SOME ARITHMETIC RAMSEY PROBLEMS AND INVERSE THEOREMS

BY

YIFAN JING

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Mathematics
in the Graduate College of the
University of Illinois Urbana-Champaign, 2021

Urbana, Illinois

Doctoral Committee:

Professor Kevin Ford, Chair
Professor József Balogh, Co-Director of Research
Professor Xiaochun Li, Co-Director of Research
Professor Lou van den Dries
Research Assistant Professor Sean English

Abstract

In this dissertation we study arithmetic Ramsey type problems and inverse problems, in various settings. This work consists of two parts.

In Part I, we study arithmetic Ramsey type problems over abelian groups. This part consists of three chapters. In Chapter 2, using hypergraph containers, we study the rainbow Erdős–Rothschild problem for sum-free sets. This is joint work with Cheng, Li, Wang, and Zhou. In Chapters 3 and 4, we study the avoidance density for (k, ℓ)-sum-free sets. The upper bound constructions are given in Chapter 3, answering a question asked by Bajnok. We also improved the lower bound for infinitely many (k, ℓ) in both chapters, and a special case of the sum-free conjecture is verified in Chapter 4. These two chapters are based on joint work with Wu.

In Part II, we study inverse problems over nonabelian topological groups. Preliminaries to topological groups are given in Chapter 5. In Chapter 6, we first obtain classifications of connected groups and sets which satisfy the equality in Kemperman's inequality, answering a question asked by Kemperman in 1964. When the ambient group is compact, we also get a near equality version of the above result with a sharp exponent bound, which confirms conjectures by Griesmer and by Tao. A measure expansion gap result for simple Lie groups is also presented. This chapter is based on joint work with Tran. In Chapter 7, we study the small measure expansion problem in noncompact locally compact groups. The question that whether there is a Brunn–Minkowski inequality was asked by Henstock and Macbeath in 1953. We obtain such an inequality and prove it is sharp for a large class of groups (including real linear algebraic groups, Nash groups, semisimple Lie groups with finite center, solvable

Lie groups, etc), answering questions by Hrushovski and by Tao. This chapter is based on joint work with Tran and Zhang.

This dissertation is based on the following papers and preprints: [41, 108, 107] (Part I), and [105, 106] (Part II).

To my family and friends.

Acknowledgments

First, I express my deepest gratitude to my advisers, József Balogh and Xiaochun Li. None of this work could have been imaginable without their enduring support and encouragement, and I am indebted to both of them for their guidance and for all the lessons that I was fortunate to receive from them. Their dedication to mathematics is the continuing motivation for me to better myself as a mathematician. I would also like to thank my Master thesis advisor Bojan Mohar from Simon Fraser University, and my Bachelor thesis advisor Jack Koolen from University of Science and Technology of China, for all the help and guidance that they have given me.

I thank my other collaborators, Jinpeng An, Yangyang Cheng, Béla Csaba, Shengwen Gan, Gennian Ge, Sebastián González Hermosillo de la Maza, Alexandr Kostochka, Lina Li, Fuhong Ma, Masood Masjoody, András Pluhár, Zhi Qiao, Akbar Rafiey, Souktik Roy, Pongpat Sittitrai, Guanghui Wang, Jingwei Xu, Zixiang Xu, Tao Zhang, and Wenling Zhou, for making the pursuit of mathematics more lively.

Special thanks to Chieu-Minh Tran, Shukun Wu, and Ruixiang Zhang, for helping me to learn model theory, harmonic analysis, and Lie theory; they have been great collaborators and friends, there is always a lot of fun to discuss mathematics with them.

I had helpful conversations with many mathematicians about the content of this thesis and about career. For that, I would like to thank Alessio Figalli, Kyle Gannon, Ben Green, John Griesmer, Zilin Jiang, Hong Liu, Jie Ma, Cosmin Pohoata, Anand Pillay, George Shakan, Terence Tao, Shi Wang, Hehui Wu, Max Wenqiang Xu, and Frank de Zeeuw. I also thank my other committee members, Lou van den Dries, Kevin Ford and Sean English for the work and time involved.

In the last three years, I have made a number of really good friends among my fellow graduate students, and I sincerely thank them for making my time at UIUC enjoyable and fulfilling. I extend special thanks to Yang Bai, Xujun Liu, Peixue Wu, Qiang Wu, and Siang Zhou, for chatting with me regularly during the pandemic period.

Finally, I would like to thank my family, especially my parents Fuchun Jing and Wenping Zhang, for their love and support. Special thanks to Sijia Liu, for her accompany through the years.

This work was supported in part by the Wind Information Fellowship, University Fellowship, the Waldemar J., Barbara G., and Juliette Alexandra Trjitzinsky Fellowship, and by Arnold O. Beckman Research Award (Campus Research Board RB21011).

Table of Contents

Symbols and Notation . . . . **x**

Chapter 1 Introduction **1**

1.1 Background 1

1.1.1 Arithmetic Ramsey theorems 2

1.1.2 Inverse theorems 4

1.2 Overview of the thesis 6

I The Abelian Groups 14

Chapter 2 Integer colorings with forbidden rainbow sums **15**

2.1 Introduction 15

2.1.1 Erdős–Rothschild problems in various settings 15

2.1.2 Erdős–Rothschild problems for sum-free sets 17

2.1.3 Our results 19

2.2 Notation and preliminaries 22

2.2.1 Basic properties of restricted sum-free sets 22

2.2.2 Structural properties of sum-free sets 23

2.2.3 Multi-color container theorem 24

2.3 Proof of Theorem 2.5 27

2.4 Proof of Theorems 2.7 and 2.8 32

2.5 Proof of Theorem 2.10 41

2.6 Concluding Remarks 45

Chapter 3 Avoidance density for (k, ℓ)-sum-free sets **47**

3.1 Introduction 47

3.2 Preliminaries 51

3.2.1 Additive combinatorics 51

3.2.2 Nonstandard analysis 53

3.2.3 Determinants of certain matrices 54

3.3 A density estimate 56

3.4 (k, ℓ)-sum-free open sets in the torus 61

3.5 Lower Bounds 63

3.5.1 Proof of Theorem 3.3 (ii) 66

3.5.2 Proof of Theorem 3.3 (iii) 70
3.6 Structure of infinite (k, ℓ)-sum-free sets 73
3.7 Upper bound constructions 78
3.8 Restricted (k, ℓ)-sum-free sets 80
3.9 Concluding Remarks 86

Chapter 4 A closer look to the largest sum-free sets 89
4.1 Introduction 89
4.2 When A is geometric 92
4.3 When A is not geometric 97

II The Nonabelian Groups 104

Chapter 5 Preliminaries to topological groups 105
5.1 Some results about topological groups 105
5.2 Measures and the modular function 107
5.3 Almost-Lie groups and the Gleason–Yamabe Theorem 109
5.4 Some results about Lie groups 110
5.5 Solvable and Semisimple Lie groups 112

Chapter 6 Minimal and nearly minimal measure expansions in connected unimodular groups 117
6.1 Introduction 118
6.1.1 Background 118
6.1.2 Statement of main results 123
6.1.3 Notation and convention 126
6.2 Outline of the argument 126
6.2.1 Overview of the strategy 126
6.2.2 The first step: Obtaining a suitable pseudometric 131
6.2.3 The second step: Constructing an appropriate homomorphism 134
6.2.4 The third step: Deducing the structure of the pair 137
6.2.5 Structure of the chapter 138
6.3 Preliminaries 140
6.3.1 Locally compact groups and Haar measures 140
6.3.2 More on Kemperman's inequality and the inverse problem 144
6.4 Reduction to nearly minimal expansion pairs 147
6.5 Reduction to sets with small measure 148
6.6 Geometry of minimal and nearly minimal expansion pairs I 153
6.6.1 Preservation of minimal expansion under quotient 153
6.6.2 Coarse versions of the main theorems 160
6.6.3 Structure control on the nearly minimal expansion sets 162
6.7 Pseudometrics and group homomorphisms into tori 168
6.7.1 Linear pseudometrics 168
6.7.2 Almost linear pseudometrics: relative sign and total weight functions 175

6.7.3 Almost linear pseudometrics: group homomorphisms into tori 184
6.8 Geometry of minimal and nearly minimal expansion pairs II 196
6.8.1 Toric transversal intersection in measure 197
6.8.2 Linear pseudometric from minimal expansions 208
6.8.3 Almost linear pseudometric from near minimal expansions 216
6.9 Proof of the main theorems 226
6.9.1 Minimal expansion pairs in noncompact groups 226
6.9.2 Nearly minimal expansion pairs in compact groups 229

Chapter 7 A nonabelian Brunn–Minkowski inequality 235
7.1 Introduction 235
7.1.1 Background 235
7.1.2 Statement of main results 238
7.1.3 Overview of the proof 242
7.2 Noncompact Lie dimension and helix dimension 246
7.3 Proof of Theorem 7.3 261
7.4 Reduction to outer terms of certain short exact sequences 268
7.5 Reduction to unimodular subgroups 275
7.6 Reduction to cocompact and codiscrete subgroups 281
7.7 Proof of Theorems 7.1, 7.2, and 7.5 285
7.7.1 A dichotomy lemma 285
7.7.2 Proofs of the main theorems 287

References 292

Symbols and Notation

$\varnothing$	the empty set
$\mathbb{N}$	the set of natural numbers
$\mathbb{Z}$	the set of integers
$\mathbb{R}^d$	the d dimensional Euclidean space
$\mathbb{T}^d$	$(\mathbb{R}/\mathbb{Z})^d$
$\mathbb{C}$	the complex field
$\mathbb{F}_p^n$	finite field of order p^n
$\mathbb{F}_p^*$	$\mathbb{F}_p \setminus \{0\}$
$\mathbb{Z}/p\mathbb{Z}$	the cyclic group of order p
${}^*\mathbb{R}$	the hyperreal field
${}^*\mathbb{Z}$	the hyperinteger ring
$K[x]$	the polynomial ring over a field K
$[n]$	for $n \in \mathbb{N}$, $[n] := [1, n] \cap \mathbb{Z}$
log	logarithm base 2
$V(G)$	the vertex set of a (hyper)graph G
$v(G)$	$\lvert V(G) \rvert$
$E(G)$	the edge set of a (hyper)graph G
$e(G)$	$\lvert E(G) \rvert$
G_0	the identity component of a group G
$[G, G]$, $G^{(1)}$	the derived subgroup of G

$G^{(k)}$	for $k \in \mathbb{N}$, the k-th derived subgroup of G
$N_G(v),\ N(v)$	the neighborhood of a vertex v in a graph G
$G \,\square\, H$	the Cartesian product of two graphs G and H
$H \cong G$	H is isomorphic to G as topological groups
$H \triangleleft G$	H is a normal subgroup of G
$N \rtimes H$	the semidirect product of N and H
id_G	the identity element of a group G
μ	the Möbius function
μ_G	the left Haar measure on a locally compact group G
ν_G	the right Haar measure on a locally compact group G
$d_G(v),\ d(v)$	the degree of a vertex v in a graph G
$\overline{d}(A)$	the upper density of a set A
$\delta(G)$	the minimum degree of a graph G
$\Delta(G)$	the maximum degree of a graph G
Δ_G	the modular function of a locally compact group G
$C_c(G)$	the space of compactly supported continuous functions over G
$\mathbb{1}_A$	the characteristic function of a set A
$e(\theta)$	$e^{2\pi i\theta}$
$e_N(\theta)$	$e^{\frac{2\pi i\theta}{N}}$
$\mathrm{ex}(n, H)$	the maximum number of edges in an n vertex graph that contains no subgraph isomorphic to H
$\exp_G,\ \exp$	the exponential map from a Lie algebra $\mathfrak{g}$ to a Lie group G
$\widehat{f}$	the Fourier coefficient of a function f
$f = O(g)$	$f \le Cg$ for some constant C
$f \ll g,\ g = \Omega(f)$	$f = O(g)$
$f \asymp g,\ f = \Theta(g)$	$f \ll g$ and $g \ll f$
$f = o(g)$	$\lim_{n\to\infty} f(n)/g(n) = 0$

$f \sim g$	$\lim_{n\to\infty} f(n)/g(n) = 1$
K_n	the complete graph on n vertices
$K_{s,t}$	the complete bipartite graph with parts of size s and t
C_n	the cycle on n vertices
P_n	the path on n vertices
$X + Y$	the sumset of X and Y, i.e., $\{x + y : x \in X, y \in Y\}$
kX	for $k \in \mathbb{N}$, the k-fold sumset of X
$k \cdot X$	dilation of X with factor k, i.e., $\{kx : x \in X\}$
XY	the productset of X and Y, i.e., $\{xy : x \in X, y \in Y\}$.
X^k	for $k \in \mathbb{N}$, the k-fold productset of X
X^{-1}	the inverse set of X, i.e., $\{x^{-1} : x \in X\}$
$P(X, Y)$	$\{P(x, y) : x \in X, y \in Y\}$ for some bivarite polynomial P
$X \triangle Y$	the symmetric difference of two sets X and Y
$Z(G)$	the center of a group G
$\langle X \rangle$	the smallest subgroup generated by X
$\overline{U}$	the closure of a set U
$(\mathbb{R}^{>0}, \times)$	topological group $\mathbb{R}^{>0}$ with multiplication as the group operation
$\mathrm{GL}(n, \mathbb{R})$	the real general linear group of degree n, i.e., the group containing $n \times n$ real invertible matrices
$\mathrm{SL}(n, \mathbb{R})$	the real special linear group of degree n, i.e., the group containing $n \times n$ real matrices with determinant 1
$\mathrm{SO}(n, \mathbb{R})$	the real special orthogonal group of degree n, i.e., the group containing $n \times n$ real orthogonal matrices with determinant 1

Chapter 1

Introduction

This thesis is a fusion of my research papers and preprints [41, 105, 106, 107, 108]. They share a common theme considering problems over some algebraic or geometric structures with certain extremal flavor, including Ramsey-type problems and inverse problems.

In this introduction, I would like to present some background of arithmetic Ramsey problems and inverse problems briefly, and why it might be interesting to explore these problems. Afterward, I will go into a more detailed description of the structure of this thesis and the main results in each chapter.

1.1 Background

The following questions arise naturally in many areas of mathematics: Let S be a certain structure (for examples, complete graphs, groups, fields), and let P be a property. How big must some substructure S' of S be to guarantee that the property P holds in S'? What can be said about the structures of S' when the property P holds or nearly holds in it?

The first type of problem refers to *Ramsey problem*, or arithmetic Ramsey problem when S has some arithmetic constraints. The second problem asks about the structural characterizations of the substructures when they have maximal (or nearly maximal) size and do not have property P; this type of problems are known as *inverse problems*.

1.1.1 Arithmetic Ramsey theorems

Let us begin the story with Ramsey-type theorems. The celebrated Ramsey Theorem [145] asserts that any finitely colored sufficiently large complete graph will contain a large monochromatic complete subgraph. This result created a new species of mathematical result — the Ramsey type theorems, and also represented the discovery of a new phenomenon in mathematics that "complete disorder is impossible".

One of the first arithmetic Ramsey results was van der Waerden's Theorem [165]: given any finite coloring of integers, one of the color classes must contain arbitrarily long arithmetic progressions. Van der Waerden's orginal proof is beautiful and elegant, but only offers poor quantitative bounds for the appearance of the first arithmetic progression of a given length. Erdős and Turán [63] then made a couple of conjectures to pursue this quantitative question further.

The first major progress of the conjectures was by Roth [148], who applied the Hardy–Littlewood circle method together with the density increment argument to establish Roth's Theorem:

Theorem 1.1 (Roth's Theorem). *If $A \subseteq [N]$ contains no nontrivial three term progressions, then $|A| \ll \frac{N}{\log\log N}$.*

On the other direction, Behrend [11] constructed a subset of $[N]$ of size $Ne^{-O(\sqrt{\log N})}$ that contains no nontrivial three term arithmetic progressions. Quantitative improvements of Roth's theorem were later obtained by a series of important papers by Bourgain [29], by Sanders [152], by Bloom [20], by Bloom and Sisask [21], and by Schoen [155]. A recent breakthrough by Bloom and Sisask [22] shows that any subset of $[N]$ of size $N/(\log N)^{1+c}$ contain a three term arithmetic progression. This confirms the first nontrivial case of the well-known Erdős sum of reciprocals conjecture.

For longer arithmetic progressions, using purely combinatorial techniques, Szemerédi established the celebrated Szemerédi's Theorem [158], which asserts that any dense subset of integers contain an arbitrarily long arithmetic progression. This confirms a conjecture by Erdős and Turán mentioned earlier in the section. One of the new ingredients in Szemerédi's proof was later formulated as Szemerédi's regularity lemma, and has became one of the most important tools in extremal combinatorics.

Szemerédi's proof required van der Waerden's Theorem, so it did not give any improved quantitative bound on that theorem. Thus, after Szemerédi proved his theorem, mathematicians were seeking to understand, reprove, and improve upon Szemerédi's Theorem in other ways. Furstenberg insightfully observed that Szemerédi's Theorem is equivalent to a multiple recurrence theorem for measure-preserving systems, and then he was able to prove Szemerédi's Theorem using ergodic theory [70]. Mathematicians then realized this new method is very powerful, and could be used to prove some other arithmetic Ramsey-type results, for instance the density Hales–Jewett Theorem [71], and the polynomial Szemerédi Theorem [14].

In parallel to these ergodic theory developments, the graph removal lemma, a breakthrough in graph theory was obtained Ruzsa and Szemerédi [149] by using the regularity lemma. They also observed that the triangle removal lemma implies Roth's Theorem (with worse quantitative bounds). This motivated the program for finding satisfactory analogue of the regularity lemma and the counting lemma for hypergraphs, and to prove Szemerédi's Theorem using purely graph theoretical techniques. These results were eventually proved by Gowers [75], and independently by Nagle, Rödl, Schacht and Skokan [141, 146].

The Fourier analytic approach, the method firstly used by Roth for the case of three term progressions, was finally revisited by Gowers [74]. In the proof, he introduced a new notion of uniformity, now known as the Gowers uniformity norms, which has many other applications and has became a standard tool in arithmetic combinatorics. It is worth noting that the Fourier analytic approach by Gowers obtained remarkably strong quantitative bounds on Szemerédi's Theorem:

Theorem 1.2 (Gowers)**.** *If $A \subseteq [N]$ contains no nontrivial k-term progressions, then there* *is* $c_k > 0$ *such that* $|A| \leq \frac{N}{(\log\log N)^{c_k}}$.

Szemerédi's Theorem studies the size of the largest subset of integers that avoids long arithmetic progressions. Although this result is possibly the most famous arithmetic Ramsey-type theorem, many other Ramsey-type problems considering avoiding patterns other than arithmetic progressions or inside other ambient structures are also well-studied. In this thesis, we will mainly focus on Schur triples, and more generally, (k, ℓ)-sums, under different settings. These results are included in Chapters 2, 3 and 4.

1.1.2 Inverse theorems

In an inverse problem, we usually begin with the sumsets (or product sets) and try to deduce information about the structure of underlying sets. For instance, suppose the ambient group is $\mathbb{Z}$, then for any nonempty set $A \subseteq \mathbb{Z}$, one can easily see that $|A+A| \geq 2|A|-1$. Freiman's Theorem gives us the additive structure of A when we know that $|A+A|$ is small:

Theorem 1.3 (Freiman's Theorem, qualitative form). *Let* $A \subseteq \mathbb{Z}$ *be nonempty. Suppose* $|A+A| \leq K|A|$ *for some constant* K, *then* A *is contained in a generalized arithmetic progression of dimension at most* $d(K)$ *and length at most* $f(K)|A|$.

Freiman's Theorem was later generalized to arbitrary abelian groups by Green and Ruzsa [80]. In their theorem, a similar conclusion still holds by replacing generalized arithmetic progression by coset progression. Generalized arithmetic progressions, and more generally, coset progressions, can be seen as a subgroup "up to a constant error" (in fact, they are *approximate groups*). Thus, intuitively, these results informally tell us, for a subset of an abelian group, if it has small cardinality expansion (that is $|A+A|/|A|$ is small), then it should look like a "subgroup" of the ambient abelian group.

The famous sum-product theorem, proved by Erdős and Szemerédi [62], asserts that a subset of a field cannot have small sumset expansion and small product set expansion simultaneously. This suggests that finite subsets of a one-dimensional space have expansion behavior under binary operations unless the situation is "controlled" by a *single abelian* group. This can also be seen from the celebrated Elekes–Rónyai Theorem [58]: A nontrivial bivarite polynomial $P(x, y)$ exhibits small expansion (that is $|P(A, A)|/|A|$ is small), it must exploit the underlying additive structure or multiplicative structure of the field (a more precised structural result was recently obtained by [104]).

For possibly nonabelian groups, Gromov's theorem on groups of polynomial growth [83] tells us that if A is finite, and there is a polynomial P such that $|A^n| < P(n)$ for every $n \in \mathbb{N}^{\geq 1}$, then $\langle A \rangle$ is virtually nilpotent. This result suggests that the small expansion property in nonabelian groups will also give us some structural information for the ambient group. This phenomenon can be also seen from a recently breakthrough by Breuillard, Green, and Tao [36]:

Theorem 1.4 (Breuillard–Green–Tao, simple form)**.** *If $A \subseteq G$ is a finite K-approximate group, then there is a $K^{O(1)}$-approximate group $A' \subseteq A^4$, such that:*

(i) *A is covered by $L(K)$ left cosets of A';*

(ii) *$\langle A' \rangle$ has a d-nilpotent subgroup of finite index, with $d \ll \log K$.*

Here, A is a K-*approximate group* means $A = A^{-1}$, and A^2 can be covered by K left cosets of A. It is clear that being a K-approximate group is a stronger condition than having small cardinality expansions (for instance, if A is a singleton union a subgroup, while $|A^2|/|A|$ is at most 4, $|A^3|/|A|$ can be arbitrarily large), a result by Tao [159] shows that a small expansion set is essentially an approximate group (by removing some elements). The Breuillard–Green–Tao Theorem can be seen as a spiritual generalization of Freiman's theorem in nonabelian groups, and Gromov's theorem for approximate groups. Their proof was built upon earlier work of Hrushovski [102] and employ tools from model theory.

The structures of "small expansion sets" are also well studied for some other notion of size. In this thesis, we will mainly focus on the case when the ambient group is a locally compact group G equipped with a Haar measure μ, and our set $A \subseteq G$ is a measurable set with finite positive measure. In this context, by small expansion, or small measure expansion, we mean that $\mu(A^2)/\mu(A)$ is small. We are interested in how A and G behave when we know that A has small measure expansion in G. For more background about small measure expansion sets, we refer to Section 6.1.1.

1.2 Overview of the thesis

In this section we give a brief chapter-by-chapter overview of this work.

Integer colorings with forbidden rainbow sums

The Erdős–Rothschild extension for sum-free sets has been recently pursued by Liu, Sharifzadeh and Staden [124] for subsets of the integers, and Hàn and Jiménez [87] for finite abelian groups. More specifically, they investigated the extremal configurations which maximize the number of sum-free r-colorings, where each of the color classes is a sum-free set, for small r. The characterization of extremal sets for $r \geq 3$ remains widely open.

