omega_k$, so this counts the number of intersections of points in A with the popular hyperbola in Ω_k. We also note that as every hyperbola in Ω_k supports (upto a factor of two) Δ_k points so we have that

$$|\Omega_k|\Delta_k \sim \sum_{u\in\Omega_k} \sum_{a\in A} \mathbb{1}_A(ua).$$

We now apply Bound (5.8) to $|\Omega_k|$ giving us

$$\sum_{u\in\Omega_k}\sum_{a\in A}\mathbb{1}_A(ua) \sim |\Omega_k|\Delta_k \ll \Delta_k \min\left(\frac{|A|^7}{\Delta_k^5}, \frac{p|A|^4}{\Delta_k^3}\right).$$

Plugging this into (5.15) gives

$$\sum_{u\in\Omega}\left(\sum_{a\in A}\mathbb{1}_A(ua)\right)^2 \ll \sum_k \Delta_k \min\left(\frac{|A|^7}{\Delta_k^4}, \frac{p|A|^4}{\Delta_k^2}\right) = \sum_k \min\left(\frac{|A|^7}{\Delta_k^3}, \frac{p|A|^4}{\Delta_k}\right).$$

We finish this section of the proof by summing over k and as $\Delta < \Delta_k$ we arrive at our goal of

$$\sum_{u\in\Omega}\left(\sum_{a\in A}\mathbb{1}_A(ua)\right)^2 \ll \min\left(\frac{|A|^7}{\Delta^3}, \frac{p|A|^4}{\Delta}\right). \tag{5.16}$$

Observe that from the definition of Δ the minimum being achieved on the second term means that $\Delta^2 < \frac{|A|^3}{p}$ and plugging in our Definition (5.11) for Δ we in fact have

$$\sigma \ll \frac{|H||A|^{5/4}}{p^{1/4}}, \tag{5.17}$$

which accounts for the corresponding additional term in the statement of Theorem 5.4.

Assuming that the minimum is instead achieved on the first term and applying the bound $E(H) \ll |H|^2 M_1$ from the forthcoming Lemma 5.1 (with the quantity M_1 having been defined in the pruning procedure at the outset) gives

$$\sigma^4 \ll |H|^2 M_1 \frac{|A|^9}{\Delta^3},$$

and so after once again plugging in our Definition (5.11) we have

$$\sigma^{10} \ll |A|^{12}|H|^8 M_1$$

and completes the proof of Estimate (5.4).

To address the real case in Estimate (5.6) in Theorem 5.4 we merely recalculate the quantity

$$\sum_{u\in\Omega}\left(\sum_{a\in A}\mathbb{1}_A(ua)\right)^2$$

using Theorem 5.1, and Estimate (5.6) follows. The details of this change are as follows, Theorem 5.1 tells us that

$$\sum_{u\in\Omega}\left(\sum_{a\in A}\mathbb{1}_A(ua)\right)^2 \lesssim \frac{|A|^4}{\Delta^3} + \frac{|A|^6}{\Delta^{11/2}}.$$

Note we now have thrown away log terms (although only in the second term) unlike in the finite field case and these are hidden in the $\lesssim$ symbol. Plugging this into Inequality (5.12), along with the same $E(H) \ll |H|^2 M_1$ from later in the thesis, we get

$$\sigma^4 \lesssim |H|^2 M_1 \left(\frac{|A|^6}{\Delta} + \frac{|A|^8}{\Delta^{7/2}}\right),$$

which after replacing the Δ's by their definition in (5.11) provides, if the first term dominates (note that the log term appears only in the second term)

$$\sigma^6 \ll |A|^7|H|^4M_1,$$

and if the second term dominates

$$\sigma^{11} \lesssim |A|^{23/2}|H|^9M_1.$$

Combining these two inequalities completes the proof of Estimate (5.6)

We now proceed towards proving Estimate (5.5), we do this by another application of Cauchy-Schwarz to the summation in A in (5.9): This yields

$$\sigma^4 \ll |A|^3 \sum_{u\in HH^{-1}HH^{-1}} r_{HH^{-1}HH^{-1}}(u) \sum_{a\in A} \mathbb{1}_A(ua).$$

As above, a positive proportion of the set of incidences must be supported on the set Ω of Möbius hyperbolae u, supporting at least

$$\Delta := \frac{\sigma^4}{3|A|^3|H|^4}$$

points of $A\times A$, thus redefining Δ. We proceed under assumption $\Delta \gg 1$, or else $\sigma \ll |A|^{3/4}|H|$.

Hence, again by Cauchy-Schwarz,

$$\sigma^8 \ll |A|^6T_4(H) \sum_{u\in\Omega} \left(\sum_{a\in A} \mathbb{1}_A(ua)\right)^2. \tag{5.18}$$

Applying (5.8) to estimate the incidence term we assume that the minimum is achieved on its first term, or else by definition of Δ one has

$$\sigma \ll (|A||H|)\left(\frac{|A|}{p}\right)^{1/8},$$

which enters the statement of Theorem 5.4 in $\mathbb{F}=\mathbb{F}_p$ case.

Hence,

$$\sum_{u\in\Omega} \left(\sum_{a\in A} \mathbb{1}_A(ua)\right) \ll \frac{|A|^{16}|H|^{12}}{\sigma^{12}}.$$

Furthermore, by Lemma 5.4, we have

$$T_4(H) \leqslant |H|^2T_3(H) \ll |H|^{6+1/3}$$

in fields of characteristic p, with $|H|<p$, and in the specific case $\mathbb{F}=\mathbb{F}_p$

$$T_4(H) \leqslant |H|^2T_3(H) \ll |H|^5M_2^2 + \begin{cases} \frac{|H|^7}{p}, & \text{if } |H|>p^{5/4} \\ p^{2/3}|H|^{5+2/3}, & \text{if } p\leqslant |H|\leqslant p^{5/4} \\ |H|^{6+1/3}, & \text{if } |H|<p \end{cases}.$$

Note that the quantity M_2, corresponding to the maximum number of translates in H, lying on a horizontal/vertical line has been redefined according to the pruning procedure at the outset of the proof.

Combining the last two estimates finishes the proof of Theorems 5.3 and 5.4.

□

5.4 Energy Bounds

In this section, we will state and prove our two energy bounds used in the previous section in the proof of Theorems 5.3 and 5.4. We will start by stating Lemma 5.1 and Lemma 5.2 before comparing them to Shkredov's Lemma 5.3, the analogue of Lemmas 5.1 and 5.2 in the proof of Theorem 5.2. Afterwards, we will provide the proofs which are taken from mine and Rudnev's paper [140]. We will then combine Lemma 5.2 with Theorem 5.7 to prove Lemma 5.4.

Our first lemma is a bound on the standard energy of the hyperbola we are interested in. As we will also deal in higher energies for our second lemma we may use both the standard notation $E(H)$ as well as the more general $T_2(H)$ from the notation of Section 2.5.3.

Lemma 5.1. *The energy of a set H of translates of the hyperbola $y = \frac{-1}{x}$ is bounded by*

$$E(H) \ll |H|^2 M,$$

where M is the maximum number of translates $(a,b) \in H$ having the same abscissa or ordinate.

Lemma 5.1 can be tight as we will show in an example after the proof although fits in our intuition for energies from Section 2.5.1 as our upper bound can vary from $|H|^2$ to $|H|^3$ depending on the arrangement of the hyperbola.

Our next lemma deals with the higher T_3 energy. Although as a bound this is better than Lemma 5.1 when dealing with very rich hyperbola (recalling that after the kth iteration we only get a nontrivial bound on the number of hyperbolae in H, which are $|A|^{1-2^{-k}}$-rich) it does show links to the Minkowski distances detailed in Section 5.2. Whilst we could use even higher energies, calculating them proves increasingly difficult as the number of variables growths quickly and such result would only improve in the case of increasingly rich hyperbolae.

Lemma 5.2. *Let H be a set of translates of the hyperbola $y = \frac{-1}{x}$. Then*

$$T_3(H) := \sum_x r^2_{HH^{-1}H}(x) \leqslant 2|H|Q(H) + 2|H|^4,$$

where

$$Q(H) = |\{(h_1,h_2,h_1',h_2') \in H^4 : D(h_1,h_1') = D(h_2,h_2')\}|,$$

with

$$D(h,h') = D((a,b),(a',b')') := (a-a')(b-b').$$

The quantity $D(h,h')$, with $h,h' \in H$ becomes the Minkowski distance D_m, defined in Section 5.2 (in particular Definition 5.2) after rotating H by 45°; in [127] quadruples in $Q(H)$ were referred to as *rectangular quadruples*, we also use this term.

These lemmas are the analogue of the following lemma due to Shkredov and can be seen as a generalisation from H being a Cartesian product to the general H. We have split the lemmas into two with Lemma 5.1 being the counterpart to Bound (5.19) and Lemma 5.2 being the counterpart to Bound (5.20). We will use Lemma 5.2 to prove the forthcoming Lemma 5.4 which is what we use in the above proof of Theorem 5.3.

Lemma 5.3. *[157, Lemma 14] For $H = B \times B$, a set of translates of the hyperbola $y = \frac{-1}{x}$, the following estimates hold.*

$$E(H) \leqslant |B|^2 E^+(B), \tag{5.19}$$

and

$$T_3(H) \leqslant |B|^2 \sum_x r^2_{(B-B)(B-B)}(x) + |B|^8. \tag{5.20}$$

In positive characteristic, for $|B| < p^{1/2}$, then

$$\sum_x r^2_{(B-B)(B-B)}(x) \lesssim |B|^5 (E^+(B))^{1/2}, \tag{5.21}$$

if $\mathbb{F} = \mathbb{F}_p$ the constraint $|B| < p^{1/2}$, can be removed by adding the extra term $|B|^8/p$ to the right-hand side of the latter estimate.

In comparison Lemma 5.3 are a specific case of our two Lemmas 5.1 and 5.2. In the Cartesian product case $|H|^2 M = |B|^5 \geqslant |B|^2 E^+(B)$ and Relation (5.20) is the specific case of Lemma 5.2 as a Cartesian product. We note that Shkredov's sum-product Bound (5.21) is stronger than the general result we prove below. The proofs are again from [140].

5.4.1 T_2 Hyperbola Energy Bound

Proof of Lemma 5.1. The proof is merely mimicking the corresponding part of the proof of Shkredov's Lemma 5.3.

We represent the hyperbola $y = a + \frac{1}{b-x}$ by the SL_2 matrix $h = \begin{pmatrix} -a & ab+1 \\ -1 & b \end{pmatrix}$. Without loss of generality, we may assume that none of the a or b are ever zero.

We have, with $w_1 = b_1 - b_2$,

$$h_1 h_2^{-1} = \begin{pmatrix} 1 + a_1 w_1 & a_1 - a_2 - a_1 a_2 w_1 \\ w_1 & 1 - a_2 w_1 \end{pmatrix}.$$

Hence, $E(H)$ can be seen to equal the number of solutions to the following set of equations (the second part by instead considering $h_1^{-1} h_2$ as $h_1 h_2^{-1} = h_1' h_2'^{-1}$ is equivalent to $h_2^{-1} h_2' = h_1^{-1} h_1'$)

gained from comparing the four entries in the matrix, we lose one such as is is dependant on the other three. Briefly, this can be done by noting that by the bottom left $w_1 = w_1'$ and thus the second line in (5.22), which when plugged into the top left and bottom right gives the remaining equations and the top right is just a combination of these.

$$\begin{array}{ccc} a_1 = a_1', & & b_1 = b_1', \\ b_1 - b_2 = b_1' - b_2', & \text{or} & a_1 - a_2 = a_1' - a_2', \\ a_2 = a_2' & & b_2 = b_2'. \end{array} \tag{5.22}$$

This completes the proof. □

You can interpret the system of Equations (5.22) as the number of parallelograms (with one pair of edges parallel to an axis) in the set H as seen in Figure 5.3 where we take the top line as $y = b_1 = b_1'$ and the bottom $y = b_2 = b_2'$. This makes it easy to see the following example is tight to the bound in Lemma 5.1. Consider first the case covered by Shkredov where the set of translates is a Cartesian product, $H = B \times B$. If B is an arithmetic progression then the number of parallelograms is exactly $|B|^5$ which is equal to both the $|H|^2 M$ of Lemma 5.1 and the $|B|^2 E^+(B)$ of Lemma 5.3. When B is not an arithmetic progression Lemma 5.1 does not capture the behaviour as well. On the other hand, we can consider the above example but where each row is translated so we no longer have a Cartesian product, but still equal spacing, then Lemma 5.1 is again tight and we cannot apply Lemma 5.3.

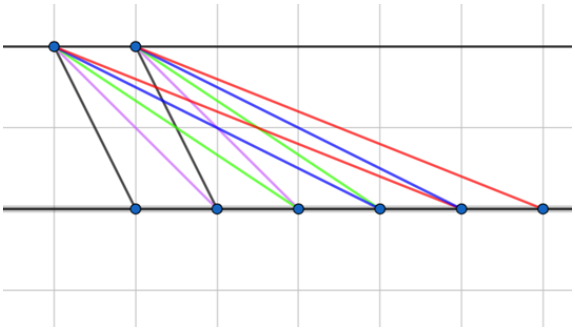


Figure 5.3: Illustrating where the bound in Lemma 5.1 arises from.

5.4.2 T_3 Hyperbola Energy Bound

Proof of Lemma 5.2. We will reuse the matrix notation from the previous proof. Note that there is always a trivial bound (recall Equation (2.6)) of $T_3(H) \leqslant |H|^2 E(H)$, thus from Lemma 5.1 $T_3(H) \leqslant |H|^4 M$.

The following argument is somewhat more involved than Shkredov's proof of (5.20), where the Cartesian product scenario enables one to easily switch between various h's appearing in the T_3 quantity.

However, the claim one ends up with is in the same spirit: in both estimates for $T_3(H)$ the first term pertains to the rectangular quadruple count, while the second term estimates separately the contribution coming from the Borel subgroup of SL_2.

To start we prove the following bound to estimate

$$X_B := \max_{g \notin B} \sum_{x \in gB} r^2_{HH^{-1}}(x) \leqslant |H|^2, \tag{5.23}$$

where B is the Borel subgroup of upper-triangular matrices. This, along with another awkward case considered below, will contribute the $|H|^4$ term.

We start to prove Estimate (5.23) by noting a left coset of B, which is not B itself, is defined by a matrix $g = \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}$, with $c \neq 0$.

$$gB = \begin{pmatrix} s & t \\ cs & ct + s^{-1} \end{pmatrix}, \qquad (s,t) \in \mathbb{F}_p^2, s \neq 0.$$

Suppose $h_1 h_2^{-1} = h_1' h_2'^{-1} \in gB$. returning to our notation from earlier that of $w_1 = b_1 - b_2$,

$$h_1 h_2^{-1} = \begin{pmatrix} 1 + a_1 w_1 & a_1 - a_2 - a_1 a_2 w_1 \\ w_1 & 1 - a_2 w_1 \end{pmatrix}.$$

This means, since $c \neq 0$ that $w_1 = w_1' \neq 0$, and therefore, by equating the diagonal entries in $g_1 g_2^{-1}$, that $a_1 = a_1'$ and $a_2 = a_2'$.

Furthermore, since $c(1 + a_1 w_1) = w_1$, we can see we cannot have $ca_1 = 1$ else we reach a contradiction as we have that $w_1 = ca_1 w_1 + c = w_1 + c$ and c is both non zero and zero. So returning to the case $ca_1 \neq 1$, given $h_1 = (a_1, b_1)$ and $h_2' = (a_2', b_2')$, we know $w_1 = w_1'$ and we can determine w_1 and thus also w_1' using $c(1 + a_1 w_1) = w_1$. Once we know these we can also calculate b_1' and b_2, as well as a_2, a_1'. This accounts for the claim in Estimate (5.23) as everything is determined after two independent choices from H.

We are now ready to finish the proof of the claim of Lemma 5.2. We start by splitting into the two cases where x is either in the Borel subgroup where we can use the above and the case where it is not. That is we partition (and give names to the two parts) the following so

$$T_3 = \sum_{x \in B} r^2_{HH^{-1}H}(x) + \sum_{x \notin B} r^2_{HH^{-1}H}(x) := Y_B + \overline{Y}_B,$$

where Y_B is the part of T_3, corresponding to $x \in B$, and $\overline{Y}_B$ the complement.