We consider a rainbow variant of the Erdős–Rothschild problem for sum-free sets on integers. For a set $A \subseteq [n]$, an r-coloring of A is *rainbow sum-free* if it contains no rainbow ordered triple (a, b, c) with $a < b < c$ and $a + b = c$. For an integer $r \geq 1$ and a set $A \subseteq [n]$, we write $g(A, r)$ for the number of rainbow sum free r-colorings of A and define $g(n, r) := \max_{A \subseteq [n]} g(A, r)$. A set $A \subseteq [n]$ is *rainbow r-extremal* if $g(A, r) = g(n, r)$. When $r \in \{1, 2\}$, it is easy to see that the only extremal set is the interval $[n]$. For more colors, the problem becomes considerably more complicated. For odd n, we define $I_1 = \left\{\frac{n-1}{2}, \frac{n-1}{2} + 1, \dots, n-1, n\right\}$. For even n, we define $I_2 = \left\{\frac{n}{2} - 1, \frac{n}{2}, \dots, n\right\}$, $I_3 = \left\{\frac{n}{2}, \frac{n}{2} + 1, \dots, n\right\}$. We first made the following conjecture.

Conjecture 1.5. *Let n, r be positive integers and $r \geq 3$.*

(i) *For $r = 3$, the interval $[n]$ is the unique rainbow r-extremal set.*

(ii) *For odd n and $r = 4$, the interval $[n]$ is the unique rainbow r-extremal set.*

(iii) *For odd n and $r \geq 5$, the set I_1 is the unique rainbow r-extremal set.*

(iv) *For even n and $r \leq 7$, the set I_2 is the unique rainbow r-extremal set.*

(v) *For even n and $r \geq 8$, the set I_3 is the unique rainbow r-extremal set.*

We resolved Conjecture 1.5 for $r = 3$ and $r \geq 8$ when n is sufficiently large. For $4 \leq r \leq 7$, we asymptotically determine the logarithm of $g(n, r)$:

Theorem 1.6. *For $4 \leq r \leq 7$ and all positive integers n, we have $g(n, r) = r^{\lceil n/2 \rceil + o(n)}$.*

Like the Gallai coloring problem, we also provide a sharp bound on the number of rainbow sum free r-colorings of $[n]$, and this determines its typical structure: we showed that for every integer $r \geq 3$, almost all rainbow sum free r-colorings of $[n]$ are 2-colorings. The proof relies on the hypergraph container method, and some ad-hoc stability analysis of sum-free sets.

Avoidance density for (k, ℓ)-sum-free sets

In 1965, Erdős [59] asked the following question: Given a set of positive integers of cardinality N, what is the size of the maximal sum-free subset of it? A set is called *sum-free*, if for every three elements x, y, z we have $x + y \neq z$. For every pair of distinct positive integers (k, ℓ), a set A is (k, ℓ)-*sum-free* if for every $k + \ell$ elements $x_1, \ldots, x_k, y_1, \ldots, y_\ell$ in A, we always have $\sum_{i=1}^{k} x_i \neq \sum_{j=1}^{\ell} y_j$.

In a recent breakthrough, Eberhard, Green, and Manners [57] showed that the size of the maximal sum-free subset of a set of size N is at most $(1/3 + o(1))N$. The upper bound for (k, ℓ)-sum-free sets was asked by Bajnok recently. The case for $(1, k)$-sum-free sets was resolved by Eberhard [56]. With Wu [108], we determined the avoidance density for all (k, ℓ)-sum-free sets. More precisely, let

$$\mathcal{M}_{(k,\ell)}(N) = \inf_{\substack{A \subseteq \mathbb{N}^{>0} \\ |A| = N}} \mathcal{M}_{(k,\ell)}(A) \quad \text{where} \quad \mathcal{M}_{(k,\ell)}(A) = \max_{\substack{S \subseteq A \\ S \text{ is } (k,\ell)\text{-sum-free}}} |S|.$$

In Chapter 3, we showed that $\mathcal{M}_{(k,\ell)}(N) \leq \frac{N}{k+\ell} + o(N)$.

A more interesting problem is to consider the lower bound on $\mathcal{M}_{(k,\ell)}(N)$. For the sum-free case (i.e. $(k,\ell) = (2,1)$), the following conjecture is well-known:

Conjecture 1.7. *There is a function $\omega(N) \to \infty$ as $N \to \infty$, such that*

$$\mathcal{M}_{(2,1)}(N) > \frac{N}{3} + \omega(N).$$

The current best bound in this direction is obtained by Bourgain [28] using Fourier analytic argument, where he showed that $\mathcal{M}_{(2,1)}(N) \geq (N+2)/3$.

One can also ask an analogue question for (k,ℓ)-sum-free set, that whether there is a function $\omega(N) \to \infty$ as $N \to \infty$, such that $\mathcal{M}_{(k,\ell)}(N) \geq N/(k+\ell) + \omega(N)$. The case when $(k,\ell) = (3,1)$ is confirmed by Bourgain [28]. In the same chapter, by generalizing Bourgain's argument, we confirm this conjecture for some other infinite families of (k,ℓ):

Theorem 1.8. *Let k, ℓ be two positive integers and $k > \ell$. Then the following hold:*

(i) *suppose $k = 5\ell$. Then*

$$\mathcal{M}_{(k,\ell)}(N) \geq \frac{N}{k+\ell} + c\frac{\log N}{\log\log N}, \tag{1.1}$$

where $c > 0$ is an absolute constant that only depends on k, ℓ.

(ii) *for every set A of N positive integers, for every positive even integer u, there is an odd integer $v < u$ such that if $k = (u+v)\ell/(u-v)$, then*

$$\mathcal{M}_{(k,\ell)}(A) \geq \frac{N}{k+\ell} + c\frac{\log N}{\log\log N}, \tag{1.2}$$

where $c > 0$ is an absolute constant that only depends on k, ℓ.

Note that case (ii) in Theorem 1.8 covers the result for $(3,1)$-sum-free sets by Bourgain [28].

A closer look to the largest sum-free sets

Recall that $\mathbb{1}_\Omega$ denote the characteristic function of Ω, and $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ is the one dimensional torus. Conjecture 1.5 is generally attacked in the literature by considering another stronger conjecture:

Conjecture 1.9. *Let $\Omega = (1/3, 2/3) \subseteq \mathbb{T}$. Then when $N \to \infty$,*

$$\max_{x\in\mathbb{T}} \sum_{n\in A} (\mathbb{1}_\Omega - 1/3)(nx) \to \infty.$$

This conjecture, if true, would also imply that a similar phenomenon occurs for $(2k, 4k)$-sum-free sets for every $k \geq 1$. In Chapter 4, we prove the latter result directly:

Theorem 1.10. *For every $k \geq 1$, there is a function $\omega(N) = \log N/\log\log N$, such that for every set A of N positive integers, there exists a maximal $(2k, 4k)$-sum-free set $\Omega(2k, 4k) \subseteq \mathbb{T}$, and we have*

$$\max_{x\in\mathbb{T}} \sum_{n\in A} \left(\mathbb{1}_{\Omega(2k,4k)} - \frac{1}{6k}\right)(nx) \gg \omega(N).$$

In particular, there is an absolute constant $c > 0$, such that

$$\mathcal{M}_{(2k,4k)}(N) \geq \frac{N}{6k} + c\,\omega(N).$$

The new ingredient of our proof is a structural analysis on the host set A. It is inspired by the upper bound construction given in Chapter 3, the Følner sequence $\{F_n\}$, that is for every integer a one have

$$\frac{|F_n \triangle (a\cdot F_n)|}{|F_n|} \to 0.$$

We then split the proof into two cases, when $|A \triangle (a\cdot A)|$ is large (multiplicative case) and when $|A \triangle (a\cdot A)|$ is small (additive case), and use different techniques for these two cases.

Minimal and nearly minimal measure expansions in connected unimodular groups

Let G be a group equipped with a reasonable notion of size s (e.g. cardinality, Haar measure, density, etc). For $A \subseteq G$ and $n \in \mathbb{N}^{>0}$, set $A^n = \{a_1 \cdot \ldots \cdot a_n \mid a_i \in A\}$. We are interested in the following questions in different settings:

- For a subset A of G with finite $s(A)$ and $n \in \mathbb{N}^{\geq 2}$, what is the strict lower bound for $s(A^n)/s(A)$ (possibly under further assumptions of A)?
- What can be said about G and A, when $s(A^n)/s(A)$ has small values?

For either a fixed group or a class of groups, the extremal expansion problems ask for the the minimum expansion rate $s(A^2)/s(A)$ (possibly under extra assumptions on A) together with structural results for sets A where the minimum expansion rate is achieved or nearly achieved.

In Chapter 6, we considered the setting when G is connected and unimodular (locally compact with left Haar measures also right Haar measure), size is given by a Haar measure μ_G, and A ranges over compact subsets of G with $0 < \mu_G(A) < \mu_G(A^2) < \mu_G(G)$. Here, an inequality by Kemperman [112] in 1964 gives us a natural lower bound:

$$\mu_G(A^2)/\mu_G(A) \geq 2.$$

The question of classifying A such that the equality happened was proposed by Kemperman in the same paper. Before our work, only the special cases for abelian G were known, proven by Kneser in [115]. Griesmer considered a related question for disconnected abelian locally compact group in [81] and conjectured that the Kneser's answer also hold when G is nonabelian. In [162], Tao considered the problem of classifying A such that equality almost happens assuming further that G is compact and abelian, and conjectured that similar conclusions holds removing the abelian assumption. All these issues were resolved in this chapter. Our answer is that for the equality to happen (or almost happens), there must be a continuous group homomorphism $\chi : G \to \mathbb{T}$ such that A is (or is almost) an inverse image of an interval in $\mathbb{T}$.

Proving these results requires a new method which allow understanding the structure of G and A when A has small expansion. In more details, we first reduce the problem into Lie groups by appropriately using the Gleason–Yamabe Theorem [73, 169], and the rest of the proof goes by induction on dimension. In the second step, we choose a proper closed subgroup H such that the intersection of A with all its cosets are small, otherwise A contains a coset of every maximal torus, and sets with this "Kakeya-type" property cannot have very small measure expansion. In the third step, we deduce from the small expansion assumption the relative shape of A with respect to H, and show that this remains unchanged under translations by elements in a small neighborhood of identity. Finally, applying induction hypothesis on cosets of H and the preceding step we show that the map $g \mapsto \mu(A \setminus gA)$ is a pseudometric with a specific property yielding a group homomorphism into either $\mathbb{T}$ or $\mathbb{R}$. With this homomorphism and ideas from the third step applying to the kernel, we get the desired structure of A. A more detailed overview is given in Section 6.2.

A nonabelian Brunn–Minkowski inequality

The Brunn–Minkowski inequality [38, 136] provides us with a good lower bound for the measure of the sumset $X + Y$ of two nonempty compact subsets X and Y of $\mathbb{R}^d$.

Theorem 1.11 (The Brunn–Minkowski inequality)**.** *Suppose X, Y are nonempty compact subsets of $\mathbb{R}^d$, and let λ be the Lebesgue measure on $\mathbb{R}^d$. Then*

$$\lambda(X + Y)^{\frac{1}{d}} \geq \lambda(X)^{\frac{1}{d}} + \lambda(Y)^{\frac{1}{d}}.$$

In Chapter 7, we consider the problem of generalizing the Brunn–Minkowski inequality to all locally compact groups, a question suggested by Henstock and Macbeath in 1953 [90] and also later asked in different variations by many others, including Hrushovski [101], McCrudden [132], and Tao [160].

We first propose the following conjecture: For a possibly nonunimodular locally compact group G with a left Haar measure μ and right Haar measure ν, and nonempty compact $X, Y \subseteq G$, we expect

$$\frac{\nu(X)^{1/n}}{\nu(XY)^{1/n}} + \frac{\mu(Y)^{1/n}}{\mu(XY)^{1/n}} \leq 1$$

where n is the *noncompact Lie dimension* of G. For a Lie group G with dimension d and maximal dimension m of a compact subgroup, its noncompact Lie dimension is simply defined as $d - m$. The noncompact Lie dimension is defined for general locally compact groups through the Gleason–Yamabe Theorem [73, 169].

The above form of the Brunn–Minkowski inequality is new, and plays an important role in our proof. Our main results include:

- A proof that the conjectural inequalities are sharp if it holds for a given G.
- A Brunn–Minkowski type inequality for all locally compact groups, and a proof of the sharpness of the exponent for a large class of groups (including all real linear algebraic groups and, more generally, Lie groups definable in an o-minimal expansion of the field of real numbers).
- A reduction of the conjecture to the case of simply connected simple Lie groups.

We next highlight some ideas from our proof. To show sharpness, we can choose the example of a small neighborhood of a maximal compact subgroup. To get the inequality for a simple Lie group, we use the Iwasawa decomposition $G = KAN$, and obtain a lower bound for the measure expansion on G from that of AN using a "proportionated average" trick. The group AN is not necessarily unimodular, so it is crucial to treat the much more involved nonunimodular case as well. Finally, the reduction to simple Lie groups uses a generalization of McCrudden's "exponent splitting" ideas, and a new observation that the expansion exponents match with the dimensions under these reductions. A detailed overview of the proof is in Section 7.1.3.

Part I

The Abelian Groups

Chapter 2

Integer colorings with forbidden rainbow sums

For a set of positive integers $A \subseteq [n]$, an r-coloring of A is rainbow sum-free if it contains no rainbow Schur triple. In this chapter we initiate the study of the rainbow Erdős–Rothschild problem in the context of sum-free sets, which asks for the subsets of $[n]$ with the maximum number of rainbow sum-free r-colorings. We show that for $r = 3$, the interval $[n]$ is optimal, while for $r \geq 8$, the set $[\lfloor n/2 \rfloor, n]$ is optimal. We also prove a stability theorem for $r \geq 4$. The proofs rely on the hypergraph container method, and some ad-hoc stability analysis. This chapter is based on joint work with Cheng, Li, Wang, and Zhou [41].

2.1 Introduction

An interesting direction of combinatorics in recent years is the study of multicolored version of classical extremal results, whose origin can be traced back to a question of Erdős and Rothschild [60] in 1974. They asked which n-vertex graph admits the maximum number of 2-edge-colorings without monochromatic triangles, and conjectured that the complete balanced bipartite graph is the optimal graph. About twenty years later, Yuster [170] confirmed this conjecture for sufficiently large n.

2.1.1 Erdős–Rothschild problems in various settings

There are many natural generalizations of the Erdős–Rothschild problem. The most obvious one may be to ask it for graphs other than the triangles, and one may also increase the number of colors used. A graph G on n vertices is called (r, F)-*extremal* if it admits the maximum number of r-edge-colorings without any monochromatic copies of F among all n-vertex graphs. Alon, Balogh, Keevash and Sudakov [1] greatly extended Yuster's result and showed that the Turán graph $T_k(n)$ is the unique (r, K_{k+1})-extremal graph for $k \geq 2$ and $r \in \{2, 3\}$. Interestingly, they also showed that Turán graphs $T_k(n)$ are no longer optimal for $r \geq 4$. Indeed, Pikhurko, and Yilma [142] later proved that $T_4(n)$ is the unique $(4, K_3)$-extremal graph, while $T_9(n)$ is the unique $(4, K_4)$-extremal graph. Determining the extremal configurations in general for $k \geq 2$ and $r \geq 4$ turned out to be a difficult problem. For further results along this line of research (when F is a non-complete graph or a hypergraph), we refer to [92, 94, 95, 119, 120, 121].

Another variant of this problem is to study edge-colorings of a graph avoiding a copy of F with a prescribed color pattern. For an r-colored graph $\hat{F}$, a graph G on n vertices is called $(r, \hat{F})$-*extremal* if it admits the maximum number of r-colorings which contain no subgraph whose color pattern is isomorphic to $\hat{F}$. This line of work was initiated by Balogh [6], who showed that the Turán graph $T_k(n)$ once again yields the maximum number of 2-colorings avoiding H_{k+1}, where H_{k+1} is any 2-coloring of K_{k+1} that uses both colors. For $r \geq 3$, the behavior of (r, H_{k+1})-extremal graphs was studied by Benevides, Hoppen, Sampaio, Lefmann, and Odermann, see [12, 96, 98, 99]. In particular, the case when $\hat{F} = \hat{K}_3$ is a triangle with rainbow pattern has recently received a lot of attention (for its relation to Gallai colorings). Hoppen, Lefmann and Odermann [98] first proved that the Turán graph $T_2(n)$ is the unique $(r, \hat{K}_3)$-extremal graph for $r \geq 5$. Very recently, Balogh and Li [7], confirming conjectures of [12] and [98], showed that the complete graph K_n is the unique $(3, \hat{K}_3)$-extremal graph, while the Turán graph $T_2(n)$ becomes optimal as $r \geq 4$.

The Erdős–Rothschild problem can also be extended to other discrete structures. In the domain of extremal set theory, Hoppen, Kohayakawa and Lefmann [93] solved the Erdős–Rothschild extension of the famous Erdős–Ko–Rado Theorem. They, for instance, showed that the optimal ℓ-intersecting families (each set is of size k) yields the maximum number of r-colorings in which every color class is ℓ-intersecting for $r \in \{2, 3\}$, and also provided a fairly complete characterization of the corresponding extremal family for $r \geq 4$. Hoppen, Lefmann and Odermann [97], and Clemens, Das and Tran [44] later studied the Erdős–Rothschild extension of the Erdős–Ko–Rado Theorem for vector spaces. Moving the problem to the context of power set lattice, recently, Das, Glebov, Sudakov and Tran [48] investigated the the Erdős–Rothschild extension of the Sperner's Theorem, and proved that the largest antichain yields the maximum number of r-colorings, in which each color class is an antichain, for $r \in \{2, 3\}$. As for many of the previous results, they demonstrated that as r grows, the largest antichain is no longer optimal. They also determined that the extremal configurations for 2-colorings without monochromatic k-chains are the largest k-chain-free family. The extremal configurations for $r \geq 3$ and $k \geq 2$ are widely unknown.

2.1.2 Erdős–Rothschild problems for sum-free sets

Given integers $n \geq m \geq 1$, write $[m, n] := \{m, \dots, n\}$ and $[n] := \{1, ..., n\}$.

Definition 2.1 (Schur triple & Sum-free set)**.** A *Schur triple* or a *sum* in an abelian group G (or in $[n]$) is a triple $\{a, b, c\}$ with $a + b = c$. A set $A \subseteq G$ (or $A \subseteq [n]$) is *sum-free* if A contains no such triple.

Given a set A of numbers, an *r-coloring* of A is a mapping $f : A \to [r]$, which assigns one color to each element of A. An r-coloring of A is called a *sum-free r-coloring* if each of the color classes is a sum-free set. Sum-free colorings are among the classical objects studied in extremal combinatorics and can be traced back to Schur's theorem, one of the seminal results in Ramsey theory.

The Erdős–Rothschild extension for sum-free sets has been pursued by Liu, Sharifzadeh and Staden [124] for subsets of the integers, and Hàn and Jiménez [87] for finite abelian groups. More specifically, they investigated the extremal configurations which maximize the number of sum-free r-colorings. In the setting of integers, it is well known that the largest sum-free set in $[n]$ has size $\lceil n/2 \rceil$. Liu, Sharifzadeh and Staden [124] determined the extremal configurations for $r = 2$.

Theorem 2.2 ([124]). *There exists $n_0 > 0$ such that for all integers $n \geq n_0$, the number of sum-free 2-colorings of a subset $A \subseteq [n]$ is at most $2^{\lceil n/2 \rceil}$. Moreover, the extremal subsets are $\{1, 3, 5, \cdots, 2\lceil n/2 \rceil - 1\}$, and $[\lfloor n/2 \rfloor + 1, n]$; and if n is even, we additionally have $[n/2, n-1]$, and $[n/2, n]$.*

Unlike the graph case, in the sum-free setting, there are extremal configurations which are not sum-free even for 2 colors. Therefore, one would expect a more sophisticated extremal behavior as r grows. Although some asymptotic bounds were obtained in [124], the characterization of extremal sets for $r \geq 3$ remains widely open.

Such problem was also studied for finite abelian groups. Let G denote a finite abelian group. Over fifty years ago, Diananda and Yap [55] determined the maximum density $\mu(G)$ of a sum-free set in G whenever $|G|$ has a prime factor $q \not\equiv 1 \mod 3$, but it was not until 2005 that Green and Ruzsa [78] completely solved this extremal question for all finite abelian group. Hàn and Jiménez [87] investigated the Erdős–Rothschild extension for sum-free sets on some special abelian groups.

Theorem 2.3 ([87]). *Let $r \in \{2, 3\}$, $q \in N$ and let G be a abelian group of sufficiently large order, which has a prime divisor q such that $q \equiv 2 \mod 3$. Then the number of sum-free r-coloring of a set $A \subseteq G$ is at most $r^{\mu(G)}$. Moreover, the maximum is only achieved by the largest sum-free set.*

For more than three colors this phenomenon does not persist in general and the problem becomes considerably more complicated. For more details, we refer the readers to [87]. For other abelian groups, despite some asymptotic bounds presented in [87], the exact extremal phenomena is unknown even for 2 colors.

2.1.3 Our results

In this chapter, we consider a rainbow variant of the Erdős–Rothschild problem for sum-free sets in $[n]$. A Schur triple or a sum $\{x, y, z\}$ is a *rainbow sum* if x, y, z are colored with different colors. Note that a rainbow sum must have three distinct elements. For convenience, sometimes we would use the following definitions, which are slightly different with the classical notations on sum-free sets.

Definition 2.4 (Restricted Schur triple & Restricted sum-free set)**.** A *restricted Schur triple* or a *restricted sum* in $[n]$ is an ordered triple (a, b, c) with $a < b < c$ and $a + b = c$. A set $A \subseteq [n]$ is *restricted sum-free* if A contains no such triple.

For any integer $n \geq 7$, it is not hard to show that the largest restricted sum-free sets in $[n]$ have size $\lfloor n/2 \rfloor + 1$. If n is even, then the only subset attaining this bound is $\left[\frac{n}{2}, n\right]$; if n is odd, then the maximum restricted sum-free sets are attained by the following four sets: $\left\{\frac{n-1}{2}, \frac{n-1}{2} + 1, \ldots, n-1\right\}$, $\left\{\frac{n-1}{2}, \frac{n-1}{2} + 2, \ldots, n\right\}$, $\left[\frac{n+1}{2}, n\right]$, and $\{1, 3, 5, \ldots, n\}$.

Given a set of positive integers $A \subseteq [n]$, an r-coloring of A is *rainbow sum-free* if it contains no rainbow sum. For a positive integer r and a set $A \subseteq [n]$, we write $g(A, r)$ for the number of rainbow sum-free r-colorings of A and define

$$g(n, r) := \max_{A \subseteq [n]} g(A, r).$$

A set $A \subseteq [n]$ is *rainbow r-extremal* if $g(A, r) = g(n, r)$. When $r \in \{1, 2\}$, it is trivial to see that $g(n, r) = r^n$ for all positive integers n, and the only extremal set is the interval $[n]$, since for every subset $A \subseteq [n]$, all r-colorings of A are rainbow sum-free. For $r \geq 3$, the characterization of the extremal sets requires substantially more work.

Our first main result is an upper bound on the number of rainbow sum-free r-colorings of dense sets.

Theorem 2.5. *For every integer $r \geq 3$, there exists n_0 such that for all $n > n_0$ the following* *holds. For a set $A \subseteq [n]$ with $|A| \geq (1 - r^{-3})n$, the number of rainbow sum-free r-colorings $g(A, r)$ satisfies*

$$g(A, r) \leq \binom{r}{2} \cdot 2^{|A|} + 2^{-\frac{n}{26 \log n}} 2^n.$$

By choosing two of the r colors and coloring the elements of $[n]$ arbitrarily with these two colors, one can easily obtain that

$$g([n], r) \geq \binom{r}{2}(2^n - 2) + r = \binom{r}{2} 2^n - (r^2 - 2r). \tag{2.1}$$

Therefore, Theorem 2.5 is asymptotically sharp for $A = [n]$ and then the typical structure of rainbow sum-free r-colorings of $[n]$ immediately follows from (2.1).

Corollary 2.6. *For every integer $r \geq 3$, almost all rainbow sum-free r-colorings of $[n]$ are 2-colorings.*

Now we turn to the extremal configurations of rainbow sum-free r-colorings. Let us first consider the case $r = 3$. Similarly as in the Gallai coloring problem, two natural candidates of the extremal sets are the maximum restricted sum-free sets and the interval $[n]$. Note that for every restrict sum-free set A, we have $g(A, 3) \leq 3^{\lfloor n/2 \rfloor + 1} \ll g([n], 3)$. Our second theorem shows that for three colors the interval $[n]$ is indeed optimal.

Theorem 2.7. *There exists n_0 such that for all $n > n_0$, among all subsets of $[n]$, the interval $[n]$ is the unique rainbow 3-extremal set.*

Just as for the Erdős-Rothschild extension for Gallai colorings [7], we may not expect that the same phenomena persists for $r \geq 4$. Define $O := \{1, 3, 5, \cdots, 2\lceil n/2 \rceil - 1\}$, and $I_0 := [\lfloor n/2 \rfloor + 1, n]$. We prove the following stability theorem.

Theorem 2.8. *For every positive integer $r \geq 4$, we have*

$$g(n, r) = r^{n/2 + o(n)}.$$

Moreover, for every $\varepsilon > 0$, there exist $\delta, n_0 > 0$ such that for all integers $n \geq n_0$ the following holds. Let A be a subset of $[n]$ with $g(A, r) \geq r^{n/2 - \delta n}$. Then

(i) *for $r \geq 5$, we have that either $|A \triangle O| \leq \varepsilon n$, or $|A \triangle I_0| \leq \varepsilon n$;*

(ii) *for $r = 4$, we have that either $|A \triangle [n]| \leq \varepsilon n$, or $|A \triangle O| \leq \varepsilon n$, or $|A \triangle I_0| \leq \varepsilon n$.*

The behavior of the exact extremal configurations not only depends on the number of colors, but also depend on the parity of n. For even n, we define

$$I_1 = \left[\frac{n}{2} - 1, n\right], \qquad I_2 = \left[\frac{n}{2}, n\right].$$

Observe that I_1 contains exactly two restricted Schur triples $(n/2-1, n/2, n-1)$, $(n/2-1, n/2+1, n)$, and it is not hard to compute that $g(I_1, r) = r^{n/2}\left(3 - 2/r\right)^2$. On the other hand, the set I_2 is a restricted sum-free set and therefore $g(I_2, r) = r^{|I_1|} = r^{n/2+1}$. For odd n, we define

$$I_3 = \left[\frac{n-1}{2}, n\right].$$

Again, the set I_3 contains exactly one restricted Schur triple $(\frac{n-1}{2}, \frac{n-1}{2} + 1, n)$, and one can show that $g(I_3, r) = r^{\lceil n/2 \rceil}\left(3 - 2/r\right)$, which is already greater than the number of colorings for any restricted sum-free set. When a set A is of size at least the size of the maximum restricted sum-free sets and not one of the above three sets, we believe that the restrictions from the triples would more than counteract the extra possibilities offered by the additional vertices. Therefore, we make the following conjecture.

Conjecture 2.9. *Let n, r be positive integers and $r \geq 4$.*

(i) *If n is even and $r \leq 7$, then $g(n, r) = r^{n/2}\left(3 - 2/r\right)^2$, and I_1 is the unique rainbow r-extremal set.*

(ii) *If n is even and $r \geq 8$, then $g(n, r) = r^{n/2+1}$, and I_2 is the unique rainbow r-extremal set.*

(iii) *If n is odd and $r = 4$, then $g(n, r) = g([n], r)$, and $[n]$ is the unique rainbow r-extremal set.*

(iv) *If n is odd and $r \geq 5$, then $g(n, r) = r^{\lceil n/2 \rceil}(3 - 2/r)$, and I_3 is the unique rainbow r-extremal set.*

Our forth main result verifies Conjecture 2.9 for $r \geq 8$ and n sufficiently large.

Theorem 2.10. *For an integer $r \geq 8$, there exists $n_0 = n_0(r)$ such that for all $n > n_0$ the following holds. Let A be a subset of $[n]$ with $|A| \geq \lceil n/2 \rceil + 1$.*

(i) *If n is even, then $g(A, r) \leq r^{\lceil n/2 \rceil + 1}$, and the equality holds if and only if $A = I_2$.*

(ii) *If n is odd, then $g(A, r) \leq r^{\lceil n/2 \rceil}(3 - 2/r)$, and the equality holds if and only if $A = I_3$.*

The chapter is organized as follows. In the next section, we list some structural results on sum-free sets, which are essential for the proof, and introduce the multi-color container theorem. In Section 2.3, we prove Theorem 2.5. In Section 2.4, we prove the stability theorem, Theorem 2.8, and determine $g(n, 3)$ for n sufficiently large. In Section 2.5, we determine $g(n, r)$ for $r \geq 8$, and describe the corresponding extremal configurations. We close the chapter with some concluding remarks in Section 2.6. All logaritheorems have base 2.

2.2 Notation and preliminaries

2.2.1 Basic properties of restricted sum-free sets

We use the following result of Staden [157] on the minimum number of additive triples among all sets of a given size.

Theorem 2.11 ([157]). *Let A be a subset of $[n]$ with $|A| > \lceil n/2 \rceil$. Then the number of Schur triples in A is at least*

$$(|A| - \lceil n/2 \rceil)(|A| - \lfloor n/2 \rfloor),$$

where the unique minimising set is $[n - |A| + 1, n]$.

For a set $A \subseteq [n]$, we write $\mathcal{S}(A)$ for the set of all restricted Schur triples in A, and let $s(A) = |\mathcal{S}(A)|$. For an integer $t \in A$, denote by $\mathcal{S}(t, A)$ the set of all triples in $\mathcal{S}(A)$ containing t, and let $s(t, A) = |\mathcal{S}(t, A)|$. Then from Theorem 2.11, we immediately obtain the following proposition.

Proposition 2.12. *Let A be a subset of $[n]$ with $|A| > \lceil n/2 \rceil$. Then*

$$s(A) \geq (|A| - \lceil n/2 \rceil)(|A| - \lfloor n/2 \rfloor) - |A|/2.$$

In particular, we have

$$s([n]) = \begin{cases} \frac{n^2-2n}{4} & \text{if } n \text{ is even;} \\ \frac{n^2-2n+1}{4} & \text{otherwise.} \end{cases}$$

2.2.2 Structural properties of sum-free sets

We will use standard definitions and notation in additive combinatorics as given in [163]. Given $A, B \subseteq \mathbb{Z}$, let

$$A + B := \{a + b \mid a \in A, b \in B\}, \quad \text{and} \quad A - B := \{a - b \mid a \in A, b \in B\}.$$

When $B = \{x\}$, we simply write $A + x$ and $A - x$.

The following lemma is known as Green's removal lemma, which was first proved by Green [76], and was later generalized to nonabelian groups by Král and Vena [117].

Lemma 2.13 ([117, 76]). *For all $\varepsilon > 0$, there exists $\delta, n_0 > 0$ such that the following holds for all integers $n \geq n_0$. Suppose that $A \subseteq [n]$ is a set containing at most δn^2 Schur triples. Then there exist $B, C \subseteq [n]$ such that $A = B \cup C$ where B is sum-free and $|C| \leq \varepsilon n$.*

We also require a very strong stability theorem for sum-free sets proved by Deshouillers, Freiman, and Odermann [52].

Lemma 2.14 ([52]). *Every sum-free set S in $[n]$ satisfies at least one of the following conditions:*

(i) $|S| \leq 2n/5$;

(ii) *S consists of odd numbers;*

(iii) $|S| \leq \min(S)$.

2.2.3 Multi-color container theorem

An important tool in our proof is the hypergraph container theorem. We use the following version from [8]. Let $\mathcal{H}$ be a k-uniform hypergraph with average degree d. The *co-degree* of a set of vertices $X \subseteq V(\mathcal{H})$ is the number of edges containing X; that is,

$$d(X) = \{e \in E(\mathcal{H}) \mid X \subseteq e\}.$$

For every integer $2 \leq j \leq k$, the j-th maximum co-degree of $\mathcal{H}$ is

$$\Delta_j(\mathcal{H}) = \max\{d(X) \mid X \subseteq V(\mathcal{H}),\ |X| = j\}.$$

When the underlying hypergraph is clear, we simply write it as Δ_j. For $0 < \tau < 1$, the *co-degree function* $\Delta(\mathcal{H}, \tau)$ is defined as

$$\Delta(\mathcal{H}, \tau) = 2^{\binom{k}{2}-1} \sum_{j=2}^{k} 2^{-\binom{j-1}{2}} \frac{\Delta_j}{d\tau^{j-1}}.$$

In particular, when $k = 3$,

$$\Delta(\mathcal{H}, \tau) = \frac{4\Delta_2}{d\tau} + \frac{2\Delta_3}{d\tau^2}.$$

Theorem 2.15 ([8])**.** *Let $\mathcal{H}$ be a k-uniform hypergraph on vertex set $[N]$. Let $0 < \varepsilon, \tau < 1/2$. Suppose that $\tau < 1/(200k!^2k)$ and $\Delta(\mathcal{H}, \tau) \leq \varepsilon/(12k!)$. Then there exists $c = c(k) \leq 1000k!^3k$ and a collection of vertex subsets $\mathcal{C}$ such that*

(i) *every independent set in $\mathcal{H}$ is a subset of some of $A \in \mathcal{C}$;*

(ii) *for every $A \in \mathcal{C}$, $e(\mathcal{H}[A]) \leq \varepsilon \cdot e(\mathcal{H})$;*

(iii) $\log |\mathcal{C}| \leq cN\tau \log(1/\varepsilon) \log(1/\tau)$.

A key concept in applying container theory to such coloring problems is the notion of *template*, which was first introduced in [64], although the concept had already appeared in [153] under the name of '2-colored multigraphs' and later in [9], simply referred as 'containers'.

Definition 2.16 (Template and palette)**.** An *r-template* of order n is a function $P : [n] \to 2^{[r]}$, associating to each element $x \in [n]$ a list of colors $P(x) \subseteq [r]$. We refer to this set $P(x)$ as the *palette* available at x.

For a set $A \subseteq [n]$, any r-coloring of A can be considered as an r-template of order n, with only one color allowed at each element in A, and no color allowed for elements not belonging to A.

Definition 2.17 (Subtemplate)**.** Let P_1, P_2 be two r-templates of order n. We say that P_1 is a *subtemplate* of P_2 (written as $P_1 \subseteq P_2$) if $P_1(x) \subseteq P_2(x)$ for each element $x \in [n]$.

For an r-template P of order n, write $RS(P)$ for the number of subtemplate of P that are rainbow restricted sums. We say that P is a *rainbow restricted sum-free* r-template if $RS(P) = 0$. Using Theorem 2.15, we obtain the following.

Theorem 2.18. *For every integer $r \geq 3$, there exists a constant $c = c(r)$ and a collection $\mathcal{C}$ of r-templates of order n such that*

(i) *every rainbow restricted sum-free r-template of order n is a subtemplate of some $P \in \mathcal{C}$;*

(ii) *for every $P \in \mathcal{C}$, $RS(P) \le n^{-1/3} s([n])$;*

(iii) $|\mathcal{C}| \le 2^{cn^{2/3}\log^2 n}$.

Proof. Let $\mathcal{H}$ be a 3-uniform hypergraph with vertex set $[n] \times \{1, 2, ..., r\}$, whose edges are all triples $\{(x_1, c_1), (x_2, c_2), (x_3, c_3)\}$ such that (x_1, x_2, x_3) forms a restricted Schur triple in $[n]$ and c_1, c_2, c_3 are all different. In other words, every hyperedge in $\mathcal{H}$ corresponds to a rainbow restricted Schur triple. Note that there are exactly $r(r-1)(r-2)$ ways to rainbow color a restricted Schur triple with r colors. Hence, the average degree d of $\mathcal{H}$ is equal to

$$d = \frac{3e(\mathcal{H})}{v(\mathcal{H})} = \frac{3r(r-1)(r-2)s([n])}{nr} \ge \frac{3(r-1)(r-2)n}{8}.$$

Now we apply Theorem 2.15 on $\mathcal{H}$. Let $\varepsilon = n^{-1/3}/r(r-1)(r-2)$ and $\tau = \sqrt{96 \cdot 3! \cdot r}n^{-\frac{1}{3}}$. Observe that $\Delta_2(\mathcal{H}) = 2(r-2), \Delta_3(\mathcal{H}) = 1$. For n sufficiently large, we can get $\tau < 1/(200 \cdot 3!^2 \cdot 3)$ and

$$\Delta(\mathcal{H}, \tau) = \frac{4\Delta_2}{d\tau} + \frac{2\Delta_3}{d\tau^2} = \frac{8(r-2)}{d\tau} + \frac{2}{d\tau^2} \le \frac{3}{d\tau^2} \le \frac{\varepsilon}{12 \cdot 3!}.$$

Hence, there is a collection of vertex subsets $\mathcal{C}$ satisfying properties (i)-(iii) of Theorem 2.15. Observe that every vertex set of $\mathcal{H}$ corresponds to an r-template of order n; every rainbow restricted sum-free r-template of order n corresponds to an independent set in $\mathcal{H}$. Therefore, $\mathcal{C}$ is a desired collection of r-templates. $\square$

Definition 2.19 (Good r-template)**.** For $A \subseteq [n]$, an r-template P of order n is a *good r-template* of A if it satisfies the following properties:

(i) For each element $i \in A$, $|P(i)| \ge 1$;

(ii) $RS(P) \le n^{-1/3} s([n])$.

For a set $A \subseteq [n]$ and a collection of templates $\mathcal{P}$, denote by $G(\mathcal{P}, A)$ the set of rainbow sum-free r-colorings of A, which is a subtemplate of some $P \in \mathcal{P}$. Let $g(\mathcal{P}, A) = |G(\mathcal{P}, A)|$. If $\mathcal{P}$ consists of a single r-template P, then we simply write $G(P, A)$ and $g(P, A)$.

2.3 Proof of Theorem 2.5

Throughout this section, we fix an integer $r \geq 3$, a sufficiently large integer n and an arbitrary set $A \subseteq [n]$ with $|A| = (1 - \xi)n$, where

$$0 \leq \xi \leq r^{-3}.$$

Let $\mathcal{C}$ be the collection of containers given by Theorem 2.18, and $\delta = 1/(24 \log n)$. We divide $\mathcal{C}$ into two classes

$$\mathcal{C}_1 = \{P \in \mathcal{C} : g(P, A) \leq 2^{(1-\delta)n}\}, \quad \mathcal{C}_2 = \{P \in \mathcal{C} : g(P, A) > 2^{(1-\delta)n}\}. \tag{2.2}$$

Note that every $P \in \mathcal{C}_2$ is a good r-template of A. The crucial part of the proof is to estimate $g(\mathcal{C}_2, A)$, which replies on the following four lemmas.

Lemma 2.20. *Let F be the collection of ordered pairs $(a, b) \in A^2$ with $a < b$ such that $\{a, b\} \nsubseteq S$ for all $S \in \mathcal{S}(A)$. Then we have $|F| \leq \xi n^2 + n/6$.*

Proof. Let

$$F_1 = \{(a, b) \in A^2 \mid a + b \in [n] \backslash A,\ b = 2a\},$$

$$F_2 = \{(a, b) \in A^2 \mid a + b > n,\ b = 2a\},$$

$$F_3 = \{(a, b) \in A^2 \mid a + b \in [n] \backslash A,\ b - a \in [n] \backslash A\},$$

$$F_4 = \{(a, b) \in A^2 \mid a + b > n,\ b - a \in [n] \backslash A\}.$$

Clearly, $|F| = \sum_{i=1}^{4} |F_i|$ and $|F_1| \le |[n]\backslash A| = \xi n$. Since every $(a,b) \in F_2$ satisfies $b = 2a \le n$ and $a + b = 3a > n$, we have $|F_2| \le n/6$. Moreover, we obtain that $|F_3| \le \binom{\xi n}{2} \le \xi^2 n^2/2$, as $a + b \in [n]\backslash A$ and $b - a \in [n]\backslash A$. Similarly, we have $|F_4| \le \xi n^2/2$, as $b > n/2$ and $b - a \in [n]\backslash A$. Finally, we conclude that $|F| \le \xi n + n/6 + \xi^2 n^2/2 + \xi n^2/2 \le \xi n^2 + n/6$. $\square$

For a template P of A, let

$$X_1 = \{x \in A \mid |P(x)| = 1\}, \quad X_2 = \{x \in A \mid |P(x)| = 2\}, \quad X_3 = \{x \in A \mid |P(x)| \ge 3\},$$

and $x_i = |X_i|$ for $i \in [3]$.

Lemma 2.21. *Suppose that P is a template of A in $\mathcal{C}_2$. Then we have*

$$\max\left\{\frac{(\xi - \delta)n + x_1}{\log r - 1},\ 0\right\} < x_3 \le 2n^{-1/3}n.$$

In particular, if $\xi \ge 2(\log r - 1)n^{-1/3} + \delta$, then $\mathcal{C}_2$ is empty.

Proof. By the definitions of $G(P,A)$ and $\mathcal{C}_2$, we have

$$2^{x_2} r^{x_3} \ge g(P,A) > 2^{(1-\delta)n}. \tag{2.3}$$

Since $x_2 = |A| - x_1 - x_3$ and $|A| = (1-\xi)n$, we obtain that

$$x_3 > \frac{(\xi - \delta)n + x_1}{\log r - 1}. \tag{2.4}$$

We first claim that $x_2 \ge (1-\xi)n/3$. Otherwise, we immediately have $x_1 + x_3 > 2(1-\xi)n/3$. Together with (2.4), we obtain that $x_3 > \frac{(2+\xi-3\delta)n}{3\log r} > \frac{n}{2\log r}$. By Lemma 2.20 and $0 \le \xi \le r^{-3}$, there are at least

$$\binom{x_3}{2} - (\xi n^2 + n/6) \ge \frac{n^2}{16\log^2 r}$$

pairs in X_3, which are contained in some restricted Schur triples in A. This contradicts the definition of good r-templates, as $RS(P) \geq n^2/(3 \cdot 16 \log^2 r) > n^{-1/3}s([n])$.

For each $a \in X_3$, let $B_a := \{b \in X_2 \mid \{a,b\} \subseteq S, \text{ for some } S \in \mathcal{S}(A)\}$. Note that every $b \in X_2 \setminus B_a$ satisfies either $|b-a| \in [n] \setminus A$ or $|b-a| = \min\{a,b\}$. Then we have $|B_a| \geq (1-\xi)n/3 - 2\xi n - 1 > n/4$. Since P is a good r-template of A, we obtain that

$$n^{-1/3}s([n]) \geq RS(P) \geq \frac{1}{2}\sum_{a \in X_3} |B_a| \geq \frac{x_3 \cdot n}{8},$$

which indicates $x_3 \leq 2n^{-1/3}n$. □

Next, we will prove a stability result on good templates with many rainbow sum-free colorings.

Lemma 2.22. *Let $0 \leq \xi < 2(\log r - 1)n^{-1/3} + \delta$. Then for every $P \in \mathcal{C}_2$, there exist two colors $\{i,j\} \in [r]$ such that the number of elements in A with palette $\{i,j\}$ is at least $(1-2\delta)n$.*

Proof. By Lemma 2.21 and (2.3), we have

$$x_2 \geq (1 - \delta - 3\log r \cdot n^{-1/3})n.$$

For $1 \leq i < j \leq r$, define $Y_{i,j} := \{x \in X_2 \mid P(x) = \{i,j\}\}$. Without loss of generality, we can assume that $|Y_{1,2}| \geq x_2/\binom{r}{2}$. Let $Y' = X_2 \setminus Y_{1,2}$. For each $a \in Y'$, let $B_a = \{b \in Y_{1,2} \mid \{a,b\} \subseteq S, \text{ for some } S \in \mathcal{S}(A)\}$. Similarly as in Lemma 2.21, we obtain that $|B_a| \geq x_2/\binom{r}{2} - 2\xi n - 1$, and then

$$n^{-1/3}s([n]) \geq RS(P) \geq \frac{1}{2}\sum_{x \in Y'} |B_a| \geq \frac{|Y'| \cdot n}{2r(r-1)}.$$

Since $\delta \gg n^{-1/3}$, we have $|Y_{1,2}| = x_2 - |Y'| \geq (1-2\delta)n$, which completes the proof. □

Lemma 2.23. *For two colors $i, j \in [r]$, denote by $\mathcal{P} = \mathcal{P}(i,j)$ the set of good r-template of*

A, in which there are at least $(1-2\delta)n$ elements in A with palette $\{i,j\}$. Then

$$g(\mathcal{P},A) \le 2^{|A|}(1+2^{-n/12}).$$

Proof. For an r-coloring $g \in G(\mathcal{P},A)$, let $S(g)$ the set of elements in A, which are not colored by i or j. By the definition of $\mathcal{P}$, we have $|S(g)| \le 2\delta n$. Define

$$\mathcal{G}_0 = \{g \in G(\mathcal{P},A) \mid S(g) = \varnothing\}, \quad \text{and} \quad \mathcal{G}_1 = \{g \in G(\mathcal{P},A) \mid |S(g)| \ge 1\}.$$

Clearly, we have $g(\mathcal{P},A) = |\mathcal{G}_0|+|\mathcal{G}_1|$ and $|\mathcal{G}_0| \le 2^{|A|}$. It remains to show that $|\mathcal{G}_1| \le 2^{|A|-n/12}$.

Let us consider the ways to color A so that the resulting colorings are in $\mathcal{G}_1$. We first choose a set $A_0 \subseteq A$ of size at most $2\delta n$, which will be colored by the colors in $[r] \setminus \{i,j\}$. The number of options is at most $\sum_{1\le k\le 2\delta n}\binom{n}{k}$, and the number of colorings is at most $r^{2\delta n}$. Once we fix A_0 and its color, take an arbitrary vertex $t \in A_0$.

Claim 1. Let $\mathcal{D}(t)$ be the collection of disjoint pairs $\{a,b\}$ in $A \setminus A_0$ such that $\{a,b,t\}$ forms a restricted Schur triple. Then $|\mathcal{D}(t)| \ge n/6$.

Proof of Claim 1. Define

$$\mathcal{S}_1 = \{(a,b) \in [n]^2 \mid a+b=t,\ a<b\},\ \textit{and}\ \mathcal{S}_2 = \{(a,b) \in [n]^2 \mid t+a=b,\ a<b\}.$$

We first observe that $|\mathcal{S}_1| = \lfloor (t-1)/2 \rfloor$ for every $t \in [n]$. Note that all pairs in $\mathcal{S}_1$ are disjoint. Therefore, if $t > 2n/5$, we have $|\mathcal{D}(t)| \ge |\mathcal{S}_1| - \xi n - |A_0| \ge |\mathcal{S}_1| - (2\delta+\xi)n \ge n/6$. If $t \le 2n/5$, observe that $|\mathcal{S}_2| = n-2t$ and all pairs in $\mathcal{S}_2$ are disjoint. Therefore, we obtain that $|\mathcal{D}(t)| \ge |\mathcal{S}_2| - \xi n - |A_0| \ge |\mathcal{S}_2| - (2\delta+\xi)n \ge n/6$. $\bowtie$

For every pair $(a,b) \in \mathcal{D}(t)$, since t is colored by some color in $[r] \setminus \{i,j\}$, and a,b can only be colored by i or j, the elements a and b must receive the same color in order to avoid the rainbow Schur triple. Therefore, together with Claim 1, the number of ways to finish the colorings is at most

$$2^{|A|-|A_0|-|\mathcal{D}(t)|} \leq 2^{|A|-n/6}.$$

Hence, we obtain that

$$|\mathcal{G}_2| \leq \sum_{1\leq k\leq 2\delta n} \binom{n}{k} r^{2\delta n} 2^{|A|-n/6} \leq 2^{|A|+2\delta n(\log n+\log r)-n/6} \leq 2^{|A|-n/12},$$

where the last inequality follows from $\delta = 1/(24\log n)$. $\square$

Now we have all ingredients to prove Theorem 2.5.