It follows from (5.23) that

$$Y_B \leqslant |H|^4.$$

This is because we can write

$$Y_B = \sum_{x \in B} \left(\sum_{h_3 \in H} r_{HH^{-1}h_3}(x) \right)^2 \leqslant |H| \sum_{x \in B, h_3 \in H} r^2_{HH^{-1}h_3}(x).$$

Observe that for each value of h_3, one has

$$r^2_{HH^{-1}h_3}(x) = \left|\{(h_1,h_2,h_1',h_2') \in H^4 : h_1h_2^{-1} = h_1'h_2'^{-1} \in h_3^{-1}B\}\right|.$$

Now apply Estimate (5.23), for each h_3, also observe that $h_3 \notin B$. That is

$$Y_B \leqslant |H| \sum_{x\in B, h_3\in H} r^2_{HH^{-1}h_3}(x) \leqslant |H| \sum_{h_3\in H} |H|^2 = |H|^4.$$

The Borel case considered above has contributed $|H|^4$ to the bound of the lemma. It remains to estimate the quantity $\overline{Y}_B$ which will contribute the first term in the lemma.

By Cauchy-Schwarz,

$$\sum_{x\notin B} r^2_{HH^{-1}H}(x) = \sum_{x\notin B}\left(\sum_{h_2\in H} r_{Hh_2^{-1}H}(x)\right)^2 \leqslant |H| \sum_{x\notin B, h_2\in H} r^2_{Hh_2^{-1}H}(x).$$

The product appearing in T_3 equals

$$h_1h_2^{-1}h_3 = \begin{pmatrix} -a_1(w_1w_2+1)-w_2 & 1+a_1w_1+b_3(w_2+a_1(1+w_1w_2)) \\ -(1+w_1w_2) & w_1+b_3(1+w_1w_2) \end{pmatrix}, \tag{5.24}$$

with an extra notation $w_2 = a_3 - a_2$. This is the generalisation of (5.22) in the previous lemma, although our Cauchy-Schwarz has allowed us to deal with five rather than six matrices as we have moved to the case where $h_2' = h_2$ rather than treat them separately. In addition, we have $1+w_1w_2 \neq 0$. And we have $h_1h_2^{-1}h_3 = h_1'h_2^{-1}h_3'$, with the corresponding notations $w_1' = b_1' - b_2$, $w_2' = a_3' - a_2$.

It follows that

$$c = w_1w_2 = w_1'w_2', \qquad (a_1'-a_1)c = w_2 - w_2' = (a_3-a_3'), \qquad (b_3-b_3')c = b_1'-b_1. \tag{5.25}$$

Since $c \neq 0$, this implies

$$(a_1-a_1')(b_1-b_1') = (a_3-a_3')(b_3-b_3').$$

This is a rectangular quadruple, with (a_2,b_2) having been eliminated.

It remains to show that given a nontrivial rectangular quadruple, there is only at most two (a_2,b_2), corresponding to it. Of course, if the quadruple is trivial, that is $h_1 = h_1'$, $h_3 = h_3'$, then there are $|H|$ choices for h_2.

Suppose, we have a fixed nontrivial quadruple (h_1,h_3,h_1',h_3'), which means from Equations (5.25) we know c. Thus (a_2,b_2) is on the intersection of H with the hyperbola $(a_3-x)(b_1-y) = c$, as well as the hyperbola $(a_3'-x)(b_1'-y) = c$. The intersection is at most two points, unless this is the same hyperbola, namely $a_3 = a_3'$, $b_1 = b_1'$.

Furthermore, from equalising the top right entries of $h_1h_2^{-1}h_3 = h_1'h_2^{-1}h_3'$, we have $b_3w_2 - b_3'w_2' = s$, where the right-hand side s is known from the quadruple. Therefore, we can determine a_2, and hence have at most two h_2, unless in addition to already having $a_3 = a_3'$, $b_1 = b_1'$ we have $b_3 = b_3'$.

But then we have $h_3 = h_3'$, and therefore $h_1 = h_1'$, so are in the trivial quadruple case. This adds another $|H|^4$ to the bound of the lemma and completes the proof. □

5.4.2.1 Minkowski Distance L^2 Bound

Next, we wish to deal with the $Q(H)$ quantity from Lemma 5.2, to do this we will make use of a result due to Murphy et al. in [110]. Explicitly we wish to prove the following lemma which is ll that remained to complete the proof of Estimate (5.5) as well as the $\mathbb{F}_p$ claims of Theorem 5.4.

Lemma 5.4. *For a set H of translates of the hyperbola $y = \frac{-1}{x}$ in $\mathbb{F}_p^2$, such that at most M translates $(a,b) \in H$ have the same abscissa or ordinate, one has*

$$T_3(H) \ll |H|^3 M^2 + \begin{cases} \frac{|H|^5}{p}, & \text{if } |H| > p^{5/4} \\ p^{2/3}|H|^{3+2/3}, & \text{if } p \leqslant |H| \leqslant p^{5/4} \\ |H|^{4+1/3}, & \text{if } |H| < p \end{cases}.$$

The bound

$$T_3(H) \ll |H|^3 M^2 + |H|^{4+1/3}.$$

holds over a general $\mathbb{F}$ of characteristic p, provided that $|H| < p$.

To this end, recall that by Lemma 5.2 we have the following bound

$$T_3(H) \lesssim |H|Q(H) + |H|^4, \tag{5.26}$$

where $Q(H)$ is the number of rectangular quadruples in H, that is

$$Q(H) = \left|\{h_1 = (a_1,b_1), h_2 = (a_2,b_2), h_1' = (a_1',b_1'), h_2' = (a_2',b_2'), : (a_1-a_1')(b_1-b_1') = (a_2-a_2')(b_2-b_2')\}\right|.$$

To continue we will use a change of variables so we can express each rectangular quadruple as an energy type of quantity where 2 pairs of points are the same Minkowski distance apart. Geometrically this is moving from equating the area of the rectangle formed by two points being opposite corners with edges parallel to the axis to equating the Minkowski distance between them instead. We use the change of variables $(a,b) \to \left(\frac{a+b}{2}, \frac{a-b}{2}\right)$, with H_m replacing H in the new variables. Now, with $h_1 = (a_1,b_1), \ldots, h_2' = (a_2',b_2')$ in H_m, $Q(H)$ equals the number of solutions of $(a_1-a_1')^2 - (b_1-b_1')^2 = (a_2-a_2')^2 - (b_2-b_2')^2$, these are the Minkowski distances between h_1 and h_1', and h_2 and h_2', respectively. That is we can instead consider $Q(H)$ as

$$Q(H) = \left|\{h_1, h_2, h_1', h_2' \text{ where } h_i = (a_i,b_i) \in H_m : D_M(h_1,h_1') = D_M(h_2,h_2')\}\right|.$$

Clearly, if -1 is a square in $\mathbb{F}$, one is free to change $-$ to $+$ in the quadratic form, becoming the analogue of the Euclidean distance $\|\cdot\|$ in $\mathbb{F}^2$. We will now state the theorem of Murphy et al. [110] that we will use to bound this quantity.

Theorem 5.7 (Theorem 4 [110])**.** *Let $H_m \subseteq \mathbb{F}_p^2$. Set*

$$Q^*(H_m) = |\{(h_1,h_2,h_1',h_2') \in H_m^4 : \|h_1 - h_1'\| = \|h_2 - h_2'\| \neq 0\}|.$$

Then

$$Q^*(H_m) \ll \begin{cases} \frac{|H_m|^4}{p}, & \text{if } |H_m| > p^{5/4} \\ p^{2/3}|H_m|^{8/3}, & \text{if } p \leqslant |H_m| \leqslant p^{5/4} \\ |H_m|^{10/3}, & \text{if } |H_m| < p \end{cases}.$$

Moreover, $Q^*(H_m,) \ll |H_m|^{10/3}$, for $|H_m| < p$ in any field of characteristic p.

Theorem 5.7 allows for -1 being a square in $\mathbb{F}$ and so we can apply it in the Minkowski distance setting. Also, note that if we are in the case $|H_m| < p$ we can always pass to an extension of $\mathbb{F}$. We will not consider this Theorems proof, however, we will consider some of the tools required to adapt it to the Minkowski distance in greater depth in Appendix A, in particular, we consider the group of Minkowski isometries (comparing these to the Euclidean group of corresponding isometries) and the Blaschke-Grünwald Kinematic Mapping which allows us to move from sets of these isometries to projective space and apply various incidence theorems.

To bound the quantity T_3 we will just need to add to the above bound on $Q^*(H_m)$ the count of rectangular quadruples in H, contributed by the case $D(h_1,h_1') = D(h_2,h_2') = 0$, the isotropic distances. If M is the maximum number of points in H on a horizontal or vertical line, their number is trivially at most $M^2|H|^2$. After multiplying by $|H|$ according to (5.26), this term will dominate the $|H|^4$ term.

We have therefore established Lemma 5.4 and consequently the last part in the proof of Theorem 5.3.

5.5 Generalisations

This penultimate section will cover recent joint work with Audie Warren from [197]. At its heart, this work generalises Theorem 5.5 to an incidence theorem between arbitrary sets of Möbius transformations and point sets in $\mathbb{F}_p^2$ (rather than for just Möbius hyperbola and a Cartesian product for the point set). We also note some asymmetric incidence results of similar flavour before using these generalisations to gain applications including a Beck's theorem (Theorem 3.2) style result for these Möbius transformations, an expander result, and a bound on the number of representations of a (non zero) number $\lambda \in \mathbb{F}_p^*$ as the product of two members of a set $A \subset \mathbb{F}_p^*$ which we shall deal with in the last section. We also cover a few other Corollaries. This Section will be split into two main parts, first stating and proving Theorem 5.8 (the generalisation of Theorem 5.5), the second dealing with the asymmetric results.

The Möbius transformations are of the form

$$f(x) = \frac{ax+b}{cx+d}, \qquad ad-bc \neq 0,$$

with $a,b,c,d \in \mathbb{F}_p$. We equate each such f with a matrix

$$M_f = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

We see these transformations as projective transforms, that is $f : \mathbb{P}(\mathbb{F}_p) \to \mathbb{P}(\mathbb{F}_p)$. As we are dealing with projective points we also need to know that $f([\frac{-d}{c};1]) = [1,0]$ and $f([1;0]) = [\frac{a}{c};1]$. Also note scalar multiples of these matrices correspond to the same transform which is to be expected and part of why we are treating these matrices projectively.

As one would hope the matrix corresponding to the composition of two functions is the product of the matrices of the individual functions. That is

$$M_{f\circ g} = M_f M_g.$$

A fair question to ask is why are we interested in these Möbius transforms? As a sort of dual question as to what is special about these transforms is can these methods be applied to other curves such as circles and other conics or even general curves? A tongue in cheek answer to the first question is because our method works for Möbius transforms. A more nuanced answer also includes that this is hopefully also a step towards further generalisations, that incidence results over finite fields for any curves are rare ([10], [157] as well as both [140] and [197] (detailed here) are some of the only such incidence results) and thus any are interesting for that reason alone. Another reason of interest, connecting back to the sum-product phenomenon and in particular growth in matrix groups (covered in detail in Section 4), is that we wish to further understand SL_2, Möbius transforms are a special case of these as the projective linear group. Assuming our field isn't too small (that is for $\mathbb{F}_q$ we have that $q \neq 2$ or 3) the special linear group is a perfect central extension of this simple group of transforms. We also note that these Möbius transforms include the affine transforms and so affine group which is studied in [121] and are a three dimensional family like the 2×2 upper triangular matrices of Theorems 4.6 and 4.9.

Moving instead to the related question of can we generalise or adapt these methods to other curves we first consider the key properties of Möbius transforms we use.

- First we use that they are composable, this is used both if you wish to deal with energies (such as the more restricted Theorem 5.11) and in particular whenever you would wish to apply bounds such as Lemma 5.1 or 5.2. We also use this in our intermediate theorem (Theorem 5.5 and the upcoming 5.8) when mapping our transforms to lines (as will be defined in Equation (5.27)) as this is a composition of two Möbius transforms along with the one being transformed. This also makes use of the fact that lines (or translations) are a subgroup of the group of Möbius transforms.

- A second useful property is the existence of a map to matrices for the transforms, this allows compositions to be reduced to calculations.

- We also make use of the fact that three (or possibly a finite number of points) define such a curve. This allows us to relate the transforms which go through a specially chosen point to lines. We would also have such a property with some other conics and as an example, circles are defined by three points however it is currently unclear to me how to (if indeed it is even possible) how to find such a map taking those circles which pass through a point to lines. As discussed in the next point, such a map would also be further hindered by the lack of injectivity.

- We use injectivity when removing our special point as we are actually removing the horizontal and vertical lines through the point. For an injective function removing the point and the bad lines are equivalent as the point is the only point on either of the lines.

Some of these may not be completely required and other methods exist for generalisations. Two potential routes include considering large collections of conics in $\mathbb{F}_p^2$ with two points in common and make use of projective transforms to map these to Möbius transforms. A second is circles,

correspondence with Thang and Mohammadi suggests that this should work as after translation (which preserves incidences) we can consider the set of circles which all pass through the origin and identify them with their centres. they then have equation $(x-a)^2+(y-b)^2=r$ which can be rewritten as $-2xa-2yb+x^2+y^2=0$ once you note we pass through the origin and this can be viewed as a line.

5.5.1 A Generalisation of Theorem 5.5

As stated above this subsection will cover the statement and proof of the following Theorem 5.8. This result is a generalisation of Theorem 5.5, which at a slight cost to the exponents, is no longer restricted to a pointset which is a Cartesian product and also now deals with general Möbius transformations rather than just the hyperbola needed for the proofs of Theorems 5.3 and 5.4.

Theorem 5.8. *For any set T of Möbius transformations, and any set of points $P \subseteq \mathbb{F}_p^2$ with $|P| \leq p^{15/13}$, we have*

$$\mathcal{I}(P,T) \ll |P|^{15/19}|T|^{15/19} + |P|^{23/19}|T|^{4/19} + |T|.$$

Furthermore, given any set P of points with $|P| \leq p^{15/26}$ and some integer $k \geq 3$, the set T_k of k-rich transformations satisfies

$$|T_k| \ll \frac{|P|^{15/4}}{k^{19/4}} + \frac{|P|^2}{k^2}.$$

Before we turn to the proof, I will comment on the result itself. In the balanced case $|P| = N = |T|$ our bound is $\ll N^{30/19}$ and in particular $30/19 > 3/2$ which is what our intuition says we should be aiming to beat as the exponent from Cauchy-Schwartz for say, points and lines, however as such transforms are uniquely defined by three rather than two points the trivial bound in this case is instead due to Hölder's theorem (Theorem 0.2) and is $N^{5/3}$ which we do improve upon and thus we are non-trivial in the balanced case. As our exponent is larger than 3/2 we must be careful when seeking applications that we remain non-trivial and there is not a Cauchy-Schwartz bound for the application.

As a generalisation of Theorem 5.5, the proof of Theorem 5.8 follows several major themes as the proof of Theorem 5.5. That is like the previous proof our outline will follow fixing a special point and considering the set of transforms that fix this point, applying a map so these transforms become lines and suitable mapping the points in the same way so we can make use of a point-line incidence result and finishing by summing over all our special points. Instead of Lemma 3.2, which works for a Cartesian point set we instead use Corollary 3.1.

Proof. We start by choosing our special point $q = (q_1, q_2) \in P$ and defining the subsets $T_{q,k} \subseteq T_q \subseteq T$ of our set of transformations. We define T_q as the transformations in T which pass through our chosen point q and $T_{q,k}$ as the k-rich transformation which also pass through q. The proof shall proceed by manipulating the transforms so the size of $T_{q,k}$ can be bounded using Corollary 3.1.

The Möbius transformations are of the form

$$f(x) = \frac{ax+b}{cx+d}, \qquad ad-bc \neq 0,$$

with $a,b,c,d \in \mathbb{F}_p$ as noted above. In particular note that a transform $f \in T_q$ satisfies $f(q_1) = q_2$ as T_q is the set of those transforms passing through q.