Proof of Theorem 2.5. First, by property (i) of Theorem 2.18 , every rainbow sum-free r-coloring of A is a subtemplate of some $P \in \mathcal{C}$. By Property (iii) of Theorem 2.18 and the definition of $\mathcal{C}_1$ (see (2.2)), we have

$$g(\mathcal{C}_1, A) \leq |\mathcal{C}_1| \cdot 2^{(1-\delta)n} \leq |\mathcal{C}| \cdot 2^{(1-\delta)n} < 2^n \cdot 2^{-n/(25\log n)}.$$

If $\xi \geq 2(\log r - 1)n^{-1/3} + \delta$, using Lemma 2.21, we are done by $g(\mathcal{C}, A) = g(\mathcal{C}_1, A)$. Otherwise, by Lemma 2.22 and Lemma 2.23, we obtain that

$$g(\mathcal{C}_2, A) = \sum_{1\leq i<j\leq r} g(\mathcal{P}(i,j), A) \leq \binom{r}{2} 2^{|A|}(1 + 2^{-n/12}).$$

Hence, we have

$$g(\mathcal{C}, A) = g(\mathcal{C}_1, A) + g(\mathcal{C}_2, A) \leq 2^n \cdot 2^{-\frac{n}{25\log n}} + \binom{r}{2} 2^{|A|}(1 + 2^{-n/12}) \leq \binom{r}{2} \cdot 2^{|A|} + 2^{-\frac{n}{26\log n}} 2^n,$$

which gives the desired upper bound on the number of rainbow sum-free r-colorings of A. $\square$

2.4 Proof of Theorems 2.7 and 2.8

The following lemma gives us a structural description of large sum-free sets.

Lemma 2.24. *Let $\varepsilon, c > 0$, $c > 10\varepsilon$, and $\varepsilon < 1/10$. Let $A, B \subseteq [n]$ such that $A \cap B = \varnothing$, B is sum-free, and $|A| = cn$. If $|B| \geq (1/2 - \varepsilon)n$, then $(A + B) \cap B \neq \varnothing$.*

Proof. Suppose, to the contrary, that $(A + B) \cap B = \varnothing$. Since $|B| > 2n/5$, by Lemma 2.14, either B only contains odd numbers, or the minimum element of B is at least $|B| \geq (1/2-\varepsilon)n$. If B only contains odd numbers, then there is $d \geq c - \varepsilon$, such that $|A \cap E| = dn$, where $E \subseteq [n]$ is the collection of all even numbers. Thus, there exists an $a \in A \cap E$ such that $dn \leq a \leq (1 - d)n$. Let P be the collection of all pairs $(i,\ i + a)$, where i is odd, and $1 \leq i \leq dn$. Observe that all pairs in P are pairwise disjoint and there are at least $dn/2$ of them. Since $(A + B) \cap B = \varnothing$, for each pair (i, j) in P, at least one of $\{i, j\}$ is not in B. This implies

$$|B| \leq \frac{n}{2} - |P| \leq \frac{n}{2} - \frac{dn}{2} \leq \frac{n}{2} - \frac{(c-\varepsilon)n}{2} \leq \left(\frac{1}{2} - 2\varepsilon\right) n,$$

which contradicts the assumption of B.

If the minimum element of B is at least $|B| \geq (1/2 - \varepsilon)n$, let b be the smallest element in B, then there is $d \geq c - 2\varepsilon$ such that $|A \cap [b - 1]| = dn$. This implies that there exists $a \in A$ with $dn/2 \leq a \leq b - dn/2$. We define P to be the collection of all pairs (i, j), where $b \leq i < (1/2 + 3\varepsilon)n$ and $j = i + a$. Then the number of pairs in P is at least $2\varepsilon n$, as $b \leq (1/2 + \varepsilon)n$. Moreover, for every (i, j) in P, we have $j < n$ since $b - dn/2 \leq (1/2 - 3\varepsilon)n$ and since $b + a > (1/2 + 3\varepsilon)n$, P is a set of disjoint pairs. Since $(A + B) \cap B = \varnothing$, for every (i, j) in P, at least one of $\{i, j\}$ is not in B. Similarly we obtain that $|B| \leq (1/2 - 2\varepsilon)n$, a contradiction. $\square$

Our next lemma says that when the number of colorings $r = 3$ and the size of A is significantly smaller than n, the number of rainbow sum-free r-colorings will be much less than 2^n. And when $r \geq 5$ and the size of A is significantly larger than $n/2$, much less than

$r^{n/2}$ when r, the number of rainbow sum-free r-colorings will be much less than $r^{n/2}$.

Lemma 2.25. *Let $\varepsilon > 0$, r be a positive integer, and let A be a subset of $[n]$. Then the followings hold.*

(i) *If $r = 3$, and $|A| \le (1-\varepsilon)n$, then there is a constant $\delta_1 = \delta_1(\varepsilon) > 0$, such that*

$$g(A, 3) \le 2^{(1-\delta_1)n}.$$

(ii) *If $r \ge 5$, and $|A| \ge \big(1/2 + \varepsilon\big)n$, then there is a constant $\delta_1 = \delta_1(\varepsilon, r) > 0$ such that*

$$g(A, r) \le r^{(1/2-\delta_1)n}.$$

Proof. Let $\mathcal{C}$ be the collection of containers given by Theorem 2.18, and let

$$g_{\max}(P, A) = \max_{P \in \mathcal{C}} g(P, A).$$

For a template $P \in \mathcal{C}$, suppose P is not a good template. Then there must be an element $i \in A$ with $|P(i)| = 0$, which immediately gives $g(P, A) = 0$. Therefore, $g_{\max}(P, A)$ is always achieved by a good template.

Let P be a good template of A. Since $RS(P) = o(n^2)$, by Green's aritheoremetic removal lemma, there is a set $E \subseteq [n]$ and a template $P' : [n] \setminus E \to 2^{[r]}$, such that $P\mid_{[n]\setminus E} = P'$, $|E| = o(n)$, and P' has no rainbow Schur triples. Define

$$X_1 = \{a \in [n] \setminus E : |P'(a)| = 1\}, \quad X_2 = \{a \in [n] \setminus E : |P'(a)| = 2\},$$

and

$$X_3 = \{a \in [n] \setminus E : |P'(a)| \ge 3\}.$$

Let $T = X_2 \cup X_3$ and let $x_i = |X_i|$ for $i = 1, 2, 3$. Therefore, we have

$$(X_3 + X_3) \cap X_3 = \varnothing, \quad (T + T) \cap X_3 = \varnothing, \quad (X_3 + T) \cap T = \varnothing. \tag{2.5}$$

As X_3 is sum-free, we have $x_3 \leq \lfloor n/2 \rfloor + 1$.

Let m be the largest element in X_3. By (2.5), for every $i < m$, at least one of $\{i,\ m - i\}$ is not in T which is also the same for X_3. Hence, we have

$$|T| \leq n - \left\lceil \frac{m-1}{2} \right\rceil, \quad x_3 \leq m - \left\lceil \frac{m-1}{2} \right\rceil. \tag{2.6}$$

Case 1: $r = 3$.

Observe that we may assume $\varepsilon < 2/5$, as otherwise we will get $g(A, 3) \leq 3^{|A|} \leq 2^{0.955n}$, which completes the proof with $\delta_1 = 0.045$. We first consider the case when $x_2 \leq (1 - 5\varepsilon/2)n$. Then we have

$$\begin{aligned} \log g(\mathcal{C}, A) &\leq \log(|\mathcal{C}| \cdot g_{\max}(P, A)) \leq cn^{2/3} \log^2 n + |E| \log 3 + x_2 + x_3 \log 3 \\ &= o(n) + x_2 + x_3 \log 3 = \left(o(n) + \frac{1}{2} \left(|T| + x_3 - \left(1 - \frac{2}{\log 3} \right) x_2 \right) \right) \log 3 \\ &\leq n + o(n) - \frac{5\varepsilon}{2} \left(1 - \frac{\log 3}{2} \right) n < (1 - \delta_1) n, \end{aligned}$$

where we take $\delta_1 = \frac{5\varepsilon}{4}(1 - \frac{\log 3}{2})$.

Now, we may assume that $x_2 > (1 - 5\varepsilon/2)n$. Then $x_3 \leq |A| - x_2 < 3\varepsilon n/2$. Thus we obtain

$$\begin{aligned} \log g(\mathcal{C}, A) &\leq o(n) + x_2 + x_3 \log 3 \leq o(n) + |A| + (\log 3 - 1) x_3 \\ &\leq n + o(n) - \frac{\varepsilon}{2} (5 - 3 \log 3)\, n < (1 - \delta_1) n, \end{aligned}$$

and we take $\delta_1 = \frac{1}{4} (5 - 3 \log 3)\, \varepsilon$.

Case 2: $r \geq 5$**.**

Since $|A| \geq (1/2+\varepsilon)n$, and P is good, we have that $x_1 + x_2 \geq |A| - x_3 - |E| \geq \varepsilon n/2$ for large enough n. We first assume that $x_2 \geq \frac{\varepsilon n}{100}$. Similarly, we get

$$\begin{aligned}\log g(\mathcal{C}, A) &\leq \log(|\mathcal{C}| \cdot g_{\max}(P, A)) \leq cn^{2/3}\log^2 n + |E|\log r + x_2 + x_3 \log r \\ &= o(n) + x_2 + x_3 \log r = \left(o(n) + \frac{1}{2}\left(|T| + x_3 - \left(1 - \frac{2}{\log r}\right)x_2\right)\right)\log r \\ &\leq \left(\frac{n}{2} + o(n) - \frac{1}{2}\left(1 - \frac{2}{\log r}\right)\frac{\varepsilon}{100}n\right)\log r < \left(\frac{n}{2} - \delta_1 n\right)\log r, \end{aligned} \tag{2.7}$$

where we take $\delta_1 = \frac{1}{300}\left(1 - \frac{2}{\log r}\right)\varepsilon$. Note that $\delta_1 > 0$ as $r \geq 5$.

Finally, we may assume that $x_2 \leq \frac{\varepsilon n}{100}$, and then $x_1 \geq \frac{\varepsilon n}{2} - x_2 \geq \frac{\varepsilon n}{3}$. We claim that $x_3 \leq (1/2 - \varepsilon/40)n$. Otherwise, by the way we construct P', we also have $(X_1 + X_3) \cap X_3 = \varnothing$ and this contradicts Lemma 2.24. Similarly as before, we can conclude that

$$\log g(\mathcal{C}, A) \leq o(n) + x_2 + x_3 \log r \leq \left(\frac{n}{2} + o(n) + \frac{\varepsilon}{100 \log r}n - \frac{\varepsilon}{80}n\right)\log r \leq \left(\frac{n}{2} - \delta_1 n\right)\log r,$$

where we take $\delta_1 = \frac{\varepsilon}{400}$. □

The case when $r = 4$ is more involved, and we will discuss it later in this section. But the result in Lemma 2.25 (i) is enough to imply Theorem 2.7.

Proof of Theorem 2.7. Observe that $g([n], 3) \geq 3 \cdot 2^n - 3$. Suppose $A \subseteq [n]$ and $A \neq [n]$. When $|A| \leq (1 - 3^{-3})n$, by Lemma 2.25 (i), there is $\delta_1 > 0$ such that $g(A, 3) \leq 2^{(1-\delta_1)n} < g([n], 3)$. Now, we have $(1 - 3^{-3})n < |A| \leq n - 1$. By Theorem 2.5, $g(A, 3) \leq (1.5 + o(1))2^n < g([n], 3)$. □

The next lemma records an easy fact about intervals for convenience in the proof of the analogue result of Lemma 2.25 when we have only 4 colors.

Lemma 2.26. *Let $\varepsilon > 0$, and let a, b be integers such that $0 < a < b < n$, $3\varepsilon n < a < n/2 - 2\varepsilon n$, and $a + 2 + 3\varepsilon n < b \leq 2a$. Suppose $A \subseteq [a+1, b]$, $B \subseteq [b+1, n]$, and*

$|A| > b - a - \varepsilon n$, $|B| > n - b - \varepsilon n$. *Then* $(A + A) \cap B \neq \varnothing$.

Proof. Let α be the smallest element in A, then $\alpha \leq a+\varepsilon n+1$. Let $J = [\alpha+1, \alpha+1+\lceil 2\varepsilon n \rceil] \subseteq [a+1, b]$. Observe that $\alpha + J \subseteq [b+1, n]$. Since $|J| = \lceil 2\varepsilon n \rceil$, and $|[a+1, b] \setminus A| < \varepsilon n$, $|[b+1, n] \setminus B| < \varepsilon n$, this implies there is $\beta \in A \cap J$ such that $\alpha + \beta \in B$. $\square$

The next lemma, Lemma 2.27, is similar to Lemma 2.25, but here we consider the case when the number of colors is 4. In order to obtain the same conclusion in Lemma 2.25, we further require that the size of A is significantly smaller than n, since if A is close to $[n]$, when we color all the elements in A by two colors, the number of colorings we obtained is also close to the extremal case. Note that if we use the same proof as in Lemma 2.25 for $r = 4$, equation (2.7) does not give us the conclusion we want. Hence the proof of Lemma 2.27 requires a more careful and complicated analysis of the structures of the containers.

Lemma 2.27. *Let* $\varepsilon > 0$ *such that* $\big(1/2+\varepsilon\big)n \leq |A| \leq (1-\varepsilon)n$. *Then there is* $\delta_2 = \delta_2(\varepsilon) > 0$ *such that*

$$g(A, 4) \leq 4^{n/2-\delta_2 n}.$$

Proof. We apply Theorem 2.18 on A. Let $\mathcal{C}$ be the collection of containers, and let $P \in \mathcal{C}$ be a good template of order n. As what we did in the proof of Lemma 2.25, we similarly apply Green's removal lemma on P, and obtain a template $P' : [n]\backslash E \to 2^{[r]}$, such that $P\mid_{[n]\backslash E} = P'$, $|E| = o(n)$, and P' is sum-free. Let $X_1, X_2, X_3 \subseteq A\backslash E$ such that $X_1 = \{a \in A \mid |P'(a)| = 1\}$, $X_2 = \{a \in A \mid |P'(a)| = 2\}$, and $X_3 = \{a \in A \mid |P'(a)| \geq 3\}$. Let $T = X_2 \cup X_3$ and $x_i = |X_i|$ for $i \in [3]$. Therefore, we have equations (2.5) still hold, and in particular, X_3 is sum-free. Thus $x_3 \leq (n+1)/2$. Since $|A| \geq (1/2+\varepsilon)n$, and P is good, we also obtain that $x_1 + x_2 \geq \varepsilon n/2$. Let $m = n - \alpha$ be the maximum element in X_3, by using the same argument, equations (2.6) still hold.

Suppose we have either

$$|T| \leq n - \left\lceil \frac{m-1}{2} \right\rceil - \frac{\varepsilon n}{1000}, \quad \text{or} \quad x_3 \leq m - \left\lceil \frac{m-1}{2} \right\rceil - \frac{\varepsilon n}{1000}.$$

Thus

$$g(\mathcal{C}, A) \leq 2^{cn^{2/3}\log^2 n} r^{|E|} 2^{x_2} r^{x_3} = r^{o(n)+\frac{1}{2}(|T|+x_3)} \leq r^{\frac{n}{2}+o(n)-\frac{\varepsilon n}{2000}} < r^{\frac{n}{2}-\delta_2 n},$$

and we take $\delta_2 = \frac{\varepsilon}{3000}$. Therefore, we may assume that

$$n - \left\lceil \frac{m-1}{2} \right\rceil - \frac{\varepsilon n}{1000} \leq |T| \leq n - \left\lceil \frac{m-1}{2} \right\rceil, \text{ and } m - \left\lceil \frac{m-1}{2} \right\rceil - \frac{\varepsilon n}{1000} \leq x_3 \leq m - \left\lceil \frac{m-1}{2} \right\rceil. \tag{2.8}$$

In the rest of the proof, we are going to show that this is impossible.

Suppose that $\alpha \leq \frac{\varepsilon n}{40}$. Note that $\max\{x_1, x_2\} \geq \frac{\varepsilon n}{4}$, and

$$(X_1 + X_3) \cap X_3 = \varnothing, \quad (X_2 + X_3) \cap X_3 = \varnothing,$$

this contradicts Lemma 2.24. Thus we have $m \leq (1 - \frac{\varepsilon}{40})n$. Since $|A| \leq (1-\varepsilon)n$, thus by the lower bound on $|T|$ in (2.8), $m \geq 3\varepsilon n/2$.

We now partition $[n]$ into three parts J_1, J_2, J_3, such that $J_1 = [n-\alpha+1, n]$, $J_2 = [1, \alpha]$, and $J_3 = [\alpha+1, n-\alpha]$. By (2.8), we obtain that

$$|J_1 \cap X_2| \geq \alpha - \frac{\varepsilon n}{1000}. \tag{2.9}$$

Take $d = \frac{\varepsilon}{400}$. Suppose $|J_2 \cap X_3| \geq dn$, then we can find $\beta \in X_3$ such that $\frac{dn}{2} \leq \beta \leq \alpha - \frac{dn}{2}$. Let $J_1' = [n-\alpha+1, n-\alpha+\frac{dn}{2}] \subseteq J_1$. Note that $(J_1'+\beta) \cap J_1' = \varnothing$, and $J_1'+\beta \subseteq J_1$, since $(X_2+X_3) \cap X_2 = \varnothing$, and for every $i \in J_1'$, there is at least one element in the pair $\{i, i+\beta\}$ that is not contained in X_2, we have that $|J_1 \cap X_2| \leq \alpha - \frac{dn}{2} = \alpha - \frac{\varepsilon n}{800}$, contradicts (2.9). Hence we may assume $|J_2 \cap X_3| \leq dn$. Therefore, by (2.8), $|J_3 \cap X_3| \geq \frac{n-\alpha}{2} - \frac{\varepsilon n}{1000} - dn$. This gives us an upper bound on α since $|J_3| \geq |J_3 \cap X_3|$, that $\alpha \leq \frac{n}{3} + \frac{\varepsilon n}{1500} + \frac{2dn}{3}$. Next, we are going to show that actually we have $\alpha \leq \frac{n}{3}$. Suppose $\alpha > \frac{n}{3}$. Note that $|J_3 \cap X_3| \geq \frac{n-\alpha}{2} - dn - \frac{\varepsilon n}{1000}$

implies

$$|J_3 \setminus X_3| \le \frac{n-3\alpha}{2} + dn + \frac{\varepsilon n}{1000} \le dn + \frac{\varepsilon n}{1000} < \frac{\varepsilon n}{250}.$$

By (2.9) and Lemma 2.26, we get $(X_3 + X_3) \cap X_2 \neq \varnothing$, this contradicts (2.5).

Let $J_3' = [n-2\alpha, n-\alpha]$. We claim that

$$|J_3' \setminus X_3| \le dn + \frac{\varepsilon n}{800} = \frac{3\varepsilon n}{800}. \tag{2.10}$$

Otherwise, observe that at least one of $\{i, m-i\}$ is not in X_3, then $|X_3 \cap (J_3 \setminus J_3')| \le \frac{n-3\alpha}{2}$ since $\alpha \le n/3$. Hence

$$x_3 \le \frac{m}{2} - \alpha + dn + \alpha - \left(dn + \frac{\varepsilon n}{800}\right) \le \frac{m}{2} - \frac{\varepsilon n}{800},$$

contradicts (2.8).

Next, let $d' = \frac{\varepsilon}{60}$, and suppose that $|(J_2+\alpha) \cap X_3| \ge d'n$. Thus there is $\gamma \in X_3$ such that $\alpha + \frac{d'n}{2} \le \gamma \le 2\alpha - \frac{d'n}{2}$. Let $J_1'' = [n-2\alpha, n-2\alpha+\frac{d'n}{2}]$. Observe that $(\gamma + J_1'') \cap J_1'' = \varnothing$, and $\gamma + J_1'' \subseteq J_1$. Since $(X_2 + X_3) \cap X_2 = \varnothing$, we have either $|X_3 \cap J_3'| \le \alpha - \frac{d'n}{4} < \alpha - \frac{3\varepsilon n}{800}$, or $|X_2 \cap J_1| \le \alpha - \frac{d'n}{4} < \alpha - \frac{\varepsilon n}{1000}$, and in either case we get a contradiction with (2.9) or (2.10). Thus, we have $|(J_2+\alpha) \cap X_3| \le d'n$. Note that $J_3' = \{m\} \cup (m - J_2)$, clearly, $|X_3 \cap (J_2 \cup J_3')| \le \alpha + 1$. Then by (2.8),

$$\frac{n-3\alpha}{2} - \frac{\varepsilon n}{1000} \le \big|X_3 \cap [\alpha, n-2\alpha]\big| \le n - 4\alpha + d'n,$$

hence $\alpha \le \frac{n}{5} + \frac{2d'n}{5} + \frac{\varepsilon n}{2500}$. Suppose now $\alpha \ge \frac{n}{5}$. We have that $|X_3 \cap [2\alpha, n-\alpha]| \ge \frac{n-\alpha}{2} - dn - d'n - \frac{\varepsilon n}{1000}$, which implies $|[2\alpha, n-\alpha] \setminus X_3| \le \frac{\varepsilon n}{1000} + dn + d'n$. By (2.9) and Lemma 2.26, we obtain that $(X_3 + X_3) \cap X_2 \neq \varnothing$, and this contradicts (2.5).

Finally, we get $\frac{\varepsilon n}{40} \le \alpha < \frac{n}{5}$. By (2.8), we have that $x_3 \le \frac{2n}{5}$. By Lemma 2.14, either X_3 consists of odd integers, or the minimum element in X_3 is at least $\frac{n-\alpha}{2} - \frac{\varepsilon n}{1000}$. The first case is impossible. Otherwise, since $|J_2| = \alpha > 2dn + \frac{\varepsilon n}{1000}$, we have

$$x_3 = |X_3 \cap J_2| + |X_3 \cap J_3| \leq dn + \frac{n - 2\alpha}{2} \leq \frac{n - \alpha}{2} - \frac{9\varepsilon n}{800},$$

and this contradicts the lower bound on x_3 in (2.8). Now, we assume $a \in X_3$ is the minimum element, and $a \geq \frac{n-\alpha}{2} - \frac{\varepsilon n}{1000}$. Observe that $|[\frac{n-\alpha}{2} + 1, n - \alpha] \setminus X_3| \leq \frac{\varepsilon n}{500}$, $\frac{n-\alpha}{2} \leq \frac{n}{2} - \frac{\varepsilon n}{80}$, and by (2.9) and Lemma 2.26, we have $(X_3 + X_3) \cap X_2 \neq \varnothing$, this contradicts (2.5). □

The final lemma consider the case when A contains many Schur triples.

Lemma 2.28. *Let $r \geq 4$ be an integer. Suppose there is $\mu > 0$, such that $s(A) \geq \mu n^2$. Then*

$$g(A, r) \leq r^{|A| - \frac{3(2\log r - \log(3r-2))}{2\log r}\mu n}.$$

Proof. Since $s(A) \geq \mu n^2$, by Pigeonhole Principle, there is $t \in A$, such that

$$s(t, A) \geq \frac{3\mu n^2}{|A|} \geq 3\mu n.$$

Let the *link graph* $L_t(A)$ to be the simple graph defined on the vertex set $A \setminus \{t\}$, such that $xy \in E(L_t(A))$ if and only if $\{t, x, y\} \in \mathcal{S}(t, A)$. Let k be the size of the maximum matching in $L_t(A)$. Observe that $\Delta(L_t(A)) \leq 2$, and $|E(L_t(A))| = s(t, A) \geq 3\mu n$. Then we have $k \geq 3\mu n/2$.

Now we consider the possible number of rainbow sum-free colorings of A. We first fix a maximum matching M of $L_t(A)$. For the elements in $A \setminus V(M)$, we color them arbitrarily. For each edge $ab \in E(M)$, in order to avoid a rainbow Schur triple, we either let a, b share the same color, or color one of a, b by the color of t, and color another vertex by a different color. In this way, a, b have exactly $r + 2(r-1)$ effective colorings. Hence, we have

$$g(A, r) \leq r^{|A|-2k-1} r(3r-2)^k \leq r^{|A|} \left(\frac{3r-2}{r^2}\right)^{\frac{3\mu n}{2}} = r^{|A| - \frac{3(2\log r - \log(3r-2))}{2\log r}\mu n},$$

as desired. $\square$

Now we can prove the stability theorem.

Proof of Theorem 2.8. The first part of the statement, that $g(n,r) \leq r^{n/2+o(n)}$, follows easily from the fact $g(A,r) \leq r^{|A|}$ when $|A| \leq n/2 + o(n)$. If $|A| \geq n/2 + \eta n$ for some constant η, the result follows from Lemma 2.25 (ii) when $r \geq 5$. For the case $r = 4$, after applying Lemma 2.27 we still have one extra case that $|A| \geq (1-\eta)n$, and this follows from Theorem 2.5.

For the second part of the statement, we will prove it by contrapositive. Let $c = \frac{3(2\log r - \log(3r-2))}{2\log r}$, clearly $c > 0$ when $r \geq 4$. Let μ be the value of $\delta(\frac{\varepsilon}{20})$ given in Lemma 2.13, and let $\varepsilon' = \min\{\frac{c\mu}{2}, \varepsilon\}$. We first consider $r \geq 5$, and suppose that we have both

$$|A \triangle O| > \varepsilon n, \quad \text{and} \quad |A \triangle I_0| > \varepsilon n. \tag{2.11}$$

In this case we take $\delta = \min\{\delta_1(\varepsilon'), \varepsilon', \frac{\varepsilon}{20}\}$, where $\delta_1(\varepsilon')$ is given in Lemma 2.25 (ii). If $|A| \geq \frac{n}{2} + \varepsilon' n$, we apply Lemma 2.25 (ii) with parameter ε', then we obtain that $g(A,r) \leq r^{n/2-\delta n}$. Thus we may assume that $|A| \leq (1/2+\varepsilon')n$. If $s(A) \geq \mu n^2$, applying Lemma 2.28, we have

$$g(A,r) \leq r^{n/2+\varepsilon' n - c\mu n} \leq r^{n/2-\varepsilon' n} \leq r^{n/2-\delta n}.$$

Finally, we have $s(A) < \mu n^2$. By Lemma 2.13, we get the partition $A = B \cup C$, where B is sum-free and $|C| < \frac{\varepsilon}{20} n$. Note that we may assume $|A| \geq \frac{n}{2} - \frac{\varepsilon}{20} n$, otherwise $g(A,r) \leq r^{|A|} \leq r^{n/2-\delta n}$. Now we have

$$|B| \geq |A| - |C| \geq \frac{n}{2} - \frac{\varepsilon n}{10} \geq \frac{2n}{5}$$

since $\varepsilon \leq 1$. We apply Lemma 2.14 on B. Hence either B contains only odd integers, or the minimum element of B is at least $|B|$. Suppose B consists of odd integers. Thus

$$|A \triangle O| \leq |C| + |O \setminus B| \leq \frac{\varepsilon}{20}n + \frac{\varepsilon}{10}n < \varepsilon n,$$

contradicts (2.11). Thus, let a be the minimum element in B, then $a \geq \frac{n}{2} - \frac{\varepsilon n}{10}$. Therefore,

$$|A \triangle I_0| \leq |C| + |B \triangle I_0| \leq \frac{\varepsilon}{20}n + \frac{\varepsilon}{10}n + \frac{\varepsilon}{5}n < \varepsilon n,$$

which also contradicts (2.11).

Next, let us consider the case when $r = 4$. Besides (2.11), we further require

$$|A \triangle [n]| > \varepsilon n. \tag{2.12}$$

We now take $\delta = \min\{\delta_2(\varepsilon'), \varepsilon', \frac{\varepsilon}{20}\}$, where $\delta_2(\varepsilon')$ is given in Lemma 2.27. The case when $|A| \leq \frac{n}{2} + \varepsilon' n$ is same as when $r \geq 5$. When $\frac{n}{2} + \varepsilon' n \leq |A| \leq n - \varepsilon' n$, by applying Lemma 2.27 we get $g(A, 4) \leq 4^{n/2 - \delta n}$. When $|A| \geq n - \varepsilon' n$, we get $|A \triangle [n]| \leq \varepsilon' n \leq \varepsilon n$, which contradicts (2.12). □

2.5 Proof of Theorem 2.10

For a subset $A \subseteq [n]$ and an integer t, recall that $L_t(A)$ is the simple graph defined on $A \setminus \{t\}$, in which $xy \in E(L_t(A))$ if and only if $\{t, x, y\} \in \mathcal{S}(t, A)$. Let $k(t, A)$ be the size of the maximum matching of $L_t(A)$. Note that $\Delta(L_t(A)) \leq 2$. Therefore we have

$$k(t, A) \geq |E(L_t(A))|/2 = s(t, A)/2. \tag{2.13}$$

Proposition 2.29. *Let $n, r, c \in N$ with $r \geq 8$ and $c > 1$. Suppose that A is a subset of $[n]$ of size $\lceil n/2 \rceil + c$.*

(i) *If there exists an element $t \in A$ such that $k(t, A) \geq 2(c-1)$, then we have $g(A, r) < r^{\lceil n/2 \rceil + 1}$.*

(ii) *If there exists an element $t \in A$ such that $k(t, A) \geq 2(c-1)+1$, then we have $g(A, r) < r^{\lceil n/2 \rceil} (3 - 2/r)$.*

Proof. First note that for $r \geq 8$ we have

$$\frac{(3r-2)^2}{r^3} < 1. \tag{2.14}$$

Let $k = k(t, A)$. Similarly as in the proof of Lemma 2.28, we obtain that

$$g(A, r) \leq r^{\lceil n/2 \rceil + c - 2k} (3r-2)^k = r^{\lceil n/2 \rceil + c} \left(\frac{3r-2}{r^2} \right)^k.$$

For $k \geq 2(c-1)$, we have

$$g(A, r) \leq r^{\lceil n/2 \rceil + c} \left(\frac{3r-2}{r^2} \right)^{2(c-1)} = r^{\lceil n/2 \rceil + 1} \left(\frac{(3r-2)^2}{r^3} \right)^{c-1} < r^{\lceil n/2 \rceil + 1},$$

where the last inequality follows from (2.14) and $c > 1$.

Similarly, for $k \geq 2(c-1)+1$, we have

$$\begin{aligned} g(A, r) &\leq r^{\lceil n/2 \rceil} \left(3 - \frac{2}{r} \right) r^{c-1} \left(\frac{3r-2}{r^2} \right)^{k-1} \leq r^{\lceil n/2 \rceil} \left(3 - \frac{2}{r} \right) r^{c-1} \left(\frac{3r-2}{r^2} \right)^{2(c-1)} \\ &= r^{\lceil n/2 \rceil} \left(3 - \frac{2}{r} \right) \left(\frac{(3r-2)^2}{r^3} \right)^{c-1} < r^{\lceil n/2 \rceil} \left(3 - \frac{2}{r} \right). \end{aligned}$$

Together with the previous inequality, this completes the proof. □

Lemma 2.30. *Let $r \geq 8$, $0 < \varepsilon \leq 1/36$, and A be a subset of $[n]$ of size $\lceil n/2 \rceil + c$, where $1 < c \leq \varepsilon n$. Suppose that there exists a partition $A = B \cup C$ such that B consists of odd numbers and $|C| \leq \varepsilon n$. Then we have $g(A, r) < r^{\lceil n/2 \rceil} (3 - 2/r)$.*

Proof. From the assumption of A, there must be an even number $t \in A$. By Proposition 2.29(ii) and $\varepsilon \leq 1/36$, it is sufficient to show that $k(t, A) \geq (1/12 - \varepsilon)n - 1$.

Recall that O is the set of all odd numbers in $[n]$. Since $|A| \geq \lceil n/2 \rceil + 1$ and $|C| \leq \varepsilon n$, we have $|O \setminus B| \leq \varepsilon n$. Then,

$$k(t, A) \geq k(t, B) \geq k(t, O) - \varepsilon n,$$

and thus it is equivalent to show that $k(t, O) \geq n/12 - 1$.

If $t \geq n/3$, we immediately have

$$k(t, O) \geq |\{(i, t - i, t), i \in O \cap [t/2 - 1]\}| \geq t/4 - 1 \geq n/12 - 1.$$

If $t < n/3$, then by (2.13) we obtain that

$$k(t, O) \geq s(t, O)/2 \geq |\{(t, i, t + i), i \in O \cap [t + 1,\ n - t]\}| /2 \geq (n - 2t)/4 \geq n/12.$$

This completes the proof. $\square$

Lemma 2.31. *Let $r \geq 8$, $0 < \varepsilon \ll 1$, and A be a subset of $[n]$ of size $\lceil n/2 \rceil + c$, where $1 < c \leq \varepsilon n$. Suppose that there exists a partition $A = B \cup C$ such that $B \subseteq I_0$ and $|C| \leq \varepsilon n$. Then the following holds.*

(i) *If n is even, then $g(A, r) < r^{\lceil n/2 \rceil + 1}$.*

(ii) *If n is odd, then $g(A, r) < r^{\lceil n/2 \rceil} (3 - 2/r)$.*

Proof. Let m be the minimum element of A, and clearly $m \leq \lfloor n/2 \rfloor - (c - 1)$. Recall that $I_0 = [\lfloor n/2 \rfloor + 1, n]$. Let $d = |I_0 \setminus A|$. From the assumption of A, we have $d \leq \varepsilon n$. We divide the proof into four cases.

Case 1: $m \leq d + 3(c - 1)$. In this case, we have $m \leq 4\varepsilon n$. Similar to the proof of

Lemma 2.30, we have

$$k(m, A) \geq k(m, B) \geq k(m, I_0) - d \geq s(m, I_0)/2 - \varepsilon n \geq (n/2 - m)/2 - \varepsilon n \geq n/4 - 3\varepsilon n,$$

which, together with Proposition 2.29(ii) and $\varepsilon \ll 1$, completes the proof.

Case 2: $d + 3(c-1) < m \leq \lceil n/2 \rceil - d - 3(c-1)$**.** Since $m \leq n/2$, each nontrivial component of $L_m(I_0)$ is a path, and there are $\min\{m, \lceil n/2 \rceil - m\} \geq d + 3(c-1)$ of them. Therefore we have

$$k(m, A) \geq k(m, I_0) - d \geq d + 3(c-1) - d = 3(c-1)$$

which, together with Proposition 2.29(ii) and $c > 1$, completes the proof.

Case 3: $\lceil n/2 \rceil - d - 3(c-1) < m \leq \lceil n/2 \rceil - 2(c-1)$**.** By the choice of m, each nontrivial component of $L_m(A)$ is a path of length 1, and the number of them is exactly

$$s(m, [m+1, n]) - |[m+1, n] \setminus A| = n - 2m - (n - |A| - (m-1)) = \lceil n/2 \rceil + (c-1) - m.$$

Therefore, we obtain that

$$k(m, A) = \lceil n/2 \rceil + (c-1) - m \geq 3(c-1),$$

which completes the proof together with Proposition 2.29(ii) and $c > 1$.

Case 4: $\lceil n/2 \rceil - 2(c-1) < m \leq \lfloor n/2 \rfloor - (c-1)$**.** Similarly as in Case 3, we obtain that $k(m, A) = \lceil n/2 \rceil + (c-1) - m$. By the choice of m, for even n, we have $k(m, A) \geq 2(c-1)$, while for odd n, $k(m, A) \geq 2(c-1) + 1$. By Proposition 2.29, this gives the desired upper bounds. $\square$

Proof of Theorem 2.10. Here we only prove (i) as the proof of (ii) is similar. If $|A| = \lceil n/2 \rceil + 1$ and $A \neq I_2$, then A must have at least one restricted Schur triple, and therefore $g(A, r) < g(I_2, r) = r^{\lceil n/2 \rceil + 1}$. When $|A| > \lceil n/2 \rceil + 1$, choose a constant $\varepsilon \ll 1$, which satisfies the assumptions of Lemmas 2.30 and 2.31. Then by Theorem 2.8, we can further assume that $|A \triangle O| \leq \varepsilon n$, or $|A \triangle I_0| \leq \varepsilon n$. Applying Lemmas 2.30 and 2.31(i) on A, for both cases, we obtain $g(A, r) < r^{\lceil n/2 \rceil + 1}$. □

2.6 Concluding Remarks

Our investigation raises many open problems. In this chapter, we determine the rainbow r-extremal sets, that is, the subsets of $[n]$ which maximize the number of rainbow sum-free r-colorings, for $r \leq 3$ and $r \geq 8$. However, for $r \in \{4, 5, 6, 7\}$, although Theorem 2.8 says the rainbow r-extremal sets should be close to what we expect, our proofs cannot give the exact structure of the extremal sets. Therefore, the most interesting question is to determine the unsolved cases of Conjecture 2.9. Recall that $I_1 = [\frac{n}{2} - 1, n]$ and $I_3 = [\frac{n-1}{2}, n]$.

Conjecture 2.32. *Let n, r be positive integers and $4 \leq r \leq 7$.*

(i) *If n is even, then $g(n, r) = r^{n/2} (3 - 2/r)^2$, and I_1 is the unique rainbow r-extremal set.*

(ii) *If n is odd and $r = 4$, then $g(n, r) = g([n], r)$, and $[n]$ is the unique rainbow r-extremal set.*

(iii) *If n is odd and $5 \leq r \leq 7$, then $g(n, r) = r^{\lceil n/2 \rceil} (3 - 2/r)$, and I_3 is the unique rainbow r-extremal set.*

Another direction is that one can consider various generalization of this problem. Recall that a sum-free set is a set forbidding the solutions of the linear equation $x_1 + x_2 = y$. It is natural to extend the Erdős–Rothschild problems on sets forbidding solutions of other linear equations, for example, the (k, ℓ)-free sets, that is, the sets without nontrivial tuples $\{x_1, \ldots, x_k, y_1, \ldots, y_\ell\}$ satisfying $\sum_{i=1}^{k} x_i = \sum_{j=1}^{\ell} y_j$. It is possible that the method used to prove Theorem 2.5 can prove the analogous results for some other (k, ℓ)-free sets. However, the stability analysis on other parts would be very involved.

One could also broaden the study of rainbow Erdős–Rothschild problems to various other extremal problems in this fashion. In the rainbow Erdős–Rothschild problems studied to date, that is, the Gallai colorings and the rainbow sum-free colorings, for $r = 3$ the configurations maximizing the number of such colorings are complete graphs or the whole intervals, while for sufficiently large r the optimal configurations are those solving the original extremal problems. It would be very interesting to determine the threshold of r to ensure that the extremal configurations for the uncolored problems are optimal for rainbow Erdős–Rothschild problems.

Chapter 3

Avoidance density for (k, ℓ)-sum-free sets

Let $\mathcal{M}_{(2,1)}(N)$ be the infimum of the largest sum-free subset of any set of N positive integers. An old conjecture in additive combinatorics asserts that there is a constant $c = c(2, 1)$ and a function $\omega(N) \to \infty$ as $N \to \infty$, such that $cN + \omega(N) < \mathcal{M}_{(2,1)}(N) < (c + o(1))N$. The constant $c(2, 1)$ is determined by Eberhard, Green, and Manners [57], while the existence of $\omega(N)$ is still wide open.

In this chapter, we study the analogous conjecture on (k, ℓ)-sum-free sets and restricted (k, ℓ)-sum-free sets. We determine the constant $c(k, \ell)$ for every (k, ℓ)-sum-free sets, answering a question asked by Bajnok [4]. We also confirm the conjecture for infinitely many (k, ℓ). This chapter is based on joint work with Wu [108].

3.1 Introduction

In 1965, Erdős asked the following question [59]. Given an arbitrary sequence A of N different positive integers, what is the size of the largest sum-free subsequence of A? By *sum-free* we mean that if $x, y, z \in A$, then $x + y \neq z$. Let

$$\mathcal{M}_{(2,1)}(N) = \inf_{\substack{A \subseteq \mathbb{N}^{>0} \\ |A| = N}} \max_{\substack{S \subseteq A \\ S \text{ is sum-free}}} |S|.$$

Using a beautiful probabilistic argument, Erdős showed that every N-element set $A \subseteq \mathbb{N}^{>0}$ contains a sum-free subset of size at least $N/3$, in other words, $\mathcal{M}_{(2,1)}(N) \geq N/3$. It turns out that it is surprisingly hard to improve upon this bound. The result was later improved by Alon and Kleitman [3], who showed that $\mathcal{M}_{(2,1)}(N) \geq (N+1)/3$. Bourgain [28], using an entirely different Fourier analytic argument, showed that $\mathcal{M}_{(2,1)}(N) \geq (N+2)/3$, which is the best lower bound on $\mathcal{M}_{(2,1)}(N)$ to date. In particular, the following conjecture has been made in a series of papers. See [59, 28, 57, 164] for example.

Conjecture 3.1. *There is a function $\omega(N) \to \infty$ as $N \to \infty$, such that*

$$\mathcal{M}_{(2,1)}(N) > \frac{N}{3} + \omega(N).$$

On the other hand, a recent breakthrough by Eberhard, Green, and Manners [57] proved that $\mathcal{M}_{(2,1)}(N) = (1/3 + o(1))N$. More precisely, they showed that for every $\varepsilon > 0$, when N is large enough, there is a set $A \subseteq \mathbb{N}^{>0}$ of size N, such that every subset of A of size at least $(1/3+\varepsilon)N$ contains x, y, z with $x + y = z$. This result is one of the first beautiful applications of the arithmetic regularity lemma. Later, using a completely different argument, the result is generalized by Eberhard [56] to k-*sum-free* set. A set A is k-sum-free if for every $y, x_1, \ldots, x_k \in A$, $y \neq \sum_{i=1}^{k} x_i$. Eberhard proved that for every $\varepsilon > 0$, there is a set $A \subseteq \mathbb{N}^{>0}$ of size N, such that every subset of A of size at least $(1/(k+1) + \varepsilon)N$ contains a k-sum. For more background we refer to the survey [164].

In this chapter, we study the analogue of the Erdős sum-free set problem for (k,ℓ)-sum-free sets. Given two positive integers k, ℓ with $k > \ell$, a set A is (k,ℓ)-*sum-free* if for every $x_1, \ldots, x_k, y_1, \ldots, y_\ell \in A$, $\sum_{i=1}^{k} x_k \neq \sum_{j=1}^{\ell} y_j$. For example, using the notation of (k,ℓ)-sum-free, sum-free is $(2,1)$-sum-free; k-sum-free is $(k,1)$-sum-free. Finding the largest (k,ℓ)-sum-free sets in some given structures has been well-studied in the past fifty years, for example, the size of the maximum (k,ℓ)-sum-free sets in finite cyclic groups was determined recently by Bajnok and Matzke [5], and the size in compact abelian groups was determined by Kravitz [118].

For every $A \subseteq \mathbb{N}^{>0}$, let

$$\mathcal{M}_{(k,\ell)}(A) = \max_{\substack{S \subseteq A \\ S \text{ is } (k,\ell)\text{-sum-free}}} |S|, \quad \text{and} \quad \mathcal{M}_{(k,\ell)}(N) = \inf_{\substack{A \subseteq \mathbb{N}^{>0} \\ |A|=N}} \mathcal{M}_{(k,\ell)}(A).$$

The problem of determining $\mathcal{M}_{(k,\ell)}(N)$ is suggested by Bajnok [4, Problem G.41]. In fact, we can also make the following conjecture for (k,ℓ)-sum-free set, which is an analogue of Conjecture 3.1.

Conjecture 3.2. *Let $k > \ell > 0$. There is a constant $c = c(k,\ell) > 0$, and a function $\omega(N) \to \infty$ as $N \to \infty$, such that*

$$cN + \omega(N) < \mathcal{M}_{(k,\ell)}(N) < (c + \varepsilon)N,$$

for every $\varepsilon > 0$.

As we mentioned above, the constant $c(k,\ell)$ in Conjecture 3.2 for $(k,\ell) = (2,1)$ is determined by Eberhard, Green, and Manners [57], and for $(k,\ell) = (k,1)$ is determined by Eberhard [56]. The conjecture for $(k,\ell) = (3,1)$ is confirmed by Bourgain [28].

Our first result determines the constant $c(k,\ell)$ in Conjecture 3.2 for every (k,ℓ) (see statements (i) and (iv) of Theorem 3.3), which answers a question asked by Bajnok [4] when the ambient group is $\mathbb{Z}$. The statement (ii) of Theorem 3.3 also confirms Conjecture 3.2 for infinitely many (k,ℓ).

Theorem 3.3. *Let k, ℓ be two positive integers and $k > \ell$. Then the following hold:*

(i) *for every k, ℓ, we have $\mathcal{M}_{(k,\ell)}(N) \geq \frac{N}{k+\ell}$.*

(ii) *suppose $k = 5\ell$. Then*
$$\mathcal{M}_{(k,\ell)}(N) \geq \frac{N}{k+\ell} + c\frac{\log N}{\log\log N}, \tag{3.1}$$
where $c > 0$ is an absolute constant that only depends on k, ℓ.

(iii) *for every set A of N positive integers, for every positive even integer u, there is an odd integer $v < u$ such that if $k = (u+v)\ell/(u-v)$, then*

$$\mathcal{M}_{(k,\ell)}(A) \geq \frac{N}{k+\ell} + c\frac{\log N}{\log\log N}, \tag{3.2}$$

where $c > 0$ is an absolute constant that only depends on k, ℓ.

(iv) *for every k, ℓ, we have $\mathcal{M}_{(k,\ell)}(N) = \left(\frac{1}{k+\ell} + o(1)\right)N$.*

We remark that Theorem 3.3 (iii) also implies estimate (3.1) when $k = 3\ell$, which in particular covers the $(3,1)$-sum-free case obtained by Bourgain. This is because when $u = 2$, the only possible value of v is 1, and this gives us $k = 3\ell$. It follows that estimate (3.2) holds for every N-element set A when $k = 3\ell$. Hence, by the definition of $\mathcal{M}_{(k,\ell)}(N)$, we prove estimate (3.1) when $k = 3\ell$.

The upper bound construction given by Eberhard, Green, and Manners [57] for $(2,1)$-sum-free set actually works in a more general setting: restricted $(2,1)$-sum-free set. A set A is *restricted (k,ℓ)-sum-free* if for every k distinct elements $a_1, \ldots, a_k$ in A, and ℓ distinct elements $b_1, \ldots, b_\ell$ in A, we have $\sum_{i=1}^k a_i \neq \sum_{j=1}^\ell b_j$. Let

$$\widehat{\mathcal{M}}_{(k,\ell)}(N) = \inf_{\substack{A \subseteq \mathbb{N}^{>0} \\ |A|=N}} \max_{\substack{S \subseteq A \\ S \text{ is restricted } (k,\ell)-\text{sum free}}} |S|.$$

Clearly, we have that $\mathcal{M}_{(k,\ell)}(N) \leq \widehat{\mathcal{M}}_{(k,\ell)}(N)$. Our next theorem gives us an upper bound on $\widehat{\mathcal{M}}_{(k,\ell)}(N)$ when $k \leq 2\ell + 1$.