As our aim is to map these transforms to lines we consider two cases. First if $c = 0$, then such transforms are lines of the form $f(x) = \frac{a}{d}x + \frac{b}{d}$ and we shall deal with the incidences these contribute separately at the end (becoming the second terms in each of the bounds of this theorem we are proving). In the second case $c \neq 0$ and for convenience we may scale so that $c = 1$ and calculate the specific form that our 'b' takes. To do this consider

$$q_2 = \frac{aq_1 + b}{q_1 + d} \implies b = q_2(q_1 + d) - aq_1.$$

So that means in the matrix setting we have

$$M_f = \begin{pmatrix} a & q_2(q_1+d) - aq_1 \\ 1 & d \end{pmatrix}.$$

Next we define our map which takes the above transforms to lines. To that end define (reminiscent of the maps used to the proof of Theorem 5.5)

$$g_1(x) = \frac{1}{q_2 - x}, \qquad g_2(x) = q_1 - \frac{1}{x}$$

and use these to define our map

$$\phi_q(f) := g_1 \circ f \circ g_2. \tag{5.27}$$

Next we check that this map does indeed send our f's to lines. to see that $\phi_q(f)$ is indeed a line consider the calculation

$$M_{g_1} M_f M_{g_2} = \begin{pmatrix} 0 & 1 \\ -1 & q_2 \end{pmatrix} \begin{pmatrix} a & q_2(q_1+d) - aq_1 \\ 1 & d \end{pmatrix} \begin{pmatrix} q_1 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} q_1 + d & -1 \\ 0 & a - q_2 \end{pmatrix}.$$

We also note that as the determinants of M_{g_1} and M_{g_2} are both one we have that $\det(M_{g_1} M_f M_{g_2}) = \det(M_f) \neq 0$. Now to see that $M_{g_1} M_f M_{g_2}$ is a line you can either note that the bottom left entry is one or write it out as the following function which is in the canonical form of a line

$$y = \frac{(q_1 + d)x - 1}{a - q_2} = \frac{(q_1 + d)}{a - q_2}x + \frac{-1}{a - q_2}.$$

Before moving on we will prove a couple properties of this ϕ_q, first that it is injective, and secondly this map does not include f's with $a = q_2$, in its domain. Starting with the second this is because such matrices would have determinate 0 (that is the determinate of $M_{g_1} M_f M_{g_2}$ is equal to that of M_f and is calculated by $(q_1 + d)(a - q_2) - 0$), equally this shows that such f's wouldn't be Möbius transform either. That our ϕ_q is injective comes from calculation, indeed assume with have two transforms f and f' with

$$f(x) = \frac{ax + (q_2(q_1+d) - aq_1)}{x + d}, \quad f'(x) = \frac{a'x + (q_2(q_1+d') - a'q_1)}{x + d'},$$

such that $\phi_q(f) = \phi_q(f')$. If we first compare the y-intercepts then

$$\frac{-1}{a - q_2} = \frac{-1}{a' - q_2},$$

and thus $a = a'$. next comparing the slope of each line (noting we have $a = a'$ already and make this substitution)

$$\frac{(q_1 + d)}{a - q_2} = \frac{(q_1 + d')}{a - q_2},$$

so $d = d'$ and thus $f = f'$.

We have now mapped our transforms to lines so next must alter our points so the number of incidences is preserved. We will have to deal with a couple of problematic lines separately (in particular points on the lines $x = q_1$ and $y = q_2$). We define a new map ψ which alters a point $p = (p_1, p_2) \in P$ as follows

$$\psi(p_1, p_2) = \left(g_2^{-1}(p_1), g_1(p_2)\right) = \left(\frac{1}{q_1 - p_1}, \frac{1}{q_2 - p_2}\right).$$

It is clear now why we deal with the problematic lines separately. We call the image of ψ P', that is

$$P' := \psi(P) = \left\{\left(\frac{1}{q_1 - p_1}, \frac{1}{q_2 - p_2}\right) : (p_1, p_2) \in P\right\}.$$

As when we introduced ϕ we now turn to some required properties of ψ, namely we want that a point $p \in P$ intersects with a curve f[4] if and only if $\psi(p)$ intersects with the line $\phi(f)$. This is done by calculation

$$p \in f \iff p_2 = f(p_1) \iff g_1(p_2) = g_1(f(p_1)) \iff g_1(p_2) = g_1(f(g_2(g_2^{-1}(p_1)))) \iff \psi(p) \in \phi(f).$$

Note that as f is a bijection the restriction that $p \neq q$ the only point on the bad lines $x = q_1$ and $y = q_2$ that a non-horizontal and non-vertical line can pass is q. nor can we have f being a horizontal or vertical line as in these cases $d = -q_1$ and $a = q_2$ respectively. If we have either of these then the determinate of f is zero which we cannot have, therefore we don't have to deal with horizontal or vertical lines. As we do not count the point corresponding to q note that a k-rich transform with respect to P is mapped to a $(k-1)$-rich line with respect to P'. We can now bound $T_{q,k}$ using Corollary 3.1 which gives us

$$|T_{q,k}| \ll \frac{|P'|^{11/4}}{(k-1)^{15/4}} + \frac{|P'|}{k-1} \ll \frac{|P|^{11/4}}{k^{15/4}} + \frac{|P|}{k}.$$

Furthermore, the transformations in T_q with $c = 0$, which are already lines, must be concurrent through q, and the number of such k-rich lines is at most $\frac{|P|}{k}$ which we absorb into the above bound.

We now wish to move from a bound for $T_{q,k}$ into a bound on the set T_k of k-rich transformations in T. We do this by summing over all $q \in P$ noting that each k-rich transform is then counted k times. This leaves us with the following bound

$$|T_k| \leq \frac{1}{k} \sum_{q \in P} |T_{q,k}| \ll \frac{|P|^{15/4}}{k^{19/4}} + \frac{|P|^2}{k^2}.$$

[4] excluding the problem lines which we shall deal with separately.

Our final step in finishing this proof is to use a dyadic decomposition in terms of a Δ to choose later in a standard manner. Let $T_{=k}$ denote the set of exactly k-rich transformations in T. Then the following calculation gives us a bound for the incidences between points in P and our set T of transforms.

$$\begin{aligned}
\mathcal{I}(P,T) &= \sum_{k=1}^{|P|} k|T_{=k}| \\
&= \sum_{k<\Delta} k|T_{=k}| + \sum_{k=\Delta}^{|P|} k|T_{=k}| \\
&\ll \Delta|T| + \sum_{i=1}^{\log|P|} \sum_{\substack{f\in T: \\ 2^i\Delta\leq|f\cap P|<2^{i+1}\Delta}} (2^{i+1}\Delta) \\
&\leq \Delta|T| + \sum_{i=1}^{\log|P|} |T_{2^i\Delta}|(2^{i+1}\Delta) \\
&\ll \Delta|T| + \sum_{i=1}^{\log|P|} \left(\frac{|P|^{15/4}}{(2^{i+1}\Delta)^{19/4}} + \frac{|P|^2}{(2^{i+1}\Delta)^2}\right)(2^{i+1}\Delta) \\
&\ll \Delta|T| + \frac{|P|^{15/4}}{\Delta^{15/14}} + \frac{|P|^2}{\Delta}
\end{aligned}$$

What remains is to choose a Δ in such a way as to optimise the first two terms. This is achieved when

$$\Delta|T| \sim \frac{|P|^{15/4}}{\Delta^{15/14}} \implies \Delta \sim \frac{|P|^{15/19}}{|T|^{4/19}}.$$

We have a slight issue though in that our bound we used for T_k requires $k \geqslant 3$ and so Δ must also be larger than three. as such take

$$\Delta = \max\left\{3, \frac{|P|^{15/19}}{|T|^{4/19}}\right\}.$$

Considering the two cases, when $\Delta = 3$, we must have

$$\frac{|P|^{15/19}}{|T|^{4/19}} \leq 3 \implies |P|^{15/4} \ll |T|,$$

and thus the bound above gives

$$\mathcal{I}(P,T) \ll |T| + |P|^{15/19} + |P|^2 \ll |T|.$$

Otherwise $\Delta = \frac{|P|^{15/19}}{|T|^{4/19}}$ and our bound becomes

$$\mathcal{I}(P,T) \ll |P|^{15/19}|T|^{15/19} + |P|^{23/19}|T|^{4/19}.$$

To conclude we combine the bounds from the above two cases to get

$$\mathcal{I}(P,T) \ll |P|^{15/19}|T|^{15/19} + |P|^{23/19}|T|^{4/19} + |T|.$$

□

This proof and that of Theorem 5.5 follow very similar methods with the main difference being which bound on the number of k-rich lines are used. This proof however also goes into greater detail showing that the transformation from hyperbola to lines both do as we say and preserve the incidences as claimed. Afterwards, the summation to ensure all such incidences are collected follow broadly the same aims but again this latter proof goes into more detail.

One of the more intriguing parts of this proof and perhaps an idea that can find further uses in other problems is this mapping of a restricted set of the k-rich transforms to $(k-1)$-rich lines whilst preserving the incidences. This is a key part of the proof and also a barrier for generalising the argument as such maps are not obvious to find or if they even exist for other curves. That by fixing one of the points on a k-rich transform leaves a $(k-1)$-rich object which is defined now by two (as the transforms are defined by three points originally and so by choosing our q as one of these it is defined by two other points) and lines are defined by two points gives a moral sort of argument as to why we may expect such a map to exist. It may be that should we find ourselves interested in objects defined by four points we could move to these Möbius transforms, or by picking a special pair of points we could move straight to lines. This also (so that incidences are preserved) uses that Möbius transform are injective, this means that when we are discounting the bad lines, that is the horizontal and vertical line through our special point q, the only points on either of these lines are the special one q.

5.5.2 Asymmetric Incidence Results

In this subsection, we consider a couple of asymmetric incidence results. The first, Theorem 5.9, is proven using an argument of Rudnev and Shkredov from [137] combined with Theorem 5.8 to prove a non-linear analogue of their Theorem 8. The second pushes further, using our energy bounds from Section 5.4 to produce Theorem 5.10, an asymmetric version of Theorem 5.3.

We shall state and then prove each result now.

Theorem 5.9. *Let $A \times B$ be a set of points in $\mathbb{F}_p^2$, and let T be any set of Möbius transformations. Then if $|A||T| \ll p^2$, we have*

$$\mathcal{I}(A \times B, T) \ll |A|^{4/5}|B|^{3/5}|T|^{4/5} + |A|^{6/5}|B|^{7/5}|T|^{1/5} + |T|$$

Furthermore, given any set $A \times B$ of points with $|A|^3|B|^2 \leq p^2$ and some integer $k \geq 3$, the set T_k of k-rich transformations satisfies

$$|T_k| \ll \frac{|A|^4|B|^3}{k^5} + \frac{|A|^2|B|^2}{k^2}.$$

Proof. We reuse the first half of the proof of Theorem 5.8 up until the point we are about to apply Corollary 3.1. That is we have shown that the k-rich transformations (with respect to P) have been mapped to $k-1$ rich lines with respect to $P \setminus \{q\}$ and so we apply Corollary 3.2. Without loss of generality we may assume $|A| \leqslant |B|$, else relabel A and B.

$$|T_{q,k}| \ll \frac{(|A|-1)^3(|B|-1)^2}{(k-1)^4} \ll \frac{|A|^3|B|^2}{k^4} + \frac{|A||B|}{k}.$$

Following the same logic as in the proof of Theorem 5.8 we move to a bound on k-rich transforms as follows

$$|T_k| \ll \frac{1}{k}\sum_{q\in P}|T_{q,k}| \ll \frac{|A|^4|B|^3}{k^5} + \frac{|A|^2|B|^2}{k^2}.$$

Plugging this into the same set of calculations as used in the proof of Theorem 5.8 we instead get an incidence bound,

$$\mathcal{I}(A\times B, T) \ll \Delta|T| + \frac{|A|^4|B|^3}{\Delta^4} + \frac{|A|^2|B|^2}{\Delta}.$$

As before our next step is to choose what our Δ should be by optimising between the first two terms, This gives that we want

$$\Delta|T| \sim \frac{|A|^4|B|^3}{\Delta^4},$$

$$\Delta \sim \frac{|A|^{4/5}|B|^{3/5}}{|T|^{1/5}}.$$

As our bound on the number of k-rich transforms only applies for $k \geqslant 2$ we choose

$$\Delta := \max\left\{2, \frac{|A|^{4/5}|B|^{3/5}}{|T|^{1/5}}\right\}.$$

Plugging in the two options again gives us two bounds we will later combine. Starting with the second we get the bound

$$\mathcal{I}(A\times B, T) \ll |A|^{4/5}|B|^{3/5}|T|^{4/5} + |A|^{4/5}|B|^{3/5}|T|^{4/5} + |A|^{6/5}|B|^{7/5}|T|^{1/5}.$$

Considering the other case when $\Delta = 2$ we instead get

$$\mathcal{I}(A\times B, T) \ll 2|T| + \frac{1}{16}|A|^4|B|^3 + \frac{1}{2}|A|^2|B|^2 \ll |T|.$$

With the last part coming from our knowledge that $2 > \frac{|A|^{4/5}|B|^{3/5}}{|T|^{1/5}}$ and so $|T| \gg |A|^4|B|^3$. □

Now having an asymmetric version of Theorem 5.5 in Theorem 5.9 leads to the question can this provide an asymmetric version of Theorem 5.3? It turns out Theorem 5.9 is not needed but we still provide the asymmetric result below, noting why we do not need Theorem 5.9 within the proof itself.

Theorem 5.10. *Let $A\times B$ be a set of points in $\mathbb{F}_p^2$ with $|B| \leq p^{1/2}$, and T be a set of Möbius transformations. Then we have*

$$\mathcal{I}(A\times B, T) \ll |A|^{7/10}|B|^{1/2}|T|^{3/5}E(T)^{1/10} + |B|^{1/2}|T|$$

where

$$E(T) := \left|\left\{(f_1,f_2,f_3,f_4)\in T^4 : f_1f_2^{-1} = f_3f_4^{-1}\right\}\right|.$$

Proof. As to be expected this proof follows the structure of the proof of Theorem 5.3 with most of the differences coming from the differences in the starting Cauchy-Schwartz argument now that the point set is not symmetric. Another version of this argument (in an asymmetric context as will follow) is seen in Rudnev and Shkredov's [137], particularly their Theorem 8 dealing with incidences between points and lines. Note that we do not use, and thus start with, a pruning argument as we do in the proof of Theorems 5.3 and 5.4 although we could if we desired.

As with all of our uses of this sort of method, we start by expressing the number of incidences as a sum in the following manner

$$\sigma = \sum_{t\in T}\sum b\in B\mathbb{1}_A(t^{-1}(b)).$$

Having a sum our go to tool is that of Cauchy-Schwartz which leaves us with

$$\sigma^2 \leqslant |B|\sum_{u\in TT^{-1}} r_{TT^{-1}}(u)\sum_{a\in A}\mathbb{1}_A(u(a)).$$

This step has skipped over some of the details which can be found in the analogous step of the proof of Theorem 5.3, in particular the steps detailed in the set of Inequalities (5.10).

As before we shall again wish to move to a popular set of transforms and so define

$$\Delta := \frac{\sigma^2}{3|B||T|^2}.$$

And then as $\sum_{u\in TT^{-1}} r_{TT^{-1}}(u) = |T|^2$ we can turn to the pigeon hole principle to define a positive proportion, Ω, of the transforms that are at Δ-rich, that is

$$\forall u\in\Omega,\ \sum_{a\in A}\mathbb{1}_A(ua)\geqslant\Delta.$$

As before we can assume $\Delta\gg 1$ else we have that $\sigma\ll|B|^{1/2}|T|$ and we are done. This provides the last term of the incidence bound of Theorem 5.10.