Theorem 3.4. *Let k, ℓ be positive integers, and $k \leq 2\ell + 1$. Then*

$$\widehat{\mathcal{M}}_{(k,\ell)}(N) = \left(\frac{1}{k+\ell} + o(1)\right)N.$$

Overview

The chapter is organized as follows. In the next section, we provide some basic definitions and properties in additive combinatorics, harmonic analysis, and model theory (or more precisely, nonstandard analysis) used later in the proof. In Section 3.3, we prove a variant of the weak Littlewood conjecture, based on the ideas introduced by Bourgain [28]. Theorem 3.3 (i) is proved by using the probabilistic argument introduced by Erdős, and some structural results for the (k, ℓ)-sum-free open set on the torus. This is included in Section 3.4. One of the main parts of the chapter is to prove Theorem 3.3 (ii) and (iii). The special case for $(3, 1)$-sum-free set is proved by Bourgain [28], but his argument relies heavily on the fact that a certain term of the Fourier coefficient of the characteristic function is multiplicative, which is not true for the other (k, ℓ). Here we introduce a different sieve function, as well as a finer control on the functions we constructed. We will discuss it in detail in Section 3.5. In Sections 3.6 and 3.7, we prove Theorem 3.3 (iv). The proof goes by showing that the constructions given by Eberhard [56] for $(k, 1)$-sum-free sets, the Følner sequence, is still the correct construction for the other (k, ℓ)-sum-free sets. The new ingredients contain structural results for the large infinite (k, ℓ)-sum-free sets, which can be viewed as a generalization of the Łuczak–Schoen Theorem [127]. We will prove Theorem 3.4 in Section 3.8. In Section 3.9, we make some concluding remarks, and pose some open problems.

3.2 Preliminaries

3.2.1 Additive combinatorics

Throughout the chapter, we use standard definitions and notation in additive combinatorics as given in [163]. Let p be a prime, and let m, n, N ranging over positive integers. Given $a, b, N \in \mathbb{N}$ and $a < b$, let $[a, b] := [a, b] \cap \mathbb{N}$, and let $[N] := [1, N]$. We use the standard Vinogradov notation. That is, $f \ll g$ means $f = O(g)$, and $f \asymp g$ if $f \ll g$ and $f \gg g$.

Given $A, B \subseteq \mathbb{Z}$, we write

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

When $A = \{x\}$, we simply write $x + B := \{x\} + B$ and $x \cdot B := \{x\}B$. Given $A \subseteq \mathbb{Z}$, let

$$kA := \{a_1 + \cdots + a_k \mid a_1, \ldots, a_k \in A\},$$

for integer $k \geq 2$. For example, $2 \cdot \mathbb{N}$ denotes the set of even natural numbers, while $2\mathbb{N}$ denotes $\mathbb{N} + \mathbb{N}$ which is still $\mathbb{N}$. Using this notation, a set A is (k, ℓ)-*sum-free* if $kA \cap \ell A = \varnothing$.

We also define the restricted sums. Let

$$\begin{aligned} A \widehat{+} B &:= \{a + b \mid a \in A, b \in B, a \neq b\}, \\ \widehat{kA} &:= \{a_1 + \cdots + a_k \mid a_1, \ldots, a_k \in A, \text{ all of them are distinct}\}. \end{aligned}$$

Thus a set A is *restricted* (k, ℓ)-*sum-free* if $\widehat{kA} \cap \widehat{\ell A} = \varnothing$.

Let $f : \mathbb{Z} \to \mathbb{C}$ be a function. Define $\widehat{f} : \mathbb{T} \to \mathbb{C}$, where $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ is the 1-dimensional torus, and for every $r \in \mathbb{T}$,

$$\widehat{f}(r) = \sum_x f(x) e(-rx),$$

where $e(\theta) = e^{2\pi i \theta}$. By Fourier Inversion, for every $x \in \mathbb{Z}$,

$$f(x) = \int_{\mathbb{T}} \widehat{f}(r) e(rx) dr.$$

Let $\mu : \mathbb{N}^{>0} \to \mathbb{C}$ be the *Möbius function*. Recall that μ is supported on the square-free integers, and $\mu(n) = (-1)^{\omega(n)}$ when n is square-free, where $\omega(n)$ counts the number of distinct prime factors of n. By Inclusive-Exclusive Principle,

$$\sum_{d|n} \mu(d) = \begin{cases} 0 & \text{if } n > 1, \\ 1 & \text{if } n = 1. \end{cases}$$

3.2.2 Nonstandard analysis

We give some basic definitions in nonstandard analysis which will be used later in the proofs. For more systematic accounts we refer to [15, 54]. Let S be a set with infinitely many elements. An *ultrafilter* $\mathcal{U}$ on S is a collection of subsets of S, such that the characteristic function $\mathbb{1}_{\mathcal{U}} : 2^S \to \{0, 1\}$ is a finitely additive $\{0, 1\}$-valued probability measure on S. An ultrafilter is *principal* if it consists of all sets containing some element $s \in S$. Let βS denotes the collection of all ultrafilters. One can embed S into βS, by mapping $x \in S$ to the principal ultrafilter generated by x. By a standard application of Zorn's Lemma, $\beta S \setminus S$ is non-empty.

Fix $\mathcal{U} \in \beta\mathbb{N} \setminus \mathbb{N}$, and let M_n be a structure for each $n \in \mathbb{N}$. The *ultraproduct* $\prod_{n\to\mathcal{U}} M_n$ is a space consists of all ultralimits $\lim_{n\to\mathcal{U}} x_n$ of sequences x_n defined in M_n, with $\lim_{n\to\mathcal{U}} x_n = \lim_{n\to\mathcal{U}} y_n$ if two sequences $\{x_n\}$ and $\{y_n\}$ agree on a set in $\mathcal{U}$. Let ${}^*\mathbb{R} := \prod_{n\to\mathcal{U}} \mathbb{R}$ be the hyperreal field. Every finite hyperreal number $\xi \in {}^*\mathbb{R}$ is infinitely close to a unique real number $r \in \mathbb{R}$, called the *standard part* of ξ. In this case, we use the notation $r = \mathrm{st}(\xi)$.

Given a sequence of finite non-empty sets F_n, let $\mu_n(X) = |X \cap F_n|/|F_n|$ be a uniform probability measure. Let $F = \prod_{n\to\mathcal{U}} F_n$ be an ultraproduct. The *Loeb measure* [125] μ_L on F is the unique probability measure on the σ-algebra generated by the Boolean algebra of internal subsets of F, such that when $X = \prod_{n\to\mathcal{U}} X_n$ is an internal subset of F, we have

$$\mu_L(X) = \mathrm{st}\Big(\lim_{n\to\mathcal{U}} \mu_n(X_n)\Big).$$

3.2.3 Determinants of certain matrices

We make use of the following lemma several times in the later proofs, which records a fact about two special matrices.

Lemma 3.5. *Let $\theta_1, \dots, \theta_n \in \mathbb{R}$. Consider two matrices*

$$A_n = \begin{pmatrix} \sin\theta_1 & \sin\theta_2 & \cdots & \sin\theta_n \\ \sin 2\theta_1 & \sin 2\theta_2 & \cdots & \sin 2\theta_n \\ \cdots & \cdots & \cdots & \cdots \\ \sin n\theta_1 & \sin n\theta_2 & \cdots & \sin n\theta_n \end{pmatrix},$$

and

$$B_n = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \cos\theta_1 & \cos\theta_2 & \cdots & \cos\theta_n \\ \cos 2\theta_1 & \cos 2\theta_2 & \cdots & \cos 2\theta_n \\ \cdots & \cdots & \cdots & \cdots \\ \cos(n-1)\theta_1 & \cos(n-1)\theta_2 & \cdots & \cos(n-1)\theta_n \end{pmatrix}.$$

Then we have the formula:

$$\det(A_n) = 2^{n-1}\Big(\prod_{k=1}^{n} \sin\theta_k\Big)\det(B_n); \tag{3.3}$$

and

$$\det(B_n) = 2^{(n-1)(n-2)/2} \prod_{1\le k<l\le n} (\cos\theta_l - \cos\theta_k). \tag{3.4}$$

As a result,

$$\det(A_n) = 2^{n(n-1)/2}\Big(\prod_{k=1}^{n} \sin\theta_k\Big) \prod_{1\le k<l\le n} (\cos\theta_l - \cos\theta_k).$$

Proof. For $k = 1, 2, \dots n-1$, we subtract the k-th row from the $(k+1)$-th row in A_n, and use the basic trigonometric identities so that

$$\begin{aligned}
\det(A_n) &= \det\begin{pmatrix} 2\sin\frac{\theta_1}{2}\cos\frac{\theta_1}{2} & \cdots & 2\sin\frac{\theta_n}{2}\cos\frac{\theta_n}{2} \\ \cdots & \cdots & \cdots \\ 2\sin\frac{\theta_1}{2}\cos\frac{(2n-1)\theta_1}{2} & \cdots & 2\sin\frac{\theta_n}{2}\cos\frac{(2n-1)\theta_n}{2} \end{pmatrix} \\
&= 2^n\Big(\prod_{k=1}^{n}\sin\frac{\theta_k}{2}\Big)\det\begin{pmatrix} \cos\frac{\theta_1}{2} & \cdots & \cos\frac{\theta_n}{2} \\ \cdots & \cdots & \cdots \\ \cos\frac{(2n-1)\theta_1}{2} & \cdots & \cos\frac{(2n-1)\theta_n}{2} \end{pmatrix} \\
&=: C_n\det(B_n').
\end{aligned}$$

For $k = 1, 2, \dots n-1$, we add the k-th row to the $k+1$-th row in B_n', and use the basic trigonometric identities again so that

$$\begin{aligned}
\det(B_n') &= \det\begin{pmatrix} \cos\frac{\theta_1}{2} & \cdots & \cos\frac{\theta_n}{2} \\ 2\cos\frac{\theta_1}{2}\cos\theta_1 & \cdots & 2\cos\frac{\theta_n}{2}\cos\theta_n \\ \cdots & \cdots & \cdots \\ 2\cos\frac{\theta_1}{2}\cos(n-1)\theta_1 & \cdots & 2\cos\frac{\theta_n}{2}\cos(n-1)\theta_n \end{pmatrix} \\
&= 2^{n-1}\Big(\prod_{k=1}^{n}\cos\frac{\theta_k}{2}\Big)\det(B_n)
\end{aligned}$$

Combining the calculations above we prove (3.3).

As for (3.4), we let T_n be the Chebyshev polynomial

$$T_n(x) = \sum_{k=0}^{\lfloor n/2\rfloor}\binom{n}{2k}(x^2-1)^k x^{n-2k}.$$

Thus, we have $T_n(\cos x) = \cos nx$. The coefficient of the leading term, x^n in $T_n(x)$ would be

$a_n = 2^{n-1}$. Combining this fact and several elementary row operations, we get

$$\begin{aligned}\det(B_n) &= 2^{(n-1)(n-2)/2} \det \begin{pmatrix} 1 & \cdots & 1 \\ \cos\theta_1 & \cdots & \cos\theta_n \\ \cdots & \cdots & \cdots \\ (\cos\theta_1)^{n-1} & \cdots & (\cos\theta_n)^{n-1} \end{pmatrix} \\ &= 2^{(n-1)(n-2)/2} \prod_{1\le k<l\le n} (\cos\theta_l - \cos\theta_k).\end{aligned}$$

The last equation comes from the determinant formula for Vandermonde martix. $\square$

3.3 A density estimate

In this section, we prove a generalization of the McGehee–Pigno–Smith theorem [135], based on the ideas given by Bourgain [28]. Recall that the weak Littlewood problem [88] is to ask to estimate

$$I(N) := \min_{A\subseteq\mathbb{Z},|A|=N} \int_{\mathbb{R}/\mathbb{Z}} \Big|\sum_{n\in A} e^{inx}\Big| \mathrm{d}\, x.$$

The conjecture, $I(N) \gg \log N$, is resolved by McGehee, Pigno, and Smith [135], and independently by Konyagin [116]. The analogous question in discrete setting is also well studied, we refer to [77, 150, 154] for the interested readers.

Let $\mathcal{N}_1$ be the set of natural numbers that only contains prime factors at least Q, where $Q \asymp (\log N)^{100}$ is a prime. We will use the following lemma from [28, Section 5].

Lemma 3.6. *Let A be a finite subset of $\mathbb{Z}^+$ with $|A| = N$. For all $R \ge 1$, we define*

$$A_R = \{m \in A : m < R\}.$$

Also, we use $\mathrm{Proj}_R \sum a_k e^{ikx}$ to denote the truncated sum $\sum_{|k|\le R} a_k e^{ikx}$. Assume $|a_n| \le 1$

and $Q > (\log N)^{20}$. *Then there is an absolute big constant* C, *such that*

$$\left\| \mathrm{Proj}_R \sum_{n \in \mathcal{N}_1, m \in A} \frac{a_n}{n} e^{imnx} \right\|_2 < CP^{-1/15} |A_R|^{1/2}.$$

Now we are able to prove our technical lemma. The proof basically follows the arguments in [135] and [28].

Lemma 3.7. *Let* $B = \{m_1, \ldots, m_M\}$ *be a finite subset of* $\mathbb{N}^{>0}$ *and let* $Q > (\log N)^{100}$. *Assume that* $w : \mathbb{N}^{>0} \to \mathbb{C}$ *is a weight. Then there exists a function* $\Phi(x)$ *with* $\|\Phi\|_\infty < 10$ *such that*

$$\left| \left\langle \sum_{j=1}^{M} e^{im_j x} w(m_j), \Phi(x) \right\rangle \right| \gg \sum_{j=1}^{M} \frac{|w(m_j)|}{j}; \tag{3.5}$$

while for any $\beta \in \mathbb{Z}$,

$$\left| \left\langle \sum_{n \in \mathcal{N}_1, m \in B} \frac{a_n}{n} e^{i\beta mnx}, \Phi(x) \right\rangle \right| \leq C(\log M)^{-2}. \tag{3.6}$$

Here c, C *are two absolute constants.*

Proof. Let k_0 be the largest natural number that $10^{6k_0} < M$. We group B into disjoint subsets $\{B_k\}_{k=0}^{k_0}$ such that for $0 \leq k \leq k_0 - 1$, $|B_k| = 10^{6k}$. Here $B_0 = \{m_1\}$, $B_1 = \{m_2, \ldots, m_{10^6+1}\}, \cdots$, and $B_{k_0} = A \setminus (\bigcup_{k \leq k_0 - 1} B_k)$. From the construction we know $|B_{k_0}| \asymp 10^{6k_0}$. Let $\tau : \mathbb{N}^{>0} \to \mathbb{S}^1$ be the argument function that $\tau(m)w(m) \geq 0$. For each B_k, we define

$$\widetilde{P}_k = \frac{1}{|B_k|} \sum_{m \in B_k} e^{imx} \tau(m).$$

Let $I_k = [a_k, b_k]$ be the interval with $a_k = \min\{m : m \in B_k\}$, $b_k = \max\{m : m \in B_k\}$, and let ξ_k be the center of I_k. We also define

$$P_k = \widetilde{P}_k * \left(e^{i\xi_k x} F_{|I_k|} \right),$$

where $F_C = \sum_{|m|\le C} \frac{C-|m|}{C} e^{imx}$ is the C-Féjer kernel. Consequently,

$$\operatorname{supp}(\widehat{P_k}) = \operatorname{supp}(\widehat{\widetilde{P_k}}) \subset I_k, \tag{3.7}$$

and for any $m \in B_k$

$$w(m)\widehat{P_k}(m) > 10^{-6k-1}|w(m)|.$$

This shows that the functions P_k are good test functions. However, the function $\sum_k P_k(x)$ has one drawback: It is not distributed evenly on the torus. That is, the L^∞-norm $\sum_k P_k(x)$ is comparably large.

To overcome this difficulty, for each P_k, we construct a function Q_k serving as a "compensator". Specifically, let $\mathcal{H}$ be the Hilbert transform in $L^2(\mathbb{R}/\mathbb{Z})$ that $\widehat{\mathcal{H}f}(n) = -i\operatorname{sgn}(n)\widehat{f}(n)$, so that when f is a real-valued function, $\mathcal{H}f$ is also real-valued. We define

$$Q_k = \left(e^{-(|\tilde{P}_k| - i\mathcal{H}[|\tilde{P}_k|])}\right) * F_{|I_k|}. \tag{3.8}$$

Since the Fourier series of $e^{-(|\tilde{P}_k| - i\mathcal{H}[|\tilde{P}_k|])}$ is supported in non-positive integers,

$$\operatorname{supp}\big(\widehat{Q}_k\big) \subset [-|I_k|, 0]. \tag{3.9}$$

Using the inequality that $|e^{-z} - 1| \le |z|$ if $z \in \mathbb{C}$ and $\operatorname{Re}(z) \ge 0$, we can easily prove

$$\|1 - Q_k\|_2 \le \|\tilde{P}_k\|_2 + \big\|\mathcal{H}[|\tilde{P}_k|]\big\|_2 < 2|B_k|^{-1/2}. \tag{3.10}$$

Thus, Q_k is approximately the identical function. In fact, $|Q_k|$ is relatively small when $|P_k|$ is relatively large, so Q_k can help us "mollify" the function P_k.

We will use the functions P_k, Q_k to construct our test function Φ. In specific, we set

$\Phi_0 = P_0$ and set

$$\Phi_k = Q_k \Phi_{k-1} + P_k, \qquad 1 \le k \le k_0. \tag{3.11}$$

Define $\Phi = \Phi_{k_0}$, which has the explicit formula

$$\Phi = P_{k_0} + P_{k_0-1} Q_{k_0} + P_{k_0-2} Q_{k_0-1} Q_{k_0} + \cdots + P_0 Q_1 \cdots Q_{k_0}. \tag{3.12}$$

We claim $\|\Phi\|_\infty < 10$. To see this, we first recall the basic inequality: $\frac{a}{10} + e^{-a} \le 1$ if $a \ge 0$. Then, observing $|P_0| = 1$ and

$$\Big\| \frac{1}{10}|P_k| + |Q_k| \Big\|_\infty \le \Big\| \big(\frac{1}{10}|\widetilde{P}_k| + e^{-|\widetilde{P}_k|} \big) * F_{|I_k|} \Big\|_\infty \le \Big\| \frac{1}{10}|\widetilde{P}_k| + e^{-|\widetilde{P}_k|} \Big\|_\infty \le 1,$$

we argue inductively using (3.11) to conclude our claim.

Next, we will verify (3.5). We will prove that for any $m \in B_k$,

$$\big| \widehat{\Phi}(m) - \widehat{P}_k(m) \big| \le 10^{-1} \big| \widehat{P_k}(m) \big| = \frac{1}{10|B_k|}. \tag{3.13}$$

In fact, using the support condition (3.9) and the equation (3.12), we have

$$\widehat{\Phi}(m) - \widehat{P}_k(m) = \widehat{P}_{k_0}(m) + \widehat{P_{k_0-1}} * \widehat{Q_{k_0}}(m) + \cdots + \widehat{P_k} * (1 - (Q_{k_0} \cdots Q_{k+1})^\wedge)(m),$$

which, combining the support condition of $\widehat{P_k}$ in (3.7), equals to

$$\sum_{j=k}^{k_0-1} \widehat{P}_j * (1 - (Q_{k_0} \cdots Q_{j+1})^\wedge)(m).$$

We estimate the above quantity using the equality

$$1 - Q_{k+1} \cdots Q_{k_0} = (1 - Q_{k+1}) + Q_{k+1}(1 - Q_{k+2}) + \cdots + (1 - Q_{k_0}) Q_{k+1} \ldots Q_{k_0-1},$$

so that

$$|\widehat{\Phi}(m) - \widehat{P}_k(m)| = \left| \sum_{j=k}^{k_0-1} \widehat{P}_j * (1 - (Q_{k_0} \cdots Q_{j+1})^\wedge)(m) \right| \le \sum_{j=k}^{k_0-1} \|P_j\|_2 \sum_{l=j}^{k_0-1} \|1 - Q_l\|_2.$$

Since $\|P_j\|_2 \le |B_j|^{-1/2}$ and since (3.10), the right hand side of the above inequality can be bounded as

$$\sum_{j=k}^{k_0-1} \|P_j\|_2 \sum_{l=j}^{k_0-1} \|1 - Q_{l+1}\|_2 \le 2 \sum_{j=k}^{k_0-1} 10^{-3j} \sum_{l=j}^{k_0-1} 10^{-3(l+1)},$$

which implies what we need that

$$|\widehat{\Phi}(m) - \widehat{P}_k(m)| \le 10^{-3k-2} \le 10^{-1} |\widehat{P_k}(m)|.$$

As a consequence of (3.13), for any $m \in B_k$,

$$\mathrm{Re}(w\widehat{\Phi})(m) > \frac{1}{2} w(m) \widehat{P}_k(m) \ge 10^{-6k-1} |w(m)|.$$

We use the above inequality to sum up all $m \in B$ to get

$$\left| \left\langle \sum_{j=1}^{M} e^{im_j x} w(m_j), \Phi(x) \right\rangle \right| \ge \sum_{j=1}^{M} \mathrm{Re}(w\widehat{\Phi})(m_j) \gg \sum_{j=1}^{M} \frac{|w(m_j)|}{j},$$

and this gives (3.5).

Finally, we remark that the proof of (3.6) is given in [108]. At this point, we complete the proof of the lemma. □

As an application of Lemma 3.7, we have the following corollary:

Corollary 3.8. *Let $B = \{m_1, \ldots, m_M\}$ be a finite subset of $\mathbb{N}^{>0}$ and let $Q > (\log N)^{100}$. Recall that $\mathcal{N}$ is the set of natural numbers that only contains prime factors at least Q.*

Assume $|a_n| \leq 1$. Then for any $\Gamma \subset \mathbb{Z}$ with $|\Gamma| \leq \log M$, we have

$$\Big\| \sum_{j=1}^{M} e^{im_j x} w(m_j) + \sum_{n \in \mathcal{N}_1, m \in B} \Big(\sum_{\beta \in \Gamma} \frac{a_n}{n} e^{i\beta m n x} \Big) \Big\|_1 \geq c \sum_{j=1}^{M} \frac{|w(m_j)|}{j} - o(1).$$

Proof. We apply Lemma 3.7 to obtain a function $\Phi(x)$ satisfying (3.5) and (3.6). Then

$$\begin{aligned} & \Big\| \sum_{j=1}^{M} e^{im_j x} w(m_j) + \sum_{n \in \mathcal{N}_1, m \in B} \Big(\sum_{\beta \in \Gamma} \frac{a_n}{n} e^{i\beta m n x} \Big) \Big\|_1 \|\Phi\|_\infty \\ & \geq \Big| \Big\langle \sum_{j=1}^{M} e^{im_j x} w(m_j), \Phi(x) \Big\rangle \Big| - \sum_{\beta \in \Gamma} \Big| \Big\langle \sum_{n \in \mathcal{N}_1, m \in B} \frac{a_n}{n} e^{i\beta m n x}, \Phi(x) \Big\rangle \Big| \\ & > c \sum_{j=1}^{M} \frac{|w(m_j)|}{j} - o(1), \end{aligned}$$

as desired. □

3.4 (k, ℓ)-sum-free open sets in the torus

In this section, we use $\mu_{\mathbb{T}}$ as the Haar probability measure on $\mathbb{T}$.