To continue we again apply Cauchy-Schwartz (specifically to the summation over u restricted to Ω) to get

$$\sigma^4 \ll |B|^2E(T)\sum_{u\in\Omega}\left(\sum_{a\in A}\mathbb{1}_A(ua)\right)^2. \tag{5.28}$$

All that remains now is to deal with $\sum_{u\in\Omega}(\sum_{a\in A}\mathbb{1}_A(ua))^2$ term. To do this we will once again turn to an intermediate theorem and proceed as in Theorem 5.3. We do not need to use Theorem 5.9 because as our $u\in TT^{-1}$, u maps A to A so the asymmetric result is not needed. However we shall still use Theorem 5.9 but with $A=B$ and so the number of k-rich transforms is bounded by

$$|T_k| \ll \frac{|A|^7}{k^5}+\frac{|A|^4}{k^2}.$$

Following the method of dyadic pigeonholing and an incidence bound presented in the proof of Theorem 5.3 to arrive at an analogue of Bound (5.16) we start by defining the required Δ_k

and Ω_k for the pigeonholing argument so that $\Delta_k := 2^k\Delta$ and Ω_k the corresponding set of Δ_k-rich hyperbola so that for $u \in \Omega_k$, $\sum_{a\in A} \mathbb{1}_A(u(a)) \sim \Delta_k$. As we are following the method previously laid out in this thesis we shall be brief with some of the details which if desired can be found in the proof of Theorem 5.3. We use our dyadic argument (as in (5.15) from above), along with our fact that $\sum_{a\in A} \mathbb{1}_A(u(a)) \sim \Delta_k$, to get that

$$\sum_{u\in\Omega}\left(\sum_{a\in A}\mathbb{1}_A(ua)\right)^2 = \sum_k \sum_{u\in\Omega_k}\left(\sum_{a\in A}\mathbb{1}_A(ua)\right)^2 \sim \sum_k \Delta_k \sum_{u\in\Omega_k}\sum_{a\in A}\mathbb{1}_A(ua). \tag{5.29}$$

Recall that

$$|\Omega_k|\Delta_k \sim \sum_{u\in\Omega_k}\sum_{a\in A}\mathbb{1}_A(ua).$$

and we can apply Theorem 5.9 to bound $|\Omega_k|$ as follows

$$|\Omega_k|\Delta \ll \frac{|A|^7}{\Delta_k^4} + \frac{|A|^4}{\Delta_k}.$$

Returning to (5.29) this gives us that

$$\sum_{u\in\Omega}\left(\sum_{a\in A}\mathbb{1}_A(ua)\right)^2 \ll \sum_k\left(\frac{|A|^7}{\Delta_k^3} + |A|^4\right).$$

We must have $\Delta_k \ll |A|$ as the number of points on a hyperbola cannot be more than all the points so the left $\frac{|A|^7}{\Delta_k^3}$ term dominates the $|A|^4$ term leaving us with

$$\sum_k \frac{|A|^7}{\Delta_k^3} \ll \frac{|A|^7}{\Delta^3}.$$

Recalling our definition of Δ and plugging both it and the above back into our bound leads to, after grouping terms,

$$\sigma^{10} \ll |A|^7|B|^5E(T)|T|^6 \ll |A|^7|B|^5|T|^8M_1.$$

The second inequality coming from Lemma 5.1 and it basically the entire proof to move from Theorem 5.10 to Theorem 5.11. All that remains is to take the tenth root so the exponents are in the form desired,

$$\sigma \ll |A|^{7/10}|B|^{1/2}|T|^{3/5}E(T)^{1/10}.$$

We are done when we add in the extra term from when $\Delta \ll 1$.

□

Comparing this to Theorem 5.3 the obvious difference is we have succeeded in our goal of an asymmetric version. Comparing exponents (taking $A = B$ else we cannot do anything meaningful) we see that the main terms of both are the same, (that is the $|A|$ has an exponent of 6/5, $|T|$ is to the power 3/5 and the energy term to a tenth) that is Theorem 5.10 is a generalisation of Theorem 5.3 and reaches this without any weakening of the result required.

Note we could have used the following bound as proven in the proof of Theorem 5.3

$$\sum_{u\in\Omega}\left(\sum_{a\in A}\mathbb{1}_A(ua)\right)^2 \ll \min\left(\frac{|A|^7}{\Delta^3}, \frac{p|A|^4}{\Delta}\right).$$

This being what would happen if we used Theorem 5.5 rather than Theorem 5.9. This would lead to the end result of the same bound when the first term dominates and (ignoring the error term for now)

$$\sigma \ll p^{1/6}|A|^{2/3}|B|^{1/2}|T|^{1/3}E(T)^{1/6}.$$

Combining these gives the full bound of

$$\mathcal{I}(A\times B,T) \ll \min\left(|A|^{7/10}|B|^{1/2}|T|^{3/5}E(T)^{1/10}, p^{1/6}|A|^{2/3}|B|^{1/2}|T|^{1/3}E(T)^{1/6}\right) + |B|^{1/2}|T|.$$

We note that we could use Lemma 5.1 to bound the energy (by $E(H)\ll|H|^2M$) which leads to the following result. As before we could use a pruning method like at the beginning of the proof of Theorem 5.3 to do better than M in the case $|H|>|A|^{3/2}$ but I will not repeat this here instead just stating the result.

Theorem 5.11. *Let $A\times B\subseteq\mathbb{F}_p^2$ be a set of points with $|B|\leq p^{1/2}$, and let H be a set of translates of the hyperbola $xy=\pm 1$. Then we have*

$$\mathcal{I}(A\times B,H) \ll |A|^{1/2}|B|^{7/10}|H|^{4/5}M^{1/10} + |B|^{1/2}|H|,$$

where M is the maximum number of translates in H having the same abscissa or ordinate.

5.6 Applications

We continue the chapter by considering some applications of our incidence bound as well as weaknesses and possible areas of further study. Some of the applications and their proofs were corollaries in the paper [140], the rest of the applications stem from the generalisations in Subsection 5.5 and as such are first found in the paper [197].

5.6.1 Erdős Unit Distance Problem

Our first corollary is to do with the Erdős unit distance problem, discussed in Section 3.3.1. The standard problem is to do with Euclidean distance and the real numbers where Erdős conjectured in [45] that given n points in the plane, the maximum number of pairs at distance say 1 is $n^{1+\frac{c}{\log\log n}}$. Erdős managed an upper bound of $n^{3/2}$ which was improved to $n^{4/3}$ by Spencer, Szemeredi and Trotter [169]. We could instead ask this question for a different distance, say the Minkowski distance between $q=(x,y)$ and $q'=(x',y')$,

$$D_m(q,q') := (x-x')^2-(y-y')^2,$$

and using our result can prove the following.

Corollary 5.1. *Let $A \subset \mathbb{F}_p$, with $|A+A|, |A-A| \leqslant K|A| < \sqrt{p}$. Then the number of realisations of a nonzero Minkowski distance between points of $A \times A$ is $O(K^{6/5}|A|^{29/10})$.*

Proof. If $|A+A|, |A-A| \leqslant K|A|$ then after the transformation $(x,y) \mapsto (\frac{x+y}{2}, \frac{x-y}{2})$, the number of realisations of a nonzero Minkowski distance is bounded via the number of incidences between $(A+A) \times (A-A)$ with $|A|^2$ translates of the hyperbola $xy = 1$. Note that really we are applying a non balanced form of Estimate (5.4) of Theorem 5.3 (in that our point set is of the form $A \times B$ rather than $A \times A$) but as both $|A+A|, |A-A| \leqslant K|A|$ we end up back where we started with the normal $|A|^{6/5}$ being instead replaced by $K^{6/5}|A|^{6/5}$, when applying Estimate (5.4) we use $M_1 = |A|$.

□

This is sadly a lot more restricted than similar results such as Zahl's [204] Theorem 3.8 and we, unfortunately, do not see a way to remove the restrictions. This is due to the "bad" example being $A = A_1 \cup A_2$, with small $A_1 + A_1$ and $A_2 + A_2$ but large $A_1 + A_2$. A specific example could be to take A_1 as the arithmetic progression $1, 2, \ldots, n-1$ and A_2 as the arithmetic progression $1, n, 2n, dots$. Should one be able to deal with the bad example or find a similar incidence result to Theorem 5.3 which is not so reliant on the group structure of the set of linear fractional transformations and the inherited restrictions to rotations breaking the Cartesian product structure of the point set, improvements could be forthcoming.

I note that I am not aware of a nontrivial bound (better than $O(|P|^{3/2}) = O(|A|^3)$) on the number of realisations of a nonzero distance between pairs of a point set $P \subset \mathbb{F}_p^2$ in positive characteristic (in the case where $|P| < p$). However, I also note that Zahl [204] shows that we cannot improve upon this 3/2 exponent in *three* (rather than the two dimensions of our setup) under the conditions that -1 is not a square in $\mathbb{F}$ and $|P| < p^2$. This is due to the following example (from Zahl's paper), where by given a set of p^2 points in $\mathbb{F}_p^3$, you can find a non-zero $r \in \mathbb{F}_p$ so that there are about p^3 solutions to

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 = r^2.$$

This provides the 3/2 exponent.

Before finishing this subsection we shall consider whether we can do better using the more general Theorem 5.11. This would allow us to avoid some of the problems encountered that stop Corollary 5.1 from being a general bound on the number of unit Minkowski distances between two points. When you follow the method through however you get the result that for $P \subseteq \mathbb{F}_p^2$ be a set of $|P| \leq p^{15/26}$ points, then the number of unit Minkowski distances between two points of P is $O(|P|^{30/19})$. The issue is that the exponent is now greater than $3/2 < 30/19$ and so the result is worse than the trivial result gained from Cauchy-Schwartz, this is an issue with other applications if there is a Cauchy-Schwartz bound to beat, however, the increased generality makes up for this in some applications.

5.6.2 Sum-Product Results

We can also gain some sum-product type results from Estimate (5.4) of Theorem 5.3. Collecting them together into one corollary we get the following.

Corollary 5.2. *Let $A \subset \mathbb{F}_p$, with $|A| < \sqrt{p}$. Then*

$$\left|\{(a_1,a_2,a_3,a_4) \in A^4 : (a_1+a_2)(a_3+a_4) = 1\}\right|,$$
$$\left|\{(a_1,a_2,a_3,a_4) \in A^4 : (a_1+a_2-a_4)(a_3+a_2+a_4) = 1\}\right|,$$
$$\left|\{(a_1,a_2,a_3,a_4) \in A^4 : (a_1+a_2)(a_3+a_2a_4) = 1\}\right|,$$
$$\left|\{(a_1,a_2,a_3,a_4) \in A^4 : (a_1+a_2+a_4)(a_3+a_2a_4) = 1\}\right|$$

are all $O(|A|^{29/10})$.

Furthermore, if $|A+A| < K|A| < \sqrt{p}$, then the number of points of $A \times A$ on the hyperbola $xy = \lambda \neq 0$ is $O(K^{6/5}|A|^{29/10})$.

Proof. The first group of estimates follows directly by applying Estimate (5.4) with $M_1 = |A|$, and $|H| = |A|^2$. Explicitly, the number of solutions to $(a_1+a_2)(a_3+a_4) = 1$ (with $a_i \in A$) is the number of points from $A \times A$ on the hyperbola $y = \frac{1}{x+a_4} - a_2$, a translate of the hyperbola $yx = 1$ and so we can apply Estimate (5.4) with $|H| = |A|^2$ and $M_1 = |A|$ as stated above noting that the first term $|A|^{5/2}$ is subsumed by the second term $|A|^{6/5+8/5+1/10} = |A|^{29/10}$, our desired bound. The other three equations can be seen as hyperbola point incidences as follows, $(a_1+a_2-a_4)(a_3+a_2+a_4) = 1$ is equivalent to $(a_1+x-y)(a_3+x+y) = 1$ so after using the same transformations as in Section 5.2 of $x' = x+y$, $y' = x-y$ we instead have $(a_1+y')(a_3+x') = 1$ and can apply Estimate (5.4) as before after seeing we have $|A|^2$ such pairs of (y',x'). The third, $(a_1+a_2)(a_3+a_2a_4) = 1$ is equivalent to $(a_1+y)(a_3+yx) = 1$ so consider $x' = yx$ to get to the now familiar $(a_1+y)(a_3+x') = 1$, noting we have $|A|^2$ such combinations of y and x'. For the last estimate, of the set of four, on the solutions of $(a_1+a_2+a_4)(a_3+a_2a_4)$, consider $y' = x+y$ and $x' = xy$, still have $|A|^2$ such pairs of (y',x') and are again in our $(a_1+y')(a_3+x') = 1$ case.

For the last estimate, without loss of generality take $\lambda = -1$, then we write

$$\left|\{(a_1,a_2) \in A^2 : a_1a_2 = -1\}\right| \leqslant \frac{1}{|A|^2}\left|\{(a_1,a_2;s_1,s_2) \in A^2 \times (A+A)^2 : (s_1-a_1)(s_2-a_2) = -1\}\right|$$

and applying Estimate (5.4) to the set $A+A$ (instead of A in the above examples), with $M_1 = |A|$, and $|H| = |A|^2$ gives us

$$\sigma(A+A,H) \ll |A+A|^{1/2}|H| + |A+A|^{\frac{6}{5}}|H|^{\frac{4}{5}}M_1^{\frac{1}{10}} \ll |A+A|^{1/2}|A|^2 + |A+A|^{\frac{6}{5}}|A|^{\frac{17}{10}}$$

Using $|A+A| < K|A|$ finishes the proof.

□

Given suitable constraints, we can, in a similar manner to Lemma 2.7, prove the following lower bound

$$|HH^{-1}H| \gg |H|^{5/3}.$$

This is in the spirit of the expander style results of Section 2.6. Explicitly we have the following corollary.

Corollary 5.3. *Let $H \subseteq \mathbb{F}_q^2$ be a set of translates of the hyperbola $y = \frac{-1}{x}$. If $|H| < p$ and $M < |H|^{2/3}$*

$$|HH^{-1}H| \gg |H|^{5/3}.$$

Proof. In the same vein as the proof of Lemma 2.7 we use Cauchy-Schwarz to get the following

$$T_3(H) = \sum_x r^2_{HH^{-1}H}(x) \geqslant \frac{\left(\sum_x r_{HH^{-1}H}(x)\right)^2}{\sum_x \mathbb{1}^2 HH^{-1}H(x)}.$$

And hence, plugging in Lemma 5.4 for the case $\mathbb{F}_p$ and noting that for a general $\mathbb{F}$ of characteristic p, provided that $|H| < p$ the bottom case holds.

$$|H|^6 \leqslant T_3(H)|HH^{-1}H| \ll |HH^{-1}H||H|^3M^2 + |HH^{-1}H| \begin{cases} \frac{|H|^5}{p}, & \text{if } |H| > p^{5/4} \\ p^{2/3}|H|^{3+2/3}, & \text{if } p \leqslant |H| \leqslant p^{5/4} \\ |H|^{4+1/3}, & \text{if } |H| < p. \end{cases}$$

From here what remains is picking our preferred conditions and rearranging. For $|H| < p$ and $M < |H|^{2/3}$ we have

$$|HH^{-1}H| \gg |H|^{5/3}.$$

For $p \leqslant |H| < p^{5/4}$ and $M < p^{1/3}|H|^{1/3} \leqslant |H|^{2/3}$ we have

$$|HH^{-1}H| \gg \frac{|H|^{7/3}}{p^{2/3}} > |H|^{23/15}.$$

For $p^{5/4} \leqslant |H|$ and $M < \frac{|H|}{\sqrt{p}}$ we have

$$|HH^{-1}H| \gg p|H|.$$

Finally for when the $|H|^3M^2$ term dominates we have

$$|HH^{-1}H| \gg \frac{|H|^3}{M^2}.$$

□

We also can get a pair of expander results from our more general results of Section 5.5. The first of this pair of corollaries provides a three-variable expander, the second a four-variable expander. We will state them both before supplying the proofs which follow the same general method of relating the expander to an incidence and then bounding them via Theorem 5.11 or Theorem 5.9 respectively.

Corollary 5.4. *For all $A \subseteq \mathbb{F}_p$ with $|A| \leq p^{1/2}$, we have*

$$\left|\left\{a + \frac{1}{b-c} : a, b, c \in A\right\}\right| \gg |A|^{6/5}.$$

Corollary 5.5. *For all $A \subseteq \mathbb{F}_p$ with $|A| \leq p^{1/2}$, we have*

$$\left|\left\{\frac{ab+c}{b+d} : a,b,c,d \in A\right\}\right| \gg |A|^{4/3}.$$

Having stated these results, we recall Corollary 5.2 which also provided several four variable expanders. Corollary 5.2 was proven in the same manner as Corollary 5.4 will be; by rearranging the set into an incidence type problem and applying one of our bounds, Theorem 5.11 (similarly Theorem 5.9 for Corollary 5.5) in this case as opposed to Estimate (5.4) from Theorem 5.3 for Corollary 5.2.

Proof of Corollary 5.4. As stated our method shall be to find an incidence setup equivalent to the set we are considering. For ease we shall call this set

$$Q := \left\{a + \frac{1}{b-c} : a,b,c \in A\right\}.$$

To this end consider the point set $Q \times A$, and the set of hyperbolas

$$H := \{(y-a)(x-c) = 1 : a,b \in A\}.$$

The use of the term point set and defining a set of hyperbola should be suggestive of the rest of the proof. Our last requirement before we may apply Theorem 5.11 is the observation that the point (q,b) lies on the hyperbola $(y-c)(x-b) = 1$. This is because for all $a,b,c \in A$, we have

$$q = a + \frac{1}{b-c} \iff (b-c)(q-a) = 1.$$

Thus we have shown that $\mathcal{I}(Q \times A, H) \geq |A|^3$ and using Theorem 5.11 as an upper bound we get

$$|A|^3 \leqslant \mathcal{I}(Q \times A, H) \ll |Q|^{1/2}|A|^{12/5} + |A|^{5/2}.$$

We may ignore the error term as it may not dominate else we are left with the contradiction that $|A|^3 \leqslant |A|^{5/2}$ and so we are left to rearrange

$$|A|^3 \ll |Q|^{1/2}|A|^{12/5},$$

which yields the required answer of

$$|Q| \gg |A|^{6/5}.$$

□

Proof of Corollary 5.5. As in the proof of Corollary 5.4 we shall define a point set and a set of transformations so the expander is an incidence. To this end define T as the set of transformations T given by

$$y = \frac{ax+c}{x+d}$$

with $a,c,d \in A$ and $ad-c \neq 0$. For the point set we shall take $A \times Q$, where we define Q as our expanders,

$$Q := \left\{ \frac{ab+c}{b+d} : a,b,c,d \in A \right\}.$$

Note that we have an incident between the point $(b,q) \in A \times Q$ and a transform $y = \frac{ax+c}{x+d}$ if $q = \frac{ab+c}{b+d}$ and so $\mathcal{I}(Q \times A, T) \geq |A|^4$.

We now move to bound the number of such incidences above before comparing to our bound below. To do this we invoke Theorem 5.9 which gives us the following;

$$\mathcal{I}(A \times Q, T) \ll |A|^{4/5}|Q|^{3/5}|T|^{4/5} + |A|^{6/5}|Q|^{7/5}|T|^{1/5} + |T|.$$

We need to bound T in terms of A (or A and Q) to continue. As there are at most $|A|^2$ triples of $a,b,c \in A$ such that $ad-c=0$ (as once you have chosen two from the last is defined), we have that $|T| \gg |A|^3$ as the number of such transforms is all $|A|^3$ possibilities minus those bad examples with $ad-c=0$ of which there are at most $|A|^2$. This then means we can replace the $|T|$'s in the above incidence bound by $|A|^3$'s which gives us the bound

$$\mathcal{I}(A \times Q, T) \ll |A|^{16/5}|Q|^{3/5} + |A|^{9/5}|Q|^{7/5} + |A|^3.$$

Combining this with our lower bound and noting the last term cannot dominate else we have the contradiction that $|A|^4 \ll |A|^3$ leaves us with

$$|A|^4 \ll |A|^{16/5}|Q|^{3/5} + |A|^{9/5}|Q|^{7/5}.$$

If the second term dominates we can rearrange to get

$$|A|^{11/7} \ll |Q|,$$

this is a stronger result than we are trying to claim as $11/7 \approx 1.57 > 4/3$. If the first term instead dominates we get that

$$|A|^{4/3} \ll |Q|$$

as desired.

□

5.6.3 Number of Representations

Moving next to a bound on the number of representations of a non-zero number $\lambda \in \mathbb{F}_p^*$ as the product of two numbers in a subset $A \subseteq \mathbb{F}_p^*$, we get the following corollary.

Corollary 5.6. *Let $A \subseteq \mathbb{F}_p^*$ satisfy $|A+A| \leqslant K|A| \leqslant p^{1/2}$. Then for all non-zero $\lambda \in AA$, we have*

$$r_{AA}(\lambda) \ll K^{6/5}|A|^{9/10}.$$

Proof. The proof of this statement follows the general method of rearranging the problem to be an incidence problem between a point set and some hyperbolas before using Theorem 5.11 to obtain the bound. Our first step is one for convenience and it is that we may assume $\lambda = 1$ without loss of generality. Next note that given a pair $(a,b) \in A^2$ such that $ab = 1$ we may rewrite the equality using any two numbers $c,d \in A$ as follows,

$$\begin{aligned} 1 &= ab \\ &= (a+c-c)(b+d-d) \\ &= (Y-c)(X-d). \end{aligned}$$

We then have that Y and X are elements in $A+A$ and the last equation defines an incidence between the point set $P = (A+A)^2$ (the point (X,Y)) and the $|A|^2$ hyperbolas H given by $(y-c)(x-d) = 1$ We can now apply Theorem 5.11, which says that (with $M = |A|$)

$$r_{AA}(1)|A|^2 \leq \mathcal{I}(P,H) \ll |A+A|^{6/5}|A|^{17/10} + |A+A|^{1/2}|A|^2.$$

Whilst we have two terms here note that if the second term dominates and

$$r_{AA}(1) \ll |A+A|^{1/2} \leqslant k^{1/2}|A|^{1/2}.$$

This is even better than we are aiming to prove so we may assume that the first term always dominates as we are done in the other case. Noting our initial assumption that $|A+A| \leq K|A| \leq p^{1/2}$ means this leads to

$$r_{AA}(1)|A|^2 \leqslant \mathcal{I}(P,H) \ll k^{6/5}|A|^{29/10},$$

and so

$$r_{AA}(1) \ll k^{6/5}|A|^{9/10}$$

as desired.

□

Note that if we instead used Theorem 5.3, due to the addition of the pruning argument which comes into action here as $|H| = |A|^2 > k^{3/2}A|^{3/2} > |A+A|^{3/2}$ if $k < |A|^{1/3}$, we instead get

$$r_{AA}(1)|A|^2 \leqslant \mathcal{I}(P,H) \ll |A+A|^{14/11}|A|^{18/11} + |A+A|^{1/2}|A|^2.$$

We still do better if the second term dominates so assume the first does which leads to

$$r_{AA}(1)|A|^2 \ll |A+A|^{14/11}|A|^{18/11},$$

and thus

$$r_{AA}(1) \ll k^{14/11}|A|^{10/11}.$$

We could further generalise this to representations in AB which provides the following result.

Corollary 5.7. *Let $A,B \subseteq \mathbb{F}_p^*$ with $|A| = |B| = N$ satisfy $|A+B| \leq KN \leq p^{1/2}$. Then for all non-zero $\lambda \in AB$, we have*

$$r_{AB}(\lambda) \ll K^{6/5}N^{9/10}.$$

The proof of this is identical bar replacing the terms $(a,b) \in A^2$ by $(a,b) \in A \times B$ leading to an incidence between the point set $P = (A+B)^2$ and the N^2 hyperbolas H. Indeed we can go further as the condition $|A| = |B|$ is not strictly necessary above, as the proof can be followed with arbitrary sizes of A and B.

Corollary 5.8. *Let $A,B \subseteq \mathbb{F}_p^*$ satisfy $|A+B| \leq KN \leq p^{1/2}$, where $N = \max\{|A|,|B|\}$. Then for all non-zero $\lambda \in AB$, we have*

$$r_{AB}(\lambda) \ll\ll K^{6/5}\frac{|A|^{17/20}|B|^{17/20}}{N^{4/5}}.$$

Proof of Corollary 5.8. As in the proof of Corollary 5.6, we may assume $\lambda = 1$. As before, given a pair $(a,b) \in A \times B$ such that $ab = 1$, we have that for any pair $(a',b') \in A \times B$,

$$\begin{aligned} 1 &= ab \\ &= (a+b'-b')(b+a'-a') \\ &= (Y-c)(X-d). \end{aligned}$$

This is an incidence between a point of $P = (A+B)^2$ and one of the $|A||B|$ hyperbolas H given by $(Y-b')(X-a') = 1$ as above. We bound the number of such incidences with Theorem 5.11 as in the proof of Corollary 5.6 which gives us

$$r_{AB}(1)|A||B| \leq I(P,H) \ll |A+B|^{6/5}|A|^{17/20}|B|^{17/20} + |A+B|^{1/2}|A||B|.$$

As before if the second term dominates we are in an even better case so assuming instead the first dominates we get our result of

$$r_{AB}(1) \ll K^{6/5}N^{-4/5}|A|^{17/20}|B|^{17/20}.$$

□

5.6.4 A Variant of Beck's Theorem

Our next application is a version of Beck's theorem (detailed in Chapter 3 as Theorem 3.2) which instead of detailing the two possible extremes of how n points can fall on the set of lines (the two cases being a large fraction of the points on a single line or a large number of lines required to join all the points) instead considers what we can instead say when lines are replaced by Möbius transforms. Although not as clean with constants as Theorem 3.2, the following corollary still has the two main extremes of a large proportion of points on a single Möbius transformation or lots of Möbius transformations containing at least three points (the number of points required to uniquely define a Möbius transformations as opposed to the two required for a line). This result and proof comes from my joint paper with Warren [197].

Corollary 5.9. *Let $P \subseteq \mathbb{F}_p^2$ with be an arbitrary point set consisting of n points, with $n \leq p^{15/13}$. Then there exist positive constants C and K such one of the following two statements is true:*

- *There is a Möbius transformation containing at least $\frac{n}{C^{7/4}}$ of the points.*
- *There exists at least $\frac{n^{12/7}}{K}$ Möbius transformations, each of which containing at least three points of P.*

Proof. A Möbius transformation is uniquely defined by three points (this is analogous to lines being uniquely defined by two points). As such, given our point set P of size n a Möbius transformation is defined by P if it passes at least three points of P. We will consider all such Möbius transformations and will split these up depending on how rich they are. Note that for this proof we will use a slightly different definition of k-rich compared to previous uses in this thesis. Explicitly, for each $k \geq 0$, we define a transformation f to be 2^k-rich if *between* 2^k and $2^{k+1}-1$ points of P lie on the transform. Specifically, if $2^k \leq |f \cap P| < 2^{k+1}$.

Having set up our required definitions we shall now outline the proof. We will define a central range of rich but not too rich transforms and turn our attention to the points not on these transforms. We will show that there is a positive proportion of such points before aiming to show these points must then be in one of two cases; either we have a positive proportion of P lying on richer transforms, which will lead to the first conclusion, or on less rich transforms, in which we must have many transforms to support all of the points of P.

Making use of our current definition of 2^k-rich transforms, we note that by Theorem 5.8 we are able to bound the number of 2^k-rich transforms by

$$O\left(\frac{n^{15/4}}{2^{19j/4}} + \frac{n^2}{2^{2j}}\right).$$

Next, we will count how many points are on these 2^k-rich transforms, to this end, note that each of these 2^k-rich transforms contain $\Omega(2^{3j})$ triples of points of P, and so at most

$$O\left(\frac{n^{15/4}}{2^{7j/4}} + n^2 2^j\right)$$

triples of points are on a 2^j-rich transform.

To define our set or fairly rich transforms, let C be a large positive constant, and consider the transforms which are 2^j-rich for

$$Cn^{3/7} \leqslant 2^j \leqslant \frac{n}{C^{7/4}}.$$

By summing the number of triples of points on all such transforms, we have at most $O\left(\frac{n^3}{C^{7/4}}\right)$ triples of points on this collection of rich but not too rich transforms. Compare this to the total number of all triples of points $\frac{n(n-1)(n-2)}{6}$ to see that by taking C sufficiently large, a positive proportion of triples do not lie on a rich transform in our above defined collection. This means

that a positive proportion of triples lie on transforms which are either richer than our collection, or poorer. This will lead to the two cases in our corollary.

Explicitly what we have shown is that either a positive proportion of triples lie on transforms with less than $2Cn^{3/7}$ points, or more than $\frac{n}{C^{7/4}}$ points. Starting with if we are in the latter case, we are done as this is the first conclusion of our corollary. As such we shall now assume we are in the former case, and none of these transforms have more than $2Cn^{3/7}$ points.

For ease of notation, let T be this set of poor transformations. To finish we shall count the number of triples on these poor transforms. That is

$$\sum_{f\in T}|f\cap P|^3 \gg |P|^3 \implies |T| \gg n^{12/7}.$$

This proves the second case of the corollary. □

5.6.5 Protectively Equivalent Subsets

Our last application is a bound on the number of times we can find a projective transformation of a set as a subset of a larger set.

Corollary 5.10. *Let $A,S\subseteq\mathbb{F}_p$ with $|S|^3|A|^2\leq p^2$. Then the number of subsets $A'\subseteq A$ which are projectively equivalent to S is $O\left(\frac{|A|^3}{|S|}\right)$.*

Proof. We shall start by providing names to our sets A and $S\subseteq\mathbb{F}_p$ to aid the intuition of this result. We shall refer to A as our 'large' set and S the pattern set which then allows us to rephrase the above statement of how many A' are projectively equivalent to S as how many subsets of A are a projective transformation of our pattern set S or in how many projective transformations t do we have such that $t(S)\subseteq A$. We will denote by T the set of all such f.

Up to a scalar multiple, these projective transforms can be written in the form of 2×2 matrices with non-zero determinant which are our Möbius transformations. Recall that we have that a triple of points determines a unique Möbius transformation as two points define a unique line so we have a trivial upper bound of $|T|\ll|A|^3$.

Seeking to do better we will construct an incidence problem which we can apply to the upper bound for k-rich transformations from Theorem 5.9. As we already have a set of transforms T, we are left to choose a point set. As a given $f\in T$ transforms our pattern set S into A, a natural point set to consider, and indeed the one we do consider is $S\times A$. As $f\in T$ maps all of S into A it is $|S|$ rich as a transform on this point set. We now have a point set upon which all our transforms in T are $|S|$-rich so by our upper bound for k-rich transformations from Theorem 5.9 we have that

$$|T|\ll\frac{|S|^4|A|^3}{|S|^5}+\frac{|S|^2|A|^2}{|S|^2}=\frac{|A|^3}{|S|}+|A|^2.$$

We are done bar the error term however as $|S|\leqslant|A|$, else we can have no subsets of A equivalent due to miss matching cardinalities, we may subsume the error term into the first term. That is if

the error term dominates then $|S| \geqslant |A|$ and then there are zero projectively equivalent subsets and so any positive bound is trivially correct.

□

5.6.6 Kloosterman Sums

Another potential application is that of Kloosterman sums. Starting with a definition, let a, b, q be natural numbers, then

$$K(a,b;q) := \sum_{\substack{0 \leq x \leq q-1 \\ \gcd(x,q)=1}} e^{\frac{2\pi i}{q}(ax+bx^{-1})}.$$

Here x^{-1} is the inverse of x modulo q. Kloosterman sums are named for Hendrik Kloosterman who introduced them in [86]. In particular, in our finite field setting $\mathbb{F}$, we will define

$$K(n,m) = \sum_{x \in \mathbb{F}^*} \exp(nx + mx^{-1}) = K(mn, 1).$$

Our interest in this analytic number theoretic quantity comes from the connection between bilinear forms of Kloosterman sums and incidences between hyperbolae and points. These bilinear forms, with two weight functions $\alpha : \mathbb{F} \to \mathbb{C}$ and $\beta : \mathbb{F} \to \mathbb{C}$, are expressions of the form

$$S(\alpha, \beta) = \sum_{n,m} \alpha(n)\beta(m)K(n,m).$$

Work in the literature on these objects includes [51, 92, 94, 97, 97, 98, 100, 157, 159] among other papers.

Shkredov [157] proves the following result using Theorem 3.6.