Proposition 3.9. *Let $A \subseteq \mathbb{T}$ be a (k, ℓ)-sum-free open set. Then $\mu_{\mathbb{T}}(A) \leq \frac{1}{k+\ell}$.*

Proof. Since A is (k, ℓ)-sum-free, we have $kA \cap \ell A = \varnothing$. In particular, $\mu_{\mathbb{T}}(kA) + \mu_{\mathbb{T}}(\ell A) \leq 1$. By Kneser's inequality [115],

$$(k + \ell)\mu_{\mathbb{T}}(A) \leq \mu_{\mathbb{T}}(kA) + \mu_{\mathbb{T}}(\ell A) \leq 1,$$

which implies that $\mu_{\mathbb{T}}(A) \leq 1/(k + \ell)$. □

Next, we construct some largest (k, ℓ)-sum-free open sets in $\mathbb{T}$. When $k - \ell \geq 2$, our construction is asymmetric, which will help us get a better lower bound on $\mathcal{M}_{(k,\ell)}(N)$. We will discuss this in details in the next section.

Lemma 3.10. *Let k, ℓ be two positive integers and $k > \ell$. For every integer $t \in [k-\ell]$, set $\Omega_t = \left(\frac{t-1}{k-\ell} + \frac{\ell}{k^2-\ell^2}, \frac{t-1}{k-\ell} + \frac{k}{k^2-\ell^2}\right)$. Then Ω_t is (k,ℓ)-sum-free.*

Lemma 3.10 is easy to verify, and we omit the details here. When $k = \ell + 1$, the following observation shows that all the possible (k,ℓ)-sum-free open sets with maximum measure are symmetric. Thus one cannot apply the method used in the next section to improve the lower bound for the cases $k = \ell + 1$.

Lemma 3.11. *Let $k = \ell + 1$. Suppose $A \subseteq \mathbb{T}$ is a maximum (k,ℓ)-sum-free open set. Then A is symmetric.*

Proof. Since $k = \ell + 1$, A is (k,ℓ)-sum-free implies that $(\ell A - \ell A) \cap A = \varnothing$. Hence $A \subseteq \mathbb{T} \setminus (\ell A - \ell A)$. By Kneser's inequality,

$$\mu_{\mathbb{T}}(\mathbb{T} \setminus (\ell A - \ell A)) \leq 1 - 2\ell \mu_{\mathbb{T}}(A).$$

By Proposition 3.9, $\mu_{\mathbb{T}}(A) = \frac{1}{2\ell+1}$. Thus $A = \mathbb{T} \setminus (\ell A - \ell A)$, and this implies that A is symmetric. $\square$

Using the argument by Erdős [59], Lemma 3.10 is able to give us the following lower bound on the maximum (k,ℓ)-sum-free subsets of any set of N integers, which proves Theorem 3.3 (i).

Proposition 3.12. *Let k, ℓ be positive integers and $k > \ell$. Then for every $A \subseteq \mathbb{N}^{>0}$ of size N, A contains a (k,ℓ)-sum-free subsets of size at least $\frac{1}{k+\ell}N$.*

Proof. Let Ω_t be as in Lemma 3.10, and let $\mathbb{1}_\Omega$ be the characteristic function of Ω in $\mathbb{T}$. Thus by Fubini's Theorem,

$$\int_{\mathbb{T}} \sum_{n \in A} \mathbb{1}_\Omega(nx) d\mu_{\mathbb{T}}(x) = \sum_{n \in A} \int_{\mathbb{T}} \mathbb{1}_\Omega(nx) d\mu_{\mathbb{T}}(x) = \frac{N}{k+\ell}.$$

Therefore, by Pigeonhole principle, there exists $x \in \mathbb{T}$ such that

$$|\{n \in A \mid nx \in \Omega\}| \geq \frac{N}{k+\ell},$$

finishes the proof. $\square$

3.5 Lower Bounds

Let k, ℓ be two positive integers with $k - \ell \geq 2$. Let $I = \{1, \ldots, k-\ell\}$ be the index set. Set

$$\Omega_t = \Big(\frac{t-1}{k-\ell} + \frac{\ell}{k^2-\ell^2}, \frac{t-1}{k-\ell} + \frac{k}{k^2-\ell^2}\Big),$$

for every $t \in I$. Let $\mathbb{1}_{\Omega_t}$ be the indicator function of Ω_t. Given $A \subseteq \mathbb{N}^{>0}$ of size N, let $\mathcal{M}(A)$ be the size of the maximum (k, ℓ)-sum-free subset of A. We have

$$\mathcal{M}(A) \geq \max_{x \in \mathbb{T}} \sum_{n \in A} \mathbb{1}_{\Omega_t}(nx), \tag{3.14}$$

since Ω_t is (k, ℓ)-sum-free for every t. Then

$$\max_{x \in \mathbb{T}} \sum_{n \in A} \mathbb{1}_{\Omega_t}(nx) = \frac{N}{k+\ell} + \max_{x \in \mathbb{T}} \sum_{n \in A} \Big(\mathbb{1}_{\Omega_t} - \frac{1}{k+\ell}\Big)(nx), \tag{3.15}$$

for every $t \in I$. We introduce a balanced function $f_t : \mathbb{T} \to \mathbb{C}$ defined by $f_t = \mathbb{1}_{\Omega_t} - \frac{1}{k+\ell}$. By orthogonality of characters we have

$$\widehat{f_t}(n) = \begin{cases} 0 & \text{if } n = 0, \\ \widehat{\mathbb{1}_{\Omega_t}}(n) & \text{else.} \end{cases}$$

By Fourier inversion, when $n > 0$,

$$\begin{aligned}\widehat{f_t}(n) &= \int_{\mathbb{T}} \mathbb{1}_{\Omega_t}(x) e(-nx) d\mu(x) \\ &= \frac{1}{2\pi i n}\Big(- e\big(- \frac{(t-1)n}{k-\ell} - \frac{nk}{k^2-\ell^2}\big) + e\big(- \frac{(t-1)n}{k-\ell} - \frac{n\ell}{k^2-\ell^2}\big)\Big).\end{aligned}$$

Simplify $\widehat{f_t}(n)$ as

$$\begin{aligned}\widehat{f_t}(n) &= \frac{1}{2\pi n} e\Big(\frac{(2t-1)n}{2(k-\ell)}\Big)\Big(\sin\big(\frac{2kn\pi}{k^2-\ell^2} - \frac{\pi n}{k-\ell}\big) - \sin\big(\frac{2\ell n\pi}{k^2-\ell^2} - \frac{\pi n}{k-\ell}\big)\Big) \\ &= \frac{1}{\pi n} e\Big(\frac{(2t-1)n}{2(k-\ell)}\Big) \sin\Big(\frac{n\pi}{k+\ell}\Big).\end{aligned}$$

Hence, for every $t \in I$ we have

$$f_t(x) = \sum_{n \neq 0} \widehat{f_1}(n) e(nx) = \sum_{n \neq 0} \frac{1}{\pi n} e\Big(\frac{(2t-1)n}{2(k-\ell)}\Big) \sin\Big(\frac{n\pi}{k+\ell}\Big) e(nx).$$

Let $F(x) := \sum_{t \in I} f_t(x)$. The sine terms cancel when summing up t as

$$\sum_{t=1}^{k-\ell} \sin\Big(\frac{(2t-1)n\pi}{k-\ell}\Big) = 0,$$

so we get

$$F(x) = \frac{1}{\pi} \sum_{n \geq 1} \frac{1}{n} \sin\Big(\frac{n\pi}{k+\ell}\Big) \alpha(n) \big(e(nx) + e(-nx)\big),$$

where $\alpha(n) : \mathbb{Z} \to \mathbb{C}$ is defined by

$$\alpha(n) = \sum_{t \in I} \cos\Big(\frac{(2t-1)n\pi}{k-\ell}\Big) = \begin{cases} 0 & \text{when } (k-\ell) \nmid n, \\ (-1)^s (k-\ell) & \text{when } n = (k-\ell)s. \end{cases}$$

Therefore, we have

$$F(x) = \frac{2}{\pi} \sum_{n \geq 1} \frac{(-1)^n}{n} \sin\left(\frac{(k-\ell)n\pi}{k+\ell}\right) \cos(2\pi(k-\ell)nx). \tag{3.16}$$

In the rest of the section, we let $k - \ell$ be an even integer. Set

$$I_1 = \{1, \ldots, (k-\ell)/2\}, \quad I_2 = \{(k-\ell)/2 + 1, \ldots, k - \ell\}.$$

We define a $(\frac{k-\ell}{2} \times \frac{k-\ell}{2})$-matrix $D = (d_{ij})$, such that

$$d_{ij} = \sin\left(\frac{i(2j-1)\pi}{k-\ell}\right)$$

for every $i, j \in I_1$.

Let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{(k-\ell)/2})$ be a vector. By Lemma 3.5, there is $\boldsymbol{\lambda} \in \mathbb{R}^{(k-\ell)/2}$, with $|\lambda_i| \leq k^k$, such that $D\boldsymbol{\lambda}^T = (0, \ldots, 0, 1)^T$. Fix this $\boldsymbol{\lambda}$, and let

$$G(x) = \sum_{j \in I_1} \lambda_j f_j(x) - \sum_{t \in I_2} \lambda_{k-\ell+1-t} f_t(x).$$

Observe that for any $n \in \mathbb{N}^{>0}$,

$$\sum_{t \in I_1} \lambda_t \cos\left(\frac{(2t-1)n\pi}{k-\ell}\right) = \sum_{t \in I_2} \lambda_{k-\ell+1-t} \cos\left(\frac{(2t-1)n\pi}{k-\ell}\right).$$

As a result, we have

$$G(x) = \frac{1}{\pi} \sum_{n \geq 1} \frac{1}{n} \sin\left(\frac{n\pi}{k+\ell}\right) \beta(n) \sin(2\pi n x),$$

where

$$\beta(n) = \sum_{j\in I_2} \lambda_{k-j} \sin\left(\frac{(2j-1)n\pi}{k-\ell}\right) - \sum_{t\in I_1} \lambda_t \sin\left(\frac{(2t-1)n\pi}{k-\ell}\right)$$
$$= \begin{cases} 0 & \text{when } n \neq \frac{k-\ell}{2}(2s-1), \\ 2(-1)^{s+1} & \text{when } n = \frac{k-\ell}{2}(2s-1). \end{cases}$$

Therefore, we get

$$G(x) = \frac{2}{\pi(k-\ell)} \sum_{n\geq 1} \frac{\gamma(n)}{n} \sin\left(\frac{(k-\ell)n\pi}{2(k+\ell)}\right) \sin(\pi(k-\ell)nx), \tag{3.17}$$

where $\gamma(n) = \beta((k-\ell)n/2)$. We now split the proof into two cases.

3.5.1 Proof of Theorem 3.3 (ii)

Now we have $k = 5\ell$. On one hand, by equation (3.16), we have

$$F(x) = -\frac{\sqrt{3}}{\pi} \sum_{n\geq 1} \frac{\psi(n)}{n} \cos(8\pi\ell nx),$$

where

$$\psi(n) = \begin{cases} 1 & \text{when } n \equiv 1,2 \pmod 6, \\ -1 & \text{when } n \equiv 4,5 \pmod 6, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\psi(n)$ is *not* a multiplicative function. Using the Möbius function μ, we define a weighted Möbius function η that

$$\eta(n) = \begin{cases} \mu(n) & \text{when } n \equiv 1, 4 \pmod 6, \\ -\mu(n) & \text{when } n \equiv 2, 5 \pmod 6, \\ 0 & \text{otherwise.} \end{cases}$$

Set $P \asymp (\log N)^{100}$ a prime. Let $\mathcal{M}_1$ be the set of square-free integers such that for every $n \in \mathcal{M}_1$ we have $3 \nmid n$, all the prime factors of n are at most P, and we further require that $1 \in \mathcal{M}_1$. Then, we have

$$\sum_{m \in \mathcal{M}_1} \frac{\eta(m)}{m} F(mx) = -\frac{\sqrt{3}}{\pi} \sum_{n \geq 1} \frac{1}{n} \cos(8\pi \ell n x) \sum_{m \in \mathcal{M}_1, m|n} \eta(m) \psi\Big(\frac{n}{m}\Big),$$

where

$$\sum_{m \in \mathcal{M}_1, m|n} \eta(m)\psi\Big(\frac{n}{m}\Big) = \sum_{\substack{m \in \mathcal{M}_1, m|n \\ 2 \nmid m}} \eta(m)\psi\Big(\frac{n}{m}\Big) + \sum_{\substack{m \in \mathcal{M}_1, m|n \\ 2 | m}} \eta(m)\psi\Big(\frac{n}{m}\Big) =: I_1 + I_2.$$

Note that $I_1 + I_2 = 0$ when $3 \mid n$. Also, recall that $\mathcal{N}$ is the set defined in Section 3.3 that contains integers only having prime factors at least P. It follows that for any odd integer $n \notin \mathcal{N}$ with $3 \nmid n$,

$$\sum_{m \in \mathcal{M}_1, m|n} \eta(m)\psi\Big(\frac{n}{m}\Big) = \psi(n) \sum_{m \in \mathcal{M}_1, m|n} \mu(m) = 0. \tag{3.18}$$

As a consequence, when n is an odd integer with $n \notin \mathcal{N}$, we have $I_1 = I_2 = 0$, unless $n = 1$; When n is an even integer with $n/2 \notin \mathcal{N}$ and $n \neq 2^d$, we have $I_1(n) = 0$ and

$I_2(n) = I_1(n/2) = 0$. When $n = 2^d$, we have

$$I_1 + I_2 = \psi(2^d) + \psi(2^{d-1}) = \begin{cases} 2 & \text{when } d = 1, \\ 0 & \text{when } d > 1. \end{cases}$$

Therefore,

$$\sum_{m\in\mathcal{M}_1} \frac{\eta(m)}{m} F\Big(\frac{mx}{4}\Big) = -\frac{\sqrt{3}}{\pi}\Bigg(\cos(2\pi\ell x) + \cos(4\pi\ell x) + \sum_{n\in\mathcal{N}} \frac{\psi(n)}{n}\cos(2\pi\ell n x) + \sum_{n\in 2\cdot\mathcal{N}} \frac{\psi(n)+\psi(n/2)}{n}\cos(2\pi\ell n x)\Bigg),$$

On the other hand, by equation (3.17), we get

$$G(x) = \frac{\sqrt{3}}{\pi}\sum_{n\geq 1}\frac{\pi(n)}{n}\sin(4\pi\ell n x),$$

where

$$\pi(n) = \begin{cases} 1 & \text{when } n \equiv \pm 1 \pmod{12}, \\ -1 & \text{when } n \equiv \pm 5 \pmod{12}, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\pi(n)$ is a multiplicative function. Let $\mathcal{M}$ be the set of square-free integers such that for every $n \in \mathcal{M}$, all the prime factors of n are at most P, and $1 \in \mathcal{M}$. Thus by the basic properties of the Möbius function, we have

$$\begin{aligned}\sum_{m\in\mathcal{M}} \frac{\mu(m)\pi(m)}{m} G(mx) &= \frac{\sqrt{3}}{\pi(k-\ell)}\sum_{n\geq 1}\frac{\pi(n)}{n}\sin(4\pi\ell n x)\sum_{m\in\mathcal{M}, m|n}\mu(m) \\ &= \frac{\sqrt{3}}{\pi(k-\ell)}\Bigg(\sin(4\pi\ell x) + \sum_{n\in\mathcal{N}}\frac{\pi(n)}{n}\sin(4\pi\ell n x)\Bigg).\end{aligned}$$

Now we are going to apply Corollary 3.8 to obtain a lower bound of $\mathcal{M}(A)$. Let

$$\pi'(n) = \begin{cases} \pi(n) & \text{when } 2 \nmid n, \\ \pi(n/2) & \text{when } 2 \mid n \text{ and } 4 \nmid n, \\ 0 & \text{otherwise,} \end{cases}$$

and let

$$\psi'(n) = \begin{cases} \psi(n) & \text{when } 2 \nmid n, \\ \psi(n) + \psi(n/2) & \text{when } 2 \mid n \text{ and } 4 \nmid n, \\ 0 & \text{otherwise.} \end{cases}$$

We have

$$\begin{aligned}
&\log N \\
&\ll \Bigg\| \sum_{m\in A} \big(\cos(2\pi\ell mx) + \cos(4\pi\ell mx) + i\sin(2\pi\ell mx) + i\sin(4\pi\ell mx)\big) \\
&\quad + \sum_{m\in A} \sum_{n\in\mathcal{N}\cup(2\cdot\mathcal{N})} \frac{1}{n}\Big(\psi'(n)\cos(2\pi\ell mnx) + i\pi'(n)\sin(2\pi\ell mnx)\Big) \Bigg\|_{L^1(\mathbb{T})} \\
&\ll \Bigg\| \sum_{t\in\mathcal{M}_1} \frac{\eta(t)}{t} \sum_{m\in A} F\Big(\frac{tmx}{4}\Big) \Bigg\|_{L^1(\mathbb{T})} + \Bigg\| \sum_{t\in\mathcal{M}} \frac{\mu(t)\pi(t)}{t} \sum_{m\in A} G\Big(\frac{tmx}{2}\Big) \Bigg\|_{L^1(\mathbb{T})} \\
&\quad + \Bigg\| \sum_{t\in\mathcal{M}} \frac{\mu(t)\pi(t)}{t} \sum_{m\in A} G(tmx) \Bigg\|_{L^1(\mathbb{T})} \\
&\ll \sum_{t\in\mathcal{M}_1} \left|\frac{\eta(t)}{t}\right| \Bigg\| \sum_{m\in A} F(mx) \Bigg\|_{L^1(\mathbb{T})} + 2\sum_{t\in\mathcal{M}} \left|\frac{\mu(t)\pi(t)}{t}\right| \Bigg\| \sum_{m\in A} G(mx) \Bigg\|_{L^1(\mathbb{T})} \\
&\ll \prod_{p\le P} \Big(1+\frac{1}{p}\Big) \Bigg(\sum_{t=1}^{k-\ell} \Bigg\| \sum_{m\in A} f_t(mx) \Bigg\|_{L^1(\mathbb{T})} + 2\sum_{t\in I_1\cup I_2} \lambda_t \Bigg\| \sum_{m\in A} f_t(mx) \Bigg\|_{L^1(\mathbb{T})} \Bigg).
\end{aligned}$$

By Mertens' estimates we get

$$\prod_{p\le P}\Big(1+\frac{1}{p}\Big) \ll \log P \asymp \log\log N.$$

Hence there is $t \in I$ such that $\big\|\sum_{m\in A} f_t(mx)\big\|_{L^1(\mathbb{T})} \gg \frac{\log N}{\log\log N}$.

Note that,

$$\int_{\mathbb{T}} \sum_{n\in A} f_t(nx)dx = 0.$$

Thus we have

$$\max_{x\in\mathbb{T}} \sum_{n\in A} f_t(nx) \ge \frac{1}{2}\Bigg\|\sum_{n\in A} f_t(nx)\Bigg\|_{L^1(\mathbb{T})}.$$

Together with (3.14) and (3.15), we get

$$\mathcal{M}(A) - \frac{N}{k+\ell} \gg \frac{\log N}{\log\log N},$$

and this proves Theorem 3.3 (ii).

3.5.2 Proof of Theorem 3.3 (iii)

Let u be an even integer, and let $t = u/2$ in this subsection. Consider the following matrix

$$X = \begin{pmatrix} \sin(\pi/u) & \sin(3\pi/u) & \cdots & \sin((2t-1)\pi/u) \\ \sin(2\pi/u) & \sin(6\pi/u) & \cdots & \sin(2(2t-1)\pi/u) \\ \cdots & \cdots & \cdots & \cdots \\ \sin(t\pi/u) & \sin(3t\pi/u) & \cdots & \sin(t(2t-1)\pi/u) \end{pmatrix}.$$

By Lemma 3.5, there is $\boldsymbol{\alpha} \in \mathbb{R}^t$, with $|\alpha_i| \le t^t$, such that $X\boldsymbol{\alpha}^T = (-1,\dots,0,0)^T$.

For each odd integer v ranging from the interval $[1,u)$, define $\mathcal{P}_v$ to be an infinite collection of pairs (k_v,ℓ_v) of positive integers such that $k_v = (u+v)\ell_v/(u-v)$. Let $F_{(k,\ell)}(x)$ and $G_{(k,\ell)}(x)$ be the function constructed in (3.16) and (3.17) with respect to the pair (k,ℓ). Note that in the current constructions, for every $(k_1,\ell_1),(k_2,\ell_2)\in\mathcal{P}_v$, we have

$$F_{(k_1,\ell_1)}\Big(\frac{x}{k_1-\ell_1}\Big)=F_{(k_2,\ell_2)}\Big(\frac{x}{k_2-\ell_2}\Big),$$

and we denote the above function by $F_v(x)$ since it only depends on v. Similarly, we also have

$$(k_1-\ell_1)G_{(k_1,\ell_1)}\Big(\frac{x}{k_1-\ell_1}\Big)=(k_2-\ell_2)G_{(k_2,\ell_2)}\Big(\frac{x}{k_2-\ell_2}\Big),$$

and we denote the above function by $G_v(x)$.

Let $\boldsymbol{F}(x)=(F_1(x),F_3(x),\dots F_{2t-1}(x))$, and we construct

$$F(x):=\boldsymbol{F}(x)\boldsymbol{\alpha}^T=\frac{2}{\pi}\sum_{n\geq 1}\frac{\Phi(n)}{n}\cos(2\pi nx),$$

where

$$\Phi(n):=\begin{cases}1 & \text{when } n\equiv 1,u-1 \pmod{2u},\\ -1 & \text{when } n\equiv u+1,-1 \pmod{2u},\\ 0 & \text{otherwise,}\end{cases}$$

since in this case $k_v-\ell_v$ is always even for every (k_v,ℓ_v) in $\mathcal{P}_v$. Note that $\Phi(n)$ is a multiplicative function. Let $\mathcal{M}$ be the set of square-free integers that only contains prime factors at most P and $1\in\mathcal{M}$, hence we have

$$\sum_{m\in\mathcal{M}}\frac{\Phi(m)\mu(m)}{m}F(mx)=\frac{2}{\pi}\Big(\cos(2\pi x)+\sum_{n\in\mathcal{N}}\frac{\Phi(n)}{n}\cos(2\pi nx)\Big).$$

Similarly, we consider the matrix

$$Y = \begin{pmatrix} \sin(\pi/2u) & \sin(3\pi/2u) & \cdots & \sin((2t-1)\pi/2u) \\ \sin(3\pi/2u) & \sin(9\pi/2u) & \cdots & \sin(3(2t-1)\pi/2u) \\ \cdots & \cdots & \cdots & \cdots \\ \sin((2t-1)\pi/2u) & \sin(3(2t-1)\pi/2u) & \cdots & \sin((2t-1)^2\pi/2u) \end{pmatrix}.$$

By Lemma 3.5, there is $\boldsymbol{\beta} \in \mathbb{R}^t$, with $|\beta_i| \leq t^t$, such that $Y\boldsymbol{\beta}^T = (1, \ldots, 0, 0)^T$. Let $\boldsymbol{G}(x) = (G_1(x), G_3(x), \ldots, G_{2t-1}(x))$, and we construct

$$G(x) := \boldsymbol{G}(x)\boldsymbol{\beta}^T = \frac{2}{\pi} \sum_{n \geq 1} \frac{\Psi(n)}{n} \sin(2\pi n x),$$

where

$$\Psi(n) := \begin{cases} 1 & \text{when } n \equiv \pm 1 \pmod{4u}, \\ -1 & \text{when } n \equiv \pm(2u-1) \pmod{4u}, \\ 0 & \text{otherwise.} \end{cases}$$

We also have $\Psi(n)$ is a multiplicative function. Hence

$$\sum_{m \in \mathcal{M}} \frac{\Psi(m)\mu(m)}{m} G(mx) = \frac{2}{\pi}\Big(\sin(2\pi x) + \sum_{n \in \mathcal{N}} \frac{\Psi(n)}{n} \sin(2\pi n x)\Big).$$

Finally, we apply Corollary 3.8. Using a similar computation employed in Section 3.5.1, we obtain that

$$\max\Big\{\Big\| \sum_{m \in A} F(mx)\Big\|_{L^1(\mathbb{T})}, \Big\| \sum_{m \in A} G(mx)\Big\|_{L^1(\mathbb{T})}\Big\} \gg \frac{\log N}{\log\log N}.$$

This implies there is an odd integer $v \in [1, u)$ such that

$$\max\Big\{\Big\| \sum_{m \in A} F_v(mx)\Big\|_{L^1(\mathbb{T})}, \Big\| \sum_{m \in A} G_v(mx)\Big\|_{L^1(\mathbb{T})}\Big\} \gg \frac{\log N}{\log\log N}.$$

Therefore, from a similar argument we used in Section 3.5.1, we can conclude that the size of the maximal (k_v, ℓ_v)-sum-free subset of A for every $(k_v, \ell_v) \in \mathcal{P}_v$ is at least

$$\frac{N}{k_v + \ell_v} + c\frac{\log N}{\log \log N}$$

for some positive c. This proves Theorem 3.3 (iii).