Theorem 5.12 (Theorem 4)**.** *Let $\alpha, \beta : \mathbb{F}_p \to \mathbb{C}$ be functions with supports on $\{1, \dots, N\} + t_1$ and $\{1, \dots, M\} + t_2$, respectively, and N or M is at most p^{1-c}, $c > 0$. Then*

$$S(\alpha, \beta) \lesssim ||\alpha||_2 ||\beta||_2 p^{1-\delta},$$

where $\delta(c) > 0$ is a positive constant. Besides, if $M^2 < pN$, then

$$S(\{1, \dots, N\} + t_1, \beta) \lesssim ||\beta||_2 \left(N^{3/7} M^{1/7} p^{13/14} + N^{3/4} p^{3/4} + N^{1/4} p^{13/12} \right).$$

Note Shkedrov proves some more general results (his Theorems 33 and 34) which this is a particular case of. We should be able to prove a similar result using our hyperbola incidence result and this could be an area of further research.

Appendix A

Bisector Energy with Minkowski Distances

In Chapter 5 we used Murphy et al.'s Theorem 5.7 in our proof of Theorem 5.3 by moving to a field extension where -1 is a square so the Euclidean and Minkowski metrics coincide. Whilst this works there is not a reason you could not prove a similar result to Theorem 5.7 replacing Euclidean by Minkowski. In this appendix, whilst we will not prove such a theorem, we will introduce and talk about the required concepts and objects to do so, comparing and contrasting them against their Euclidean counterparts. In particular, we will consider Minkowski Isometrics and the Blaschke-Grünwald Kinematic Mapping. This corresponds to Sections 3.1 and 3.2 in Murphy et al. [110].

A.1 Minkowski Isometries

An Isometry is a distance preserving map, as such a Minkowski Isometry is a map preserving the Minkowski distance which we recall from Chapter 5 is defined[1] as

$$D_M(q,q') := (x-x')^2 - (y-y')^2.$$

In this section, we will consider the group of such isometries as well as comparing this with the group of Euclidean isometries to further our intuition. We will take our terminology from the most famous example of these isometries - those of four dimensions used in the theory of special relativity, before adapting the definition to our two-dimensional case as well as changing the field from the reals to arbitrary finite fields. For further details, see the book Modern Geometry Methods and Applications Part 1 [22]. As in this physical example, we will use the notation of time and space dimensions.

The group of all Minkowski spacetime isometries is the Poincare group, $IO(1,3)$. That is, in the classical case, the Poincare group is the group of affine transformations on $\mathbb{R}^4$ which preserve the Minkowski metric. This is the counterpart to the Euclidean group, E_n, which is the group

[1] See Definition 5.2.

generated by reflections, rotations, and translations of Euclidean space. The Poincare group can be generalised to general dimension, d, as

$$IO(1,d-1)(\mathbb{F}) := \mathbb{F}^{1,d-1} \rtimes \mathrm{O}(1,d-1)(\mathbb{F})$$

where we will define $\mathbb{F}^{1,d-1}$ and $\mathrm{O}(1,d-1)(\mathbb{F})$ below. The group operation is defined as

$$(\alpha, f)\cdot(\beta, g) = (\alpha + f\cdot\beta,\ f\cdot g).$$

We note that $IO(1,d-1)(\mathbb{F})$ is a linear algebraic group by Varadarajan and Virtanen [193].

The next groups we define are those covering what in the Euclidean setting would be seen as rotations. In the Euclidean setting, this would be the special orthogonal group $SO(d)$. In the physical setting of Minkowski space, the corresponding subgroup of the Poincare group is the proper Lorentz subgroup $SO(1,3)$, of the Lorentz subgroup $O(1,3)$ which consists of the rotations in the space like vectors $SO(3) \hookrightarrow SO(1,3)$ and the hyperbolic rotations or "Boosts" to the physicists. And so to the definitions, the isometries which are not translations in the Poincare group form the Lorentz group, $\mathrm{O}(1,d-1)(\mathbb{F})$. We note that image of the $SO(d-1) \hookrightarrow \mathrm{O}(1,d-1)$ are the rotations in space. In particular, $SO(1) \hookrightarrow \mathrm{O}(1,1)$ is either reflections in the x axis or the y depending on which of the space or time like dimensions it is embedded in. We define as the group of translations of the Artinian plane or equivalently the 2-D version of the Minkowski space $\mathbb{R}^{1,1}$ or $\mathbb{F}^{1,1}$ depending on the field. This is the equivalent of the translations of Euclidean space $\mathbb{R}^n$ or $\mathbb{F}^n$. We will call the two-dimensional group, $IO(1,1)(\mathbb{F})$, the Poincare group and this is what we shall mean in the following.

The Poincare group consists of three main types of isometry, translations in both the time and space dimensions, (which form the abelian lie group of translations on space-time), rotations just in the space or time dimensions (which is the rotation group $SO(d-1)$ in the space dimensions (which is Euclidean) and $SO(1)$ in the time dimensions (which is practically just reflections of the time axis about the origin)) and finally the transformations connecting two uniformly moving bodies (in two dimensions this will be the hyperbolic rotations). As above we will mean the 2 dimensional Lorentz group for the rest of the appendix.

We have some subgroups of the Lorentz group we may refer to, the proper Lorentz group $SO(d-1,1)$ is the subgroup of elements which have determinant $+1$ (as elements $SO(d-1,1) \hookrightarrow GL_d$ of the general linear group), which is analogous to $SO(n) \subset O(n)$ (noting that if -1 is a square, then it coincides with $SO(2)$ and circles and hyperbola are equivalent). As this terminology originates in physics, it mostly refers to these groups acting on real or complex numbers. In these cases we also the proper orthochronous (or restricted) Lorentz group $SO^+(d-1,1)$ which is the further subgroup of elements that also do not act by reflection along the timelike axis.

As a smooth manifold, the Lorentz group $O(d-1,1)(\mathbb{R})$ has four connected components. These are

$$SO^+(d-1,1)(\mathbb{R})$$

$$SO^+(d-1,1)(\mathbb{R})\cdot P$$

$$SO^+(d-1,1)(\mathbb{R})\cdot T$$

$$SO^+(d-1,1)(\mathbb{R})\cdot PT$$

where, as matrices $P := \text{diag}(1,-1,-1,\ldots,-1)$ is a reflection in space, $T := \text{diag}(-1,1,1,\ldots,1)$ is a reflection in time, and $PT = TP = \text{diag}(-1,-1,\ldots,-1) = -\text{id}$.

In our finite fields setting we do not have such separation as we do not have the distinction between positive and negative numbers and thus also don't have the distinction between orthochronous and non-orthochronous. As such we have that $ISO^*(1,1)(\mathbb{F})$ is an order 2 subgroup (as would be desire following the methodology in [110] where $SF_2(\mathbb{F})$ is also an order 2 subgroup of the group of all distance (Euclidean) preserving transformations and this is what we are seeking to emulate) rather than an order 4 subgroup as is the case over infinite fields. In the infinite case our methods would not work however those of you interested can find more work in this direction due to Roche-newton and Rudnev [127]. Explicitly in our two dimensional, finite fields case, the hyperbolic rotations over F_p as a group is cyclic with $p-1$ elements.

Looking at the two-dimensional case in more detail, we can express elements as 3x3 matrices as they are the semidirect product of some rotations and transformations, explicitly elements are of the form

$$\begin{pmatrix} u & v & s \\ v & u & t \\ 0 & 0 & 1 \end{pmatrix}, \text{ where } u^2 - v^2 = 1.$$

;

As stated above, this is reminiscent of the Euclidean $SF_2(\mathbb{F})$ of positively oriented rigid motions of $\mathbb{F}^2$ is generated by $SO_2(\mathbb{F})$ and $T_2(\mathbb{F})$; an analogue of the special Euclidean group $SE_2(\mathbb{R})$. And thus elements of $SF_2(\mathbb{F})$ can be expressed in the form

$$\begin{pmatrix} u & -v & s \\ v & u & t \\ 0 & 0 & 1 \end{pmatrix}, \text{where } u^2 + v^2 = 1.$$

From these, it is clear where the Euclidean x^2+y^2 quadratic form and the Minkowski Quadratic form x^2-y^2 appear. We can re-frame the above using hyperbolic trigonometric functions (or standard trigonometric functions in the case of $SF_2(\mathbb{F})$).

$$ISO^*(1,1)(\mathbb{F}) := \begin{pmatrix} \cosh\theta & \sinh\theta & s \\ \sinh\theta & \cosh\theta & t \\ 0 & 0 & 1 \end{pmatrix}.$$

A.1.1 Axial Symmetries

Apart from getting a greater understanding of where the differences between the Euclidean and Minkowski cases are, the above groups will allow the definition of a Minkowski axial symmetry

which would be required for a proof on a Minkowski version of Theorem 5.7. In [110] they define axial symmetries as

> " a reflection through a non-isotropic line an axial symmetry. "

They also note that coset of $SF_2(\mathbb{F})$ (order 2 in the group of all distance-preserving transformations) consists of compositions of reflection through some (non-isotropic) line, and a translation parallel to this line, in particular these axial symmetries. This has an obvious problem when we first seek to generalise this into our hyperbolic case, reflections are not necessarily isometries and thus we cannot just take all reflections through non-isotropic lines. However, we do at least an order two subgroup in the group of all distance-preserving transformation mimicking $SF_2(\mathbb{F})$. What we can do is take reflections about the x-axis and thus can apply a rotation so our line of reflection is the x-axis and rotate back as necessary. These will be what we call axial symmetries in our case and, along with compositions with a translation, form the other coset of $ISO^*(1,1)(\mathbb{F})$ and we continue to fit in with the narrative. For two points, x and y we will denote by $x \sim_l y$ that x is axially symmetric to y relative to a line l, again following the notation of [110].

Having got a framework of what we need a Minkowski Axial Symmetry, it is useful to be able to express them concretely. Unsurprisingly we turn to matrices for this task. To this end we remember that a hyperbolic rotation through an angle of θ can be given by

$$\begin{pmatrix} \cosh\theta & \sinh\theta \\ \sinh\theta & \cosh\theta \end{pmatrix},$$

and a reflection in the x-axis (equivalently can be seen as a rotation in the one-dimensional space-like coordinates) is given by

$$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Multiplying these gives

$$\begin{pmatrix} \cosh\theta & \sinh\theta \\ \sinh\theta & \cosh\theta \end{pmatrix}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \cosh\theta & -\sinh\theta \\ -\sinh\theta & \cosh\theta \end{pmatrix} = \begin{pmatrix} -\cosh 2\theta & \sinh 2\theta \\ -\sinh 2\theta & \cosh 2\theta \end{pmatrix}.$$

We note this assumes the rotations are about the origin and thus our line l passes through the origin. If this is not the case we can apply a translation beforehand and afterwards to deal with this.

It is also of use to be able to calculate the composition of two axial symmetries after sorting translations so the point of intersection between the two lines, l and l', is at the origin. The calculation is as follows

$$\begin{pmatrix} -\cosh 2\theta & \sinh 2\theta \\ -\sinh 2\theta & \cosh 2\theta \end{pmatrix}\begin{pmatrix} -\cosh 2\phi & \sinh 2\phi \\ -\sinh 2\phi & \cosh 2\phi \end{pmatrix} = \begin{pmatrix} \cosh(2\phi - 2\theta) & \sinh(2\phi - 2\theta) \\ \sinh(2\phi - 2\theta) & \cosh(2\phi - 2\theta) \end{pmatrix}.$$

To follow the ideas of [110], the reason behind defining such objects is so the set of axial symmetries can be mapped $ISO^*(1,1)(\mathbb{F})$ to which there exists a kinematic mapping (see next section) so that the image of these axial symmetries is a plane and we can make use of point plane incidence theorems such as Theorem 3.3.

The way we can map the set of axial symmetries to $ISO^*(1,1)(\mathbb{F})$ is by composing axial symmetries with a fixed axial symmetry ρ relative to a non-isotropic subspace ℓ_τ. We can take ℓ_τ to be the x-axis, which means we can calculate this by translating the point on the origin we are rotating about to the origin, rotate it and then translate it back. In matrices that is

$$\begin{pmatrix} 1 & 0 & x \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u & v & 0 \\ v & u & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -x \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

In this case, the image of the axial symmetries under this mapping which we will denote by R_τ is

$$R_\tau = \left\{ \begin{pmatrix} u & v & x(1-u) \\ v & u & -vx \\ 0 & 0 & 1 \end{pmatrix} : u^2 - v^2 = 1, \text{ and } u,v,x \in \mathbb{F} \right\}.$$

This object is then amenable to a Blaschke-Grünwald kinematic style mapping from the next section which allows us to map this to geometric objects, in particular, this R_τ is contained in a plane.

A.2 Blaschke-Grünwald Kinematic Mapping

The idea of this section is to define a mapping from $ISO(1,1)(\mathbb{F})$, the isometries from the previous section into $\mathbb{PF}^3$ which allows geometric incidence results to be brought to the fore. The quintessential example is the Blaschke-Grünwald kinematic mapping discovered independently by Blaschke [8] and Grünwald [65], which provides a mapping from the group of isometries in the plane to three dimensional projective space. Explicitly the mapping we which to tweak is

$$\kappa' : \begin{pmatrix} \cos\theta & \sin\theta & a \\ \sin\theta & \cos\theta & b \\ 0 & 0 & 1 \end{pmatrix} \mapsto [X_0 : X_1 : X_2 : X_3]$$

where we define

$$X_0 := 2\cos\theta, X_1 := 2\sin\theta, X_2 := a\sin\theta + b\cos\theta, X_3 := a\cos\theta - b\sin\theta.$$

For Cayley-Klein Geometries the geometries can be described by the homogeneous Clifford algebra model and a kinematic mapping can be constructed following methods in [85] and Appendix A of [110].

Note another example is Study's [173] mapping from isometries in three dimensions to seven-dimensional projective space, this is often referred to as Study's quadric. This has been studied concerning a Minkowski equivalent by Uğurlu and Çalişkan [192].

The Blaschke-Grünwald kinematic mapping is based on half angle formulas which also exist for hyperbolic trigonometric functions, as such in our Minkowski setting we would instead define $\kappa : ISO(1,1)(\mathbb{F}) \to \mathbb{PF}^3$

$$\kappa : \begin{pmatrix} \cosh\theta & \sinh\theta & a \\ \sinh\theta & \cosh\theta & b \\ 0 & 0 & 1 \end{pmatrix} \mapsto [X_0 : X_1 : X_2 : X_3],$$

where we now define

$$X_0 := 2\cosh\theta, X_1 := 2\sinh\theta, X_2 := a\sinh\theta - b\cosh\theta, X_3 := a\cosh\theta - b\sinh(\theta).$$

Equivalently

$$\cosh(2\theta) = \frac{X_0^2 + X_1^2}{X_0^2 - X_1^2}, \sinh(2\theta) = \frac{2X_0X_1}{X_0^2 - X_1^2}, a = \frac{2X_1X_2 - 2X_0X_3}{X_0^2 - X_1^2}, b = \frac{2X_1X_3 - 2X_0X_2}{X_0^2 - X_1^2}.$$

Under this kinematic mapping, the image of R_τ (from the previous section) is contained in the plane $X_2 = 0$ which allows point plane incidences to be brought to bear.

The required properties of this and other such mappings can be proven using Clifford Algebras, examples of this may be found in [85].

Bibliography

[1] E. ACKERMAN, *On topological graphs with at most four crossings per edge*, Computational Geometry, 85 (2019), p. 101574.

[2] P. AGARWAL AND B. ARONOV, *Counting facets and incidences*, Discrete & Computational Geometry, 7 (1992), pp. 359–369.

[3] B. ARONOV, J. PACH, M. SHARIR, AND G. TARDOS, *Distinct distances in three and higher dimensions*, Combinatorics, Probability and Computing, 13 (2004), p. 283–293.

[4] A. BALOG, O. ROCHE-NEWTON, AND D. ZHELEZOV, *Expanders with superquadratic growth*, Electronic Journal of Combinatorics, 24 (2016).

[5] A. BALOG AND E. SZEMER´EDI, *A statistical theorem of set addition*, Combinatorica, 14 (1994), pp. 263–268.

[6] A. BALOG AND T. D. WOOLEY, *A low-energy decomposition theorem*, The Quarterly Journal of Mathematics, 68 (2016), pp. 207–226.

[7] J. BECK, *On the lattice property of the plane and some problems of Dirac, Motzkin and Erdős in combinatorial geometry*, Combinatorica, 3 (1983), pp. 281–297.

[8] W. BLASCHKE, *Euklidische kinematik und nichteuklidische geometrie*, Zeitschr. Math. Phys., 60 (1911), pp. 61–91 and 203–204.

[9] J. BOURGAIN, *On the Erdős-Volkmann and Katz-Tao ring conjectures*, Geometric and Functional Analysis - GEOM FUNCT ANAL, 13 (2003), pp. 334–365.

[10] ——, *A modular Szemerédi–Trotter theorem for hyperbolas*, Comptes Rendus Mathematique, 350 (2012), pp. 793 – 796.

[11] J. BOURGAIN AND A. GAMBURD, *Uniform expansion bounds for Cayley graphs of $SL_2(\mathbb{F}_p)$*, Annals of Mathematics, 167 (2008), pp. 625–642.

[12] J. BOURGAIN AND M. Z. GARAEV, *On a variant of sum-product estimates and explicit exponential sum bounds in prime fields*, Mathematical Proceedings of the Cambridge Philosophical Society, 146 (2009), p. 1–21.

[13] J. BOURGAIN, A. GLIBICHUK, AND S. KONYAGIN, *Estimates for the number of sums and products and for exponential sums in fields of prime order*, Journal of the London Mathematical Society, 73 (2006), pp. 380 – 398.

[14] J. BOURGAIN, N. KATZ, AND T. TAO, *A sum-product estimate in finite fields, and applications*, Geom. funct. anal., 14 (2004), pp. 27–57.

[15] J. BOURGAIN AND S. KONYAGIN, *Estimates for the number of sums and products and for exponential sums over subgroups in fields of prime order*, Comptes Rendus Mathematique, 337 (2003), pp. 75–80.

[16] P. BRASS, W. O. J. MOSER, AND J. PACH, *Research Problems in Discrete Geometry*, Springer London, Limited, 2007.

[17] E. BREUILLARD, *Lecture on approximate groups*, June 2010.

[18] E. BREUILLARD AND B. GREEN, *Approximate groups, III: the unitary case*, Turkish Journal of Mathematics, 36 (2010).

[19] ——, *Approximate groups, I The torsion-free nilpotent case*, Journal of the Institute of Mathematics of Jussieu, 10 (2011), p. 37–57.

[20] ——, *Approximate groups, II: The solvable linear case*, Q. J. Math., 62 (2011), pp. 513–521.

[21] E. BREUILLARD, B. GREEN, AND T. TAO, *Approximate subgroups of linear groups*, Geom. Funct. Anal., 21 (2011), pp. 774–819.

[22] R. BURNS, B. DUBROVIN, A. FOMENKO, AND S. NOVIKOV, *Modern Geometry — Methods and Applications: Part I: The Geometry of Surfaces, Transformation Groups, and Fields*, Graduate Texts in Mathematics, Springer New York, 1991.

[23] M.-C. CHANG, *A sum–product theorem in semi-simple commutative banach algebras*, Journal of Functional Analysis, 212 (2004), pp. 399–430.

[24] ——, *A sum-product estimate in algebraic division algebras*, Israel Journal of Mathematics, 150 (2005), pp. 369–380.

[25] ——, *Additive and multiplicative structure in matrix spaces*, Combinatorics, Probability & Computing, 16 (2007), pp. 219–238.

[26] J. CHAPMAN, M. ERDOGAN, D. HART, A. IOSEVICH, AND D. KOH, *Pinned distance sets, k-simplices, Wolff's exponent in finite fields and sum-product estimates*, Mathematische Zeitschrift, 271 (2009), pp. 63–93.

[27] C. CHEN, B. KERR, AND A. MOHAMMADI, *A new sum–product estimate in prime fields*, Bulletin of the Australian Mathematical Society, 100 (2019), p. 268–280.

[28] F. CHUNG, *The number of different distances determined by n points in the plane*, Journal of Combinatorial Theory, Series A, 36 (1984), pp. 342–354.

[29] F. CHUNG, E. SZEMERÉDI, AND W. TROTTER, *The number of different distances determined by a set of points in the Euclidean plane*, Discrete & Computational Geometry, 7 (1992), pp. 1–11.

[30] J. CILLERUELO, A. IOSEVICH, O. ROCHE-NEWTON, AND M. RUDNEV, *Elementary methods for incidence problems in finite fields*, Acta Arithmetica, 177 (2014).

[31] K. CLARKSON, H. EDELSBRUNNER, L. GUIBAS, M. SHARIR, AND E. WELZL, *Combinatorial complexity bounds for arrangements of curves and spheres*, Discrete and Computational Geometry, 5 (1990), pp. 99–160.

[32] M. CORDES, T. HARTNICK, AND V. TONI'C, *Foundations of geometric approximate group theory*, 2020.

[33] O. DINAI, *Expansion properties of finite simple groups*, PhD thesis, The Hebrew University of Jerusalem, Sep 2009.

[34] T. T. DO, *Extending erdős-beck's theorem to higher dimensions*, Comput. Geom., 90 (2020), p. 101625.

[35] Z. DVIR, *On the size of Kakeya sets in finite fields*, Journal of the American Mathematical Society, 22 (2008).

[36] S. EBERHARD, B. MURPHY, L. PYBER, AND E. SZABÓ, *Growth in linear groups*, 2021.

[37] H. EDELSBRUNNER, *Algorithms in Combinatorial Geometry*, Springer Publishing Company, Incorporated, 1st ed., 2012, ch. 6.5 Lower bounds for many cells.

[38] H. EDELSBRUNNER AND E. WELZL, *On the maximal number of edges of many faces in an arrangement*, Journal of Combinatorial Theory, Series A, 41 (1986), pp. 159–166.

[39] G. EDGAR AND C. MILLER, *Borel subrings of the reals*, Proceedings of the American Mathematical Society, 131 (2001).

[40] G. ELEKES, *On linear combinatorics I. concurrency - an algebraic approach.*, Combinatorica, 17 (1997), pp. 447–458.

[41] ——, *On linear combinatorics II. structure theorems via additive number theory*, Combinatorica, 18 (1998), pp. 13–25.

[42] ——, *Sums versus products in Number Theory, Algebra and Erdős Geometry — a survey*, Bolyai Math. Soc. Stud 11, 2002, p. 241–290.

[43] G. ELEKES AND C. TÓTH, *Incidences of not-too-degenerate hyperplanes*, in Proceedings of the Annual Symposium on Computational Geometry, 01 2005, pp. 16–21.

[44] M. ERDOGAN, *A bilinear Fourier extension theorem and applications to the distance set problem*, International Mathematics Research Notices, 2005 (2005), p. 1411.

[45] P. ERDŐS, *On sets of distances of n points*, The American Mathematical Monthly, 53 (1946), pp. 248–250.

[46] ——, *Some remarks on number theory (in hebrew)*, Riveon Lematematika, 9 (1955), pp. 45–48.

[47] ——, *An asymptotic inequality in the theory of numbers, (in russian)*, Vestnik Leningrad. Univ., 15 (1960), pp. 41–49.

[48] ——, *On sets of distances of n points in Euclidean space)*, Magyar Tud. Akad. Mat. Kut. Int. Közl., 5 (1960), pp. 165–169.

[49] P. ERDŐS AND E. SZEMERÉDI, *On sums and products of integers*, in Studies in pure mathematics, Springer, 1983, pp. 213–218.

[50] P. ERDŐS AND B. VOLKMANN, *Additive gruppen mit vorgegebener Hausdorffscher dimension.*, Journal für die reine und angewandte Mathematik (Crelles Journal), 1966 (1966), pp. 203–208.

[51] ÉTIENNE FOUVRY, E. KOWALSKI, AND P. MICHEL, *Algebraic trace functions over the primes*, Duke Mathematical Journal, 163 (2014), pp. 1683 – 1736.

[52] K. J. FALCONER, *On the Hausdorff dimensions of distance sets*, Mathematika, 32 (1985), p. 206–212.

[53] ——, *Dimensions of intersections and distance sets for polyhedral norms.*, Real analysis exchange, 30 (2005), pp. 719–726.

[54] K. FORD, *Sums and products from a finite set of real numbers*, The Ramanujan Journal, 2 (1998), pp. 59–66.

[55] ——, *The distribution of integers with a divisor in a given interval*, Annals of Mathematics, 168 (2008), pp. 367–433.

[56] G. A. FREIMAN, *Foundations of a structural theory of set addition*, American Mathematical Society, Providence, R. I., 1973.
Translated from the Russian, Translations of Mathematical Monographs, Vol 37.

[57] M. Z. GARAEV, *An Explicit Sum-Product Estimate in* $\mathbb{F}_p$, International Mathematics Research Notices, (2007).

[58] ——, *The sum-product estimate for large subsets of prime fields*, in Proceedings of the American Mathematical Society, vol. 136, American Mathematical Society, 2008, pp. 2735–2739.

[59] J. GARIBALDI, A. IOSEVICH, AND S. SENGER, *The Erdős distance problem*, American Mathematical Society, 2011.

[60] N. GILL AND H. HELFGOTT, *Growth in solvable subgroups of $GL_r(\mathbb{Z}/p\mathbb{Z})$*, Math. Ann., 360 (2014), pp. 157–208.

[61] T. GOWERS, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, GAFA, 8 (1998), pp. 529–551.

[62] B. GREEN AND I. Z. RUZSA, *Freiman's theorem in an arbitrary abelian group*, Journal of the London Mathematical Society, 75 (2007), pp. 163–175.

[63] B. GREEN AND T. TAO, *On sets defining few ordinary lines*, Discrete & Computational Geometry, 50 (2013), pp. 409–468.

[64] C. GROSU, *$\mathbb{F}_p$ is locally like $\mathbb{C}$*, Journal of the London Mathematical Society, 89 (2014), pp. 724–744.

[65] J. GRÜNWALD, *Ein abbildungsprinzip, welches die ebene geometrie und kinematikmit der räumlichen geometrie verknüpft*, Sitzber. Ak. Wiss. Wien., 20 (1911), pp. 677–741.

[66] L. GUTH, A. IOSEVICH, Y. OU, AND H. WANG, *On Falconer's distance set problem in the plane*, Inventiones mathematicae, 219 (2018), pp. 779–830.

[67] L. GUTH AND N. KATZ, *On the Erdős distinct distance problem in the plane*, Annals of Mathematics, 181 (2010).

[68] D. HART, A. IOSEVICH, D. KOH, AND M. RUDNEV, *Averages over hyperplanes, sum-product theory in vector spaces over finite fields and the Erdős-Falconer distance conjecture*, Transactions of the American Mathematical Society, 363 (2007).

[69] D. HART, A. IOSEVICH, AND J. SOLYMOSI, *Sum-product estimates in finite fields via Kloosterman sums*, International Mathematics Research Notices, 2007 (2007).

[70] N. HEGYVÁRI AND F. HENNECART, *A structure result for bricks in Heisenberg groups*, Journal of Number Theory, 133 (2013), pp. 2999–3006.

[71] ——, *Substructure for product sets in the Heisenberg groups*, Moscow Journal of Combinatorics and Number Theory, 3 (2013), pp. 57–68.

[72] ——, *Expansion for cubes in the Heisenberg group*, Forum Mathematicum, 30 (2018), pp. 227–236.

[73] H. HELFGOTT, *Growth and generation in $SL_2(\mathbb{Z}/p\mathbb{Z})$*, Ann. of Math. (2), 167 (2008), pp. 601–623.

[74] ——, *Growth in $SL_3(\mathbb{Z}/p\mathbb{Z})$*, J. Eur. Math. Soc. (JEMS), 13 (2011), pp. 761–851.

[75] ——, *Growth in groups: ideas and perspectives*, Bull. Amer. Math. Soc. (N.S.), 52 (2015), pp. 357–413.

[76] H. HELFGOTT AND M. RUDNEV, *An explicit incidence theorem in* $\mathbb{F}_p$, Mathematika, 57 (2011), pp. 135 – 145.

[77] H. HELFGOTT AND A. SERESS, *On the diameter of permutation groups*, Ann. of Math. (2), 179 (2014), pp. 611–658.

[78] A. IOSEVICH, *On the unit distance problem*, 2017.
arXiv:1709.08048 math.CA.

[79] A. IOSEVICH AND M. RUDNEV, *Erdős distance problem in vector spaces over finite fields*, Transactions of the American Mathematical Society, 359 (2005), pp. 6127–6142.

[80] T. JONES, *Explicit incidence bounds over general finite fields*, Acta Arithmetica, 150 (2010).

[81] H. KAPLAN, J. MATOUŠEK, Z. SAFERNOVÁ, AND M. SHARIR, *Unit distances in three dimensions*, Combinatorics, Probability and Computing, 21 (2012), p. 597–610.

[82] N. H. KATZ AND C. YEN SHEN, *A slight improvement to Garaevs sum product estimate*, Proc. Amer. Math. Soc, (2008), pp. 2499–2504.

[83] N. N. KATZ AND G. TARDOS, *A new entropy inequality for the Erdős distance problem*, Contemporary Mathematics, 342 (2004), pp. 119–126.

[84] L. M. KELLY AND W. O. J. MOSER, *On the number of ordinary lines determined by n points*, Canadian Journal of Mathematics, 10 (1958), p. 210–219.

[85] D. KLAWITTER AND M. HAGEMANN, *Kinematic mappings for Cayley-Klein geometries via Clifford algebras*, Beiträge zur Algebra und Geometrie / Contributions to Algebra and Geometry, 54 (2013).

[86] H. D. KLOOSTERMAN, *On the representation of numbers in the form* $ax^2 + by^2 + cz^2 + dt^2$, Acta Mathematica, 49 (1926), pp. 407 – 464.

[87] D. KOH, T. PHAM, AND L. VINH, *Extension theorems and a connection to the Erdős-Falconer distance problem over finite fields*, 2020.
arXiv:1809.08699v9 math.CA.

[88] J. KOLLÁR, *Szemerédi–Trotter-type theorems in dimension 3*, Advances in Mathematics, 271 (2015).

[89] S. KONYAGIN AND I. SHKREDOV, *On sum sets of sets, having small product set*, Proceedings of the Steklov Institute of Mathematics, 290 (2015).

[90] ——, *New results on sums and products in* $\mathbb{R}$, Proceedings of the Steklov Institute of Mathematics, 294 (2016), pp. 78–88.

[91] S. KONYAGIN AND I. ŁABA, *Distance sets of well-distributed planar sets for polygonal norms*, Israel Journal of Mathematics, 152 (2006), pp. 157–179.

[92] M. KOROLEV AND I. SHPARLINSKI, *Sums of algebraic trace functions twisted by arithmetic functions*, Pacific Journal of Mathematics, 304 (2020), pp. 505–522.

[93] D. KOUKOULOPOULOS, *Divisors of shifted primes*, International Mathematics Research Notices, 24 (2010), pp. 4585–4627.

[94] E. KOWALSKI, P. MICHEL, AND W. SAWIN, *Bilinear forms with Kloosterman sums and applications*, Annals of Mathematics, 186 (2017), pp. 413–500.

[95] L. LI, *Slightly improved sum-product estimates in fields of prime order*, Acta Arithmetica, 147 (2009).

[96] L. LI AND O. ROCHE-NEWTON, *An improved sum-product estimate for general finite fields*, SIAM J. Discrete Math., 25 (2011), pp. 1285–1296.

[97] K. LIU, I. E. SHPARLINSKI, AND T. ZHANG, *Cancellations between Kloosterman sums modulo a prime power with prime arguments*, Mathematika, 65 (2019), p. 475–487.

[98] S. MACOURT AND I. E. SHPARLINSKI, *Double sums of Kloosterman sums in finite fields*, Finite Fields and Their Applications, 60 (2019), p. 101575.

[99] P. MATTILA, *Spherical averages of Fourier transforms of measures with finite energy; dimensions of intersections and distance sets*, Mathematika, 34 (1987), p. 207–228.

[100] P. MICHEL, V. BLOMER, E. FOUVRY, E. KOWALSKI, AND D. MILICEVIC, *On moments of twisted l-functions*, American Journal of Mathematics, 139 (2017).

[101] H. MINKOWSKI, *Raum und zeit*, Physikalische Zeitschrift, 10 (1909), pp. 75–88.

[102] ——, *Die grundgleichungen für die elektromagnetischen vorgänge in bewegten körpern*, Mathematische Annalen, 68 (1910), pp. 472–525.

[103] A. MOHAMMADI AND S. STEVENS, *Attaining the exponent* 5/4 *for the sum-product problem in finite fields*, 2021.
arXiv:2103.08252v3 math.CO.

[104] L. MOSER, *On the different distances determined by n points*, The American Mathematical Monthly, 59 (1952), pp. 85–91.

[105] A. MUDGAL, *Sum-product estimates for diagonal matrices*, Bulletin of the Australian Mathematical Society, 103 (2020), pp. 1–10.

[106] B. MURPHY, *Upper and lower bounds for rich lines in grids*, Amer. J. of Math., (2017).

[107] B. MURPHY AND G. PETRIDIS, *A point-line incidence identity in finite fields, and applications*, Moscow Journal of Combinatorics and Number Theory,, (2016).
30 pages.

[108] ——, *Products of difference over arbitrary finite fields*, Discrete Anal., (2018), pp. Paper No. 18, 42.

[109] ——, *An example related to the Erdős-Falconer question over arbitrary finite fields*, Bull. Hellenic Math. Soc., 63 (2019), pp. 38–39.

[110] B. MURPHY, G. PETRIDIS, T. PHAM, M. RUDNEV, AND S. STEVENS, *On the pinned distances problem over finite fields*, 2020.
arXiv:2003.00510 math.CO.

[111] B. MURPHY, G. PETRIDIS, O. ROCHE-NEWTON, M. RUDNEV, AND I. SHKREDOV, *New results on sum-product type growth over fields*, Mathematika, 65 (2017), pp. 588–642.

[112] B. MURPHY, O. ROCHE-NEWTON, AND I. SHKREDOV, *Variations on the sum-product problem*, SIAM Journal on Discrete Mathematics, 29 (2013).

[113] B. MURPHY AND J. WHEELER, *Growth in some finite three-dimensional matrix groups*, SIAM J. Discret. Math., 34 (2020), pp. 1984–1998.

[114] M. NATHANSON, *On sums and products of integers*, Proceedings of the American Mathematical Society, 125 (1997).

[115] M. NATHANSON AND G. TENENBAUM, *Inverse theorems and the number of sums and products*, Asterisque- Societe Mathematique de France, 258 (1999).

[116] J. PACH, R. RADOICIC, G. TARDOS, AND G. TOTH, *Improving the crossing lemma by finding more crossings in sparse graphs*, Discrete & Computational Geometry, 36 (2006), pp. 527–552.

[117] J. PACH AND M. SHARIR, *On the number of incidences between points and curves*, Combinatorics, Probability and Computing, 7 (1998), p. 121–127.

[118] J. PACH AND G. TÓTH, *Graphs drawn with few crossings per edge*, Combinatorica, 17 (1997), pp. 427–439.

[119] G. PETRIDIS, *New proofs of Plünnecke-type estimates for product sets in groups*, Combinatorica, 32 (2011).

[120] ——, *Pinned algebraic distances determined by cartesian products in* $\mathbb{F}_p^2$, 2017.
arXiv:1610.03172 math.CO.

[121] G. PETRIDIS, O. ROCHE-NEWTON, M. RUDNEV, AND A. WARREN, *An Energy Bound in the Affine Group*, International Mathematics Research Notices, (2020).
rnaa130.

[122] N. D. PHUONG, P. V. THANG, AND L. A. VINH, *Incidences between points and generalized spheres over finite fields and related problems*, 2016.
arXiv:1410.7899 math.CO.

[123] H. PLÜNNECKE, *Eine zahlentheoretische anwendung der graphentheorie.*, Journal für die reine und angewandte Mathematik (Crelles Journal), 1970 (1970), pp. 171–183.

[124] M. H. POINCARÉ, *Sur la dynamique de l'électron*, Rendiconti del Circolo Matematico di Palermo (1884-1940), 21 (1906), pp. 129–175.

[125] L. PYBER AND E. SZABÓ, *Growth in linear groups*, in Thin groups and superstrong approximation, vol. 61 of Math. Sci. Res. Inst. Publ., Cambridge Univ. Press, Cambridge, 2014, pp. 253–268.

[126] A. A. RAZBOROV, *A product theorem in free groups*, Annals of Mathematics, 179 (2014), pp. 405–429.

[127] O. ROCHE-NEWTON AND M. RUDNEV, *On the Minkowski distances and products of sum sets*, Israel J. Math., 209 (2015), pp. 507–526.

[128] O. ROCHE-NEWTON, M. RUDNEV, AND I. SHKREDOV, *New sum-product type estimates over finite fields*, Advances in Mathematics, 293 (2016), pp. 589–605.

[129] O. ROCHE-NEWTON AND A. WARREN, *New expander bounds from affine group energy*, Discrete & Computational Geometry, (2019), pp. 1–23.

[130] M. RUDNEV, *An improved sum-product inequality in fields of prime order*, International Mathematics Research Notices, 2012 (2010).

[131] ——, *On distinct cross-ratios and related growth problems*, Moscow Journal of Combinatorics and Number Theory„ 7 (2017), pp. 51—-65.

[132] ——, *On the number of incidences between points and planes in three dimensions*, Combinatorica, 38 (2018), pp. 219–254.

[133] ——, *On the number of incidences between points and planes in three dimensions*, Combinatorica, 38 (2018), pp. 219–254.

[134] ——, *Point-plane incidences and some applications in positive characteristic*, Combinatorics and Finite Fields, (2019), pp. 211–240.

[135] M. RUDNEV, *On incidences of lines in regular complexes*, 2020.
arXiv:2003.04744 math.CO.

[136] M. RUDNEV, G. SHAKAN, AND I. SHKREDOV, *Stronger sum-product inequalities for small sets*, Proceedings of the American Mathematical Society, 148 (2019), p. 1.

[137] M. RUDNEV AND I. SHKREDOV, *On growth rate in $SL_2(\mathbf{F}_p)$, the affine group and sum-product type implications*, 2018.
arXiv:1812.01671 math.CO.

[138] M. RUDNEV, I. SHKREDOV, AND S. STEVENS, *On the energy variant of the sum-product conjecture*, Revista Matemática Iberoamericana, 36 (2016).

[139] M. RUDNEV AND S. STEVENS, *An update on the sum-product problem*, 2020.
arXiv:2005.11145 math.NT.

[140] M. RUDNEV AND J. WHEELER, *Incidence bounds with Möbius hyperbolae in positive characteristic*, 2021.
arXiv:2104.10534 math.CO.

[141] I. RUZSA, *Addendum to: An application of graph theory to additive number theory*, Scientia. Series A: Mathematical Sciences. New Series, 3 (1989).

[142] ——, *Generalized arithmetical progressions and sumsets*, Acta Math. Hungar., 65 (1994), pp. 379–388.

[143] ——, *Sums of finite sets*, Number Theory: New York Seminar 1991–1995, (1996).

[144] S. SAFIN, *Powers of sets in free groups*, Russian Academy of Sciences Sbornik Mathematics, 202 (2010).

[145] C. SCHADE, *Exakte Maximalzahlen gleicher Abstände*, diploma thesis, Techn. Univ. Braunschwei, 1993.

[146] T. SCHOEN AND I. SHKREDOV, *Higher moments of convolutions*, Journal of Number Theory, 133 (2011).

[147] G. SHAKAN, *On higher energy decompositions and the sum–product phenomenon*, Mathematical Proceedings of the Cambridge Philosophical Society, 167 (2019), p. 599–617.

[148] G. SHAKAN AND I. SHKREDOV, *Breaking the 6/5 threshold for sums and products modulo a prime*, 2018.
arXiv:1806.07091v1 math.CO.

[149] M. SHARIR AND N. SOLOMON, *Incidences between points on a variety and planes in* $\mathbb{R}^3$, 2017.
arXiv:1603.04823 math.CO.

[150] M. SHARIR AND J. ZAHL, *Cutting algebraic curves into pseudo-segments and applications*, Journal of Combinatorial Theory, Series A, 150 (2016).

[151] I. SHKREDOV, *Some new results on higher energies*, Transactions of the Moscow Mathematical Society, 74 (2012).

[152] ——, *An introduction to higher energies and sumsets*, arXiv e-prints 1512.00627, (2015).
arXiv:1512.00627 math.CO.

[153] ——, *On sums of Szemerédi-Trotter sets*, Proceedings of the Steklov Institute of Mathematics, 289 (2015), pp. 300–309.

[154] ——, *Some remarks on the asymmetric sum–product phenomenon*, Moscow Journal of Combinatorics and Number Theory, 8 (2017).

[155] ——, *On asymptotic formulae in some sum-product questions*, Trans. Moscow Math. Soc., 79 (2018), pp. 231–281.

[156] ——, *Some remarks on products of sets in the Heisenberg group and in the affine group*, Forum Math., 32 (2020), pp. 189–199.

[157] ——, *Modular hyperbolas and bilinear forms of Kloosterman sums*, J. Number Theory, 220 (2021), pp. 182–211.

[158] ——, *On an application of higher energies to sidon sets*, 2021. arXiv:2103.14670 math.NT.

[159] I. SHPARLINSKI AND T. ZHANG, *Cancellations amongst Kloosterman sums*, Acta Arithmetica, 176 (2016).

[160] J. SOLYMOSI, *On sum-sets and product-sets of complex numbers*, Journal de Théorie des Nombres de Bordeaux, 17 (2005).

[161] ——, *On the number of sums and products*, Bulletin of the London Mathematical Society, 37 (2005), pp. 491 – 494.

[162] ——, *Bounding multiplicative energy by the sumset*, Advances in Mathematics, 222 (2009), pp. 402–408.

[163] J. SOLYMOSI AND T. TAO, *An incidence theorem in higher dimensions*, Discrete & Computational Geometry, 48 (2011).

[164] J. SOLYMOSI AND G. TARDOS, *On the number of k-rich transformations*, in Computational geometry (SCG'07), ACM, New York, 2007, pp. 227–231.

[165] J. SOLYMOSI AND C. TÓTH, *Distinct distances in the plane*, Discrete and Computational Geometry, 25 (2001), pp. 629–634.

[166] J. SOLYMOSI AND V. VU, *Near optimal bounds for the Erdős distinct distances problem in high dimensions*, Combinatorica, 28 (2008), pp. 113–125.

[167] ——, *Sum-product estimates for well-conditioned matrices*, Bulletin of The London Mathematical Society - BULL LOND MATH SOC, 41 (2009), pp. 817–822.

[168] J. SOLYMOSI AND C. WONG, *An application of kissing number in sum-product estimates*, Acta Mathematica Hungarica, 155 (2017).

[169] J. SPENCER, E. SZEMERÉDI, AND W. TROTTER, *Unit distances in the Euclidean plane*, Graph Theory and Combinatorics, (1984), pp. 293–303.

[170] S. STEVENS, *Incidence Geometry in the Plane and Applications to Arithmetic Combinatorics*, PhD thesis, University of Bristol, 2019.

[171] S. STEVENS AND A. WARREN, *On sum sets of convex functions*, 2021. arXiv:2102.05446 math.CO.

[172] S. STEVENS AND F. ZEEUW, *An improved point-line incidence bound over arbitrary fields: An improved point-line incidence bound*, Bulletin of the London Mathematical Society, 49 (2017).

[173] E. STUDY, *Geometrie der dynamen. Die zusammensetzung von kräften und verwandte gegenstände der geometrie bearb.*, B.G. Teubner, Leipzig, 1903.

[174] L. SZÉKELY, *Crossing numbers and hard Erdős problems in discrete geometry*, Combinatorics, Probability and Computing, 6 (1997), pp. 353 – 358.

[175] ——, *Crossing numbers and hard Erdős problems in discrete geometry*, Combinatorics, Probability and Computing, 6 (1997), pp. 353 – 358.

[176] E. SZEMERÉDI AND W. TROTTER, *Extremal problems in discrete geometry*, Combinatorica, 3 (1983), pp. 381–392.

[177] T. TAO, *Product set estimates for non-commutative groups*, Combinatorica, 28 (2008), pp. 547–594.

[178] ——, *The sum-product phenomenon in arbitrary rings*, Contributions to Discrete Mathematics, 4 (2008).

[179] ——, *Freiman's theorem for solvable groups*, Contrib. Discrete Math., 5 (2009).

[180] ——, *The sum-product phenomenon in arbitrary rings*, Contrib. Discrete Math., 4 (2009), pp. 59–82.

[181] ——, *The Szemerédi-Trotter theorem and the cell decomposition*, Jun 2009. Blog post.

[182] ——, *Notes on simple groups of Lie type*, Sep 2013. Blog post.

[183] ——, *Notes on the classification of complex Lie algebras*, April 2013. Blog post.

[184] ——, *Expansion in Finite Simple Groups of Lie Type*, American Mathematical Society, 2015.

[185] T. TAO AND V. H. VU, *Additive Combinatorics*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2006.

[186] G. TARDOS, *On distinct sums and distinct distances*, Advances in Mathematics - ADVAN MATH, 180 (2001).

[187] G. TENENBAUM, *Sur la probabilité qu'un entier possède un diviseur dans un intervalle donné*, Compositio Mathematica, 51 (1984).

[188] P. THANG AND L. VINH, *Erdős–Rényi graph, Szemerédi–Trotter type theorem, and sum-product estimates over finite rings*, Forum Mathematicum, 0 (2012).

[189] M. C. H. TOINTON, *Freiman's theorem in an arbitrary nilpotent group*, Proc. Lond. Math. Soc. (3), 109 (2014), pp. 318–352.

[190] ——, *Introduction to approximate groups*, vol. 94 of London Mathematical Society Student Texts, Cambridge University Press, Cambridge, 2020.

[191] C. TÓTH, *The Szemerédi-Trotter theorem in the complex plane*, Combinatorica, 35 (2003).

[192] H. UĞURLU AND A. ÇALIŞKAN, *The Study mapping for directed space-like and time-like lines in Minkowski 3-space* $\mathbb{R}^3_1$, Mathematical & Computational Applications, 1 (1996).

[193] V. VARADARAJAN AND J. VIRTANEN, *Structure, classification, and conformal symmetry, of elementary particles over non-archimedean space-time*, Letters in Mathematical Physics, 89 (2010).

[194] L. VINH, *The Szemerédi–Trotter type theorem and the sum-product estimate in finite fields*, European Journal of Combinatorics, 32 (2011), pp. 1177–1181.

[195] ——, *On point-line incidences in vector spaces over finite fields*, Discrete Applied Mathematics, 177 (2014), p. 146–151.

[196] A. WALFISZ, *Weylsche Exponentialsummen in der neueren Zahlentheorie*. Mathematische Forschungsberichte. 16. Berlin: VEB Deutscher Verlag der Wissenschaften. 231 S. (1963)., 1963.

[197] A. WARREN AND J. WHEELER, *Incidences of Möbius transformations in* $\mathbb{F}_p$, 2021. arXiv:2107.12286 math.CO.

[198] T. WOLFF, *Recent work connected with the Kakeya problem*, in Prospects in mathematics (Princeton, NJ, 1996, pp. 129–162.

[199] ——, *Decay of circular means of Fourier transforms of measures*, International Mathematics Research Notices, 1999 (1999), pp. 547–567.

[200] E. YAZICI, B. MURPHY, M. RUDNEV, AND I. SHKREDOV, *Growth estimates in positive characteristic via collisions*, International Mathematics Research Notices, 2017 (2015).

[201] J. ZAHL, *An improved bound on the number of point-surface incidences in three dimensions*, Contributions to Discrete Mathematics, 8 (2011).

[202] ——, *A Szemerédi–Trotter type theorem in* $\mathbb{R}^4$, Discrete & Computational Geometry, 54 (2015).

[203] ——, *Breaking the 3/2 barrier for unit distances in three dimensions*, International Mathematics Research Notices, 2019 (2017).

[204] ——, *Sphere tangencies, line incidences and Lie's line-sphere correspondence*, Mathematical Proceedings of the Cambridge Philosophical Society, (2021), p. 1–21.