3.6 Structure of infinite (k, ℓ)-sum-free sets

Given $A \subseteq \mathbb{N}^{>0}$, the *upper density* of A is defined as

$$\overline{d}(A) = \limsup_{N \to \infty} \frac{|A \cap [N]|}{N}.$$

We also define the *upper density on multiples* of A by

$$\widetilde{d}(A) = \limsup_{N \to \infty} \limsup_{n \to \infty} \frac{|A \cap (N! \cdot [n])|}{n}.$$

In this section, we will prove the following theorem, which will be used in the next section when constructing the upper bound estimate for Theorem 1 (iv).

Theorem 3.13. *Suppose that $A \subseteq \mathbb{N}^{>0}$, and A is (k, ℓ)-sum-free. Then $\widetilde{d}(A) \leq \frac{1}{k+\ell}$.*

We break the proof of this theorem into three lemmas. The first lemma says that if a (k, ℓ)-sum-free set A contains a certain long arithmetic progression, then the upper density of A is bounded.

Lemma 3.14. *Let $A \subseteq \mathbb{N}^{>0}$ be a (k, ℓ)-sum-free set. Let x, s, d, m be positive integers, such that $s \in \ell A - (k-1)A$, $x + d \cdot [m] \subseteq A$, and s is in the coset $x + d \cdot \mathbb{Z}$. Then*

$$\overline{d}(A) \leq \frac{m + k + \ell - 2}{(k+\ell)m + 2(k + \ell - 2)}.$$

Proof. Since $s \in \ell A-(k-1)A$ and A is (k,ℓ)-sum-free, we have $s \notin A$. We will only consider $s \leq x$, and the case when $s \geq x+m$ follows from the same proof. Since $x+d\cdot[m] \subseteq A$, then $\big(x+d\cdot[m]\big)\cap\big(\ell A-(k-1)A\big)=\varnothing$. Thus, there is $s_0 \in x+d\cdot\mathbb{Z}$, such that $s_0 \in \ell A-(k-1)A$, and

$$\big(s_0+d\cdot[m]\big)\cap\big(\ell A-(k-1)A\big)=\varnothing. \tag{3.19}$$

Let $s_0=\sum_{i=1}^{\ell} a_i-\sum_{j=1}^{k-1} b_j$, where $a_i, b_j \in A$ for every $1 \leq i \leq \ell$ and $1 \leq j \leq k-1$.

Let $B \subseteq A$ such that

$$B := \{b \in A \mid \big(b+d\cdot[m]\big)\cap A \neq \varnothing\}.$$

Set $a_0=b_0=0$. Given integers $1 \leq u \leq k-1$ and $2 \leq v \leq \ell$, let

$$\mathcal{C}(u)=B+\sum_{j=1}^{k-u} b_j+(u-1)a_1, \qquad \mathcal{D}(v)=B+\sum_{j=0}^{\ell-v} a_j+\sum_{i=0}^{v-1} b_i,$$

and $\mathcal{C}(k)=A+(k-1)a_1$, $\mathcal{D}(1)=A+\sum_{j=1}^{\ell-1} a_j$. Let $\mathcal{F}=\{\mathcal{C}(u),\mathcal{D}(v) \mid u \in [k], v \in [\ell]\}$ be the collection of all $\mathcal{C}(u)$ and $\mathcal{D}(v)$.

Claim 2. Elements in $\mathcal{F}$ are pairwise disjoint.

Proof of Claim 2. Observe that for every $u \in [k]$ and $v \in [\ell]$, $\mathcal{C}(u)\cap\mathcal{D}(v)=\varnothing$. Otherwise, we will get $kA\cap\ell A \neq \varnothing$, contradicts that A is (k,ℓ)-sum-free. Let $u_1, u_2 \in [k]$ and $u_1 < u_2$. Suppose that $\mathcal{C}(u_1)\cap\mathcal{C}(u_2)\neq\varnothing$. Then there exist $y_1 \in B$ and $y_2 \in A$, such that

$$y_1+\sum_{j=k-u_2+1}^{k-u_1} b_j=y_2+(u_2-u_1)a_1.$$

Then

$$s_0=\sum_{i=1}^{\ell} a_i-\sum_{j=1}^{k-1} b_j$$

$$= y_1 + \sum_{i=2}^{\ell} a_i - y_2 - (u_2 - u_1 - 1)a_1 - \sum_{j\in[1,k-u_2]\cup[k-u_1+1,k-1]} b_j.$$

Since $y_1 \in B$, thus there is $r \in [m]$ such that $y_1 + rd \in A$. This implies $s_0 + rd \in \ell A - (k-1)A$, contradicts (3.19).

Suppose $\mathcal{D}(v_1) \cap \mathcal{D}(v_2) \neq \varnothing$ for some $v_1, v_2 \in [\ell]$ and $v_1 < v_2$. Similarly, there exist $y_1 \in A$ and $y_2 \in B$, such that

$$y_1 + \sum_{j=\ell-v_2+1}^{\ell-v_1} a_j = y_2 + \sum_{i=v_1}^{v_2-1} b_i.$$

Let $c_0 = 0$, and let $c_1, \ldots, c_{v_2-v_1-1} \in A$ if $v_2 > v_1 + 1$. Therefore

$$s_0 = y_2 + \sum_{j\in[0,\ell-v_2]\cup[\ell-v_1+1,\ell]} a_j + \sum_{t=0}^{v_2-v_1-1} c_t - y_1 - \sum_{i\in[0,v_1-1]\cup[v_2,k-1]} b_i - \sum_{t=0}^{v_2-v_1-1} c_t.$$

Observe $y_2 \in B$ implies that there is $r \in [m]$, such that $y_2 + rd \in A$. Hence $s_0 + rd \in \ell A - (k-1)A$, which contradicts (3.19). $\bowtie$

By Claim 2, we obtain

$$(k + \ell - 2)\overline{d}(B) + 2\overline{d}(A) \leq 1. \tag{3.20}$$

On the other hand, let $\mathcal{N}(t) = A \setminus B + td$ for every $t \in [m]$, and let

$$\mathcal{G} = \Big\{A, A - (k-1)x + \sum_{i=1}^{\ell-1} a_i, \mathcal{N}(t) \;\Big|\; t \in [m]\Big\}.$$

Claim 3. Elements in $\mathcal{G}$ are pairwise disjoint.

Proof of Claim 3. Suppose there are $u, v \in [m]$, $u < v$, such that $\mathcal{N}(u) \cap \mathcal{N}(v) \neq \varnothing$. Thus we have $c \in A \setminus B$ such that $c_1 + (u-v)d \in A$, and this contradicts the assumption of B. Same conclusion holds if $A \cap \mathcal{N}(u) \neq \varnothing$. Observe that if $A \cap (A - (k-1)x + \sum_{i=1}^{\ell-1} a_i) \neq \varnothing$, it will contradict that A is (k, ℓ)-sum-free. Finally, we assume that there are $c_1, c_2 \in A$, $u \in [m]$

such that

$$c_1 + ud = c_2 - (k-1)x + \sum_{i=1}^{\ell-1} a_i.$$

Thus, $c_1 + x + ud + (k-2)x = c_2 + \sum_{i=1}^{\ell-1} a_i$. Since $x + d \cdot [m] \subseteq A$, this contradicts A is (k,ℓ)-sum-free. $\bowtie$

Thus, by Claim 3, we get

$$(m+2)\overline{d}(A) - m\overline{d}(B) \leq 1.$$

Together with (3.20), this finishes the proof. $\square$

The next lemma is a finite version of the Szemerédi Theorem [158], and we will use it to find the arithmetic progression in Lemma 3.14.

Lemma 3.15 ([158])**.** *For every $\varepsilon > 0$ and $m \in \mathbb{N}^{>0}$, there is $L = L(\varepsilon, m) > 0$ such that every set $A \subseteq \mathbb{N}^{>0}$ with $\overline{d}(A) > \varepsilon$, there exist $x \in \mathbb{N}$, $d < L$, and $x + d \cdot [m] \subseteq A$.*

Our final lemma says that a (k,ℓ)-sum-free set A with large upper density should be periodic. This structural result can be viewed as a generalization of the Łuczak–Schoen Theorem [127].

Lemma 3.16. *Let $\varepsilon > 0$. Then there is $D > 0$ such that the following holds. Let $A \subseteq \mathbb{N}^{>0}$ be a (k,ℓ)-sum-free set, and $\overline{d}(A) > \frac{1}{k+\ell} + \varepsilon$. Then A is contained in a periodic (k,ℓ)-sum-free set with period D.*

Proof. We pick $m \in \mathbb{N}^{>0}$ such that

$$\frac{m+k+\ell-2}{(k+\ell)m + 2(k+\ell-2)} < \frac{1}{k+\ell} + \varepsilon. \tag{3.21}$$

Let $L = L(\varepsilon, m)$ be as in Lemma 3.15. Let $D = L!$. Suppose the lemma fails. Let $C \subseteq \mathbb{N}^{>0}$ be a periodic set with period D, consists of all positive integers in every coset $a + D \cdot \mathbb{Z}$ for

$a \in A$. Thus C is not (k,ℓ)-sum-free. This means, there are $a_1, \ldots, a_\ell$ and $b_1, \ldots, b_k$ in C such that $\sum_{i=1}^{\ell} a_i = \sum_{j=1}^{k} b_j$. Let P be the "(k,ℓ)-sum-free" part of C. That is,

$$P = C \setminus \big(\ell C - (k-1)C\big).$$

Set $a_0 = b_0 = 0$. For every $u \in [k]$ and $v \in [\ell]$, let

$$\mathcal{M}(u) = P + \sum_{j=0}^{k-u} b_j + (u-1)a_1, \qquad \mathcal{N}(v) = P + \sum_{i=0}^{\ell-v} a_i + (v-1)b_1.$$

Let $\mathcal{F}$ be the collection of all $\mathcal{M}(u)$ and $\mathcal{N}(v)$.

Claim 4. Elements in $\mathcal{F}$ are pairwise disjoint.

Proof of Claim 4. Observe that for every $u \in [k]$ and $v \in [\ell]$, $\mathcal{M}(u) \cap \mathcal{N}(v) = \varnothing$. Otherwise there are $p_1, p_2 \in P$, such that

$$p_1 = p_2 + \sum_{i=0}^{\ell-v} a_1 + (v-1)b - \sum_{j=0}^{k-u} b_j - (u-1)a_1 \in \ell C - (k-1)C,$$

contradicts the assumption of P. Now, suppose $u_1, u_2 \in [k]$, $u_1 < u_2$, such that $\mathcal{M}(u_1) \cap \mathcal{M}(u_2) \neq \varnothing$. The case that $\mathcal{N}(v_1) \cap \mathcal{N}(v_2) \neq \varnothing$ can be proved in the same way. Thus, there exist $p_1, p_2 \in P$, such that

$$p_1 + \sum_{j=k-u_2+1}^{k-u_1} b_j = p_2 + (u_2 - u_1)a_1.$$

This implies

$$0 = \sum_{j=1}^{k} b_j - \sum_{i=1}^{\ell} a_i = p_2 + (u_2 - u_1 - 1)a_1 + \sum_{j \in [0,k-u_2] \cup [k-u_1+1,k]} b_j - \sum_{i=2}^{\ell} a_i - p_1,$$

hence $P \cap (\ell C - (k-1)C) \neq \varnothing$, contradiction. $\bowtie$

By Claim 4, we obtain that $\overline{d}(P) \le \frac{1}{k+\ell}$. This means, $\overline{d}(A \setminus P) \ge \varepsilon$. By Lemma 3.15, $A \setminus P$ contains a progression $x + d \cdot [m]$, and $d < L$. By the way we construct P, there are $s_1, \dots, s_\ell$ and $t_1, \dots, t_{k-1}$ in C such that

$$x + dm = \sum_{i=1}^{\ell} s_i - \sum_{j=1}^{k-1} t_j.$$

Hence there are $e_1, \dots, e_\ell$ and $f_1, \dots, f_{k-1}$ in A, such that for every $i \in [\ell]$ and $j \in [k-1]$, we have that $e_i \in s_i + D \cdot \mathbb{Z}$, and $f_j \in t_j + D \cdot \mathbb{Z}$. Let $s = \sum_{i=1}^{\ell} e_i - \sum_{j=1}^{k-1} f_j$, thus $s \in \ell A - (k-1)A$, and $s \in x + D \cdot \mathbb{Z}$. Since $d \mid D$, we have $s \in x + d \cdot \mathbb{Z}$. By Lemma 3.14, we have that

$$\overline{d}(A) \le \frac{m + k + \ell - 2}{(k+\ell)m + 2(k + \ell - 2)},$$

and this contradicts (3.21). □

Now we can prove the main result of this section.

Proof of Theorem 3.13. Let $A/N! := \{a \mid aN! \in A\}$. Thus $\widetilde{d}(A) > 0$ implies that $A/N!$ contains a multiple of every natural number. In particular, $A/N!$ is not contained in a periodic (k, ℓ)-sum-free set. By Lemma 3.16, $\overline{d}(A/N!) \le \frac{1}{k+\ell}$. Observe that $\widetilde{d}(A) = \limsup_{N\to\infty} \overline{d}(A/N!)$, thus $\widetilde{d}(A) \le \frac{1}{k+\ell}$. □

3.7 Upper bound constructions

Recall a *Følner sequence* in $(\mathbb{N}, \cdot)$ is any sequence $\Phi : m \mapsto \Phi_m$ of finite non-empty subsets of $\mathbb{N}$, such that for every $a \in \mathbb{N}$,

$$\lim_{m\to\infty} \frac{|\Phi_m \triangle (a \cdot \Phi_m)|}{|\Phi_m|} = 0.$$

Følner sequence has been used as some good constructions in many additive combinatorics problems, see [139, 100] for example. In this section, we will show that the sets in Følner sequence will never have large (k,ℓ)-sum-free subsets. In fact, we will prove the following theorem.

Theorem 3.17. *Let $\Phi = \{\Phi_m\}$ be a Følner sequence in $(\mathbb{N},\cdot)$. Suppose there are infinitely many m such that Φ_m has a (k,ℓ)-sum-free set of size at least $\delta|\Phi_m|$ for some positive real number $\delta \leq 1$. Then there exists a (k,ℓ)-sum-free set $A \subseteq \mathbb{N}$ such that $\widetilde{d}(A) \geq \delta$.*

Theorem 3.3 (iv) follows easily from Theorem 3.17 and Theorem 3.13.

Proof of Theorem 3.17. By passing to a subsequence, we may assume for every $\Phi_m \in \Phi$, there is a (k,ℓ)-sum-free set $\phi_m \subseteq \Phi_m$, such that $|\phi_m|/|\Phi_m| \geq \delta$. Let $\beta\mathbb{N}$ be the collection of ultrafilters, and let $\mathfrak{U} \in \beta\mathbb{N} \setminus \mathbb{N}$ be a non-principal ultrafilter. Let ${}^*\mathbb{Z} = \prod_{m\to\mathfrak{U}} \mathbb{Z}$ be the ultrapower of $\mathbb{Z}$. Let Σ be the Loeb σ-algebra on ${}^*\mathbb{Z}$. Let μ_L be the Loeb measure induced by μ_m, where $\mu_m(X) = |X \cap \Phi_m|/|\Phi_m|$ for every $X \subseteq \mathbb{Z}$. Let $\phi = \prod_{m\to\mathfrak{U}} \phi_m$ be the internal set. Then by Łoś's Theorem, ϕ is (k,ℓ)-sum-free, and

$$\mu_L(\phi) = \operatorname{st}\left(\lim_{m\to\mathfrak{U}} \mu_m(\phi_m)\right) \geq \delta.$$

Claim 5. For every $a \in \mathbb{N}$, the map $x \mapsto ax$ is Σ-measurable and μ_L-preserving.

Proof of Claim 5. Note that $x \mapsto ax$ sends internal sets to internal sets, thus it is Σ-measurable. For every $X \subseteq \mathbb{Z}$, since

$$\mu_m(X) - \mu_m(a\cdot X) = \frac{|X\cap\Phi_m| - |(a\cdot X)\cap\Phi_m|}{|\Phi_m|} \leq \frac{|(a\cdot\Phi_m)\triangle\Phi_m|}{|\Phi_m|} \to 0$$

as $m \to \infty$, it preserves the Loeb measure μ_L. ⋈

Now we are able to apply the probabilistic argument used in the proof of Proposition 3.9 on the set ϕ. For every $x \in {}^*\mathbb{Z}\setminus\{0\}$, let $A_x := \{a \in \mathbb{N} \mid ax \in \phi\}$. Thus A_x is (k,ℓ)-sum-free. By Claim 5, $\widetilde{d}(A_x)$ is Σ-measurable on x. Suppose x is chosen uniformly at random with respect to the measure μ_L. By Fatou's Lemma,

$$\begin{aligned}\mathbb{E}(\widetilde{d}(A_x)) &\geq \limsup_{N\to\infty}\limsup_{n\to\infty}\mathbb{E}\left(\frac{|A_x\cap(N!\cdot[n])|}{n}\right)\\&=\limsup_{N\to\infty}\limsup_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\mathbb{P}(jN!x\in\phi).\end{aligned}$$

By Claim 5, we have

$$\mathbb{E}(\widetilde{d}(A_x))\geq\limsup_{N\to\infty}\limsup_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\mathbb{P}(x\in\phi)=\mu_L(\phi)\geq\delta.$$

Thus by Pigeonhole Principle, there exists a set $A_x\subseteq\mathbb{N}$ for some $x\in{}^*\mathbb{Z}\setminus\{0\}$ such that $\widetilde{d}(A_x)\geq\delta$. $\square$

3.8 Restricted (k,ℓ)-sum-free sets

In this section, we prove Theorem 3.4. Since restricted (k,ℓ)-sum-free can be expressed by first order formula, once we prove the conclusion in Theorem 3.13 also works for restricted (k,ℓ)-sum-free sets, Theorem 3.4 follows by using the same proof in Theorem 3.17. More precisely, in the proof of Theorem 3.17, if $A_x=\{a\in\mathbb{N}\mid ax\in\phi\}$ is not restricted (k,ℓ)-sum-free for some $x\in{}^*\mathbb{Z}\setminus\{0\}$, since the map $a\mapsto ax$ is injective, we also have that ϕ is not restricted (k,ℓ)-sum-free.

We first consider the analogue of Lemma 3.14 for restricted (k,ℓ)-sum-free sets. The similar argument also works here, with a different and more involved constructions of sets $\mathcal{C}(u)$, $\mathcal{D}(v)$, and $\mathcal{N}(t)$, and a more careful analysis. These new constructions will lead a slightly different structure for the large infinite restricted (k,ℓ)-sum-free sets in Lemma 3.19, compared to the non-restricted setting.

Lemma 3.18. *Let k,ℓ be positive integers, and $\ell<k\leq 2\ell+1$. Suppose $A\subseteq\mathbb{N}^{>0}$ be a restricted (k,ℓ)-sum-free set. Define $W\subseteq\mathbb{N}^{>0}$, satisfies that for every $w\in W$, there are*

ℓ distinct elements $y_1, \ldots, y_\ell \in A$, and $k-1$ distinct elements $z_1, \ldots, z_{k-1} \in A$, such that $w \neq z_i$ for $i \in [k-1]$, and $w = \sum_{j=1}^{\ell} y_j - \sum_{i=1}^{k-1} z_i$. Let x, s, d, m be integers, such that $s \in W$, $m > k+\ell$, $x + d \cdot [m] \subseteq A$, and s is in the coset $x + d \cdot \mathbb{Z}$. Then

$$\overline{d}(A) \leq \frac{m-2}{(k+\ell)(m-k-\ell)+2(k+\ell-2)}.$$

Proof. $s \in W$ implies that $s \notin A$ since A is restricted (k,ℓ)-sum-free. We only consider the case when $s < x$. Since $A \cap W = \varnothing$, there is $s_0 \in x + d \cdot \mathbb{Z}$ such that $s_0 \in W$ and $(s_0 + d \cdot [m]) \cap W = \varnothing$. Thus there are ℓ distinct elements $a_1, \ldots, a_l \in A$, and $k-1$ distinct elements $b_1, \ldots, b_{k-1} \in A$, $s_0 \neq b_j$ for every $j \in [k-1]$, and $s_0 = \sum_{i=1}^{\ell} a_i - \sum_{j=1}^{k-1} b_j$. Let $\mathcal{E}$ consists of $k-1$ distinct elements $e_1, \ldots, e_{k-1} \in A$, and all of them are disjoint from $\{a_i\}_{i=1}^{\ell}$, $\{b_j\}_{j=1}^{k-1}$, $\{s_0\}$ and $s_0 + d \cdot [m]$. Let

$$A' = A \setminus \Big(\bigcup_{i=1}^{\ell} \{a_i\} \cup \bigcup_{j=1}^{k-1} \{b_j\} \cup \mathcal{E} \cup \{s_0\} \cup (s_0 + d \cdot [m]) \Big). \tag{3.22}$$

Observe that

$$(s_0 + d \cdot [m]) \cap \{b_j\}_{j=1}^{k-1} = \varnothing, \tag{3.23}$$

since $b_j \in W$ for every $j \in [k-1]$. Let $m' = m - k - \ell$, we claim that

$$(s_0 + d \cdot [m']) \cap \{a_i\}_{i=1}^{\ell} = \varnothing. \tag{3.24}$$

Otherwise, suppose there is $r \in [m']$ such that $s_0 + rd = a_t$ for some $t \in [\ell]$. Then

$$x' + \sum_{j=1}^{k-1} b_j = x' + rd + \sum_{j=1, j \neq t}^{\ell} a_j.$$

By taking $x' \in x + d \cdot [0, m-r]$, then both x' and $x' + rd$ are in A. Since $m - r \geq k + \ell$, there is $\alpha \in [0, m-r]$ such that $x + \alpha d \notin \{b_j\}_{j=1}^{k-1}$, and $x + (\alpha + r)d \notin \{a_i\}_{i=1}^{\ell}$. This contradicts that A is restricted (k,ℓ)-sum-free.

Let $B = \{b \in A' \mid (b + d \cdot [m']) \cap A \neq \varnothing\}$, and let

$$B' = B \setminus \left(\Big(\bigcup_{i=1}^{\ell} \{a_i\} \cup \mathcal{E} \Big) - d \cdot [m'] \right).$$

Let $c_0 = 0$, $c_i = a_i$ when $i \in [\ell]$, and $c_j = a_{j-\ell}$ when $j \in [\ell+1, k-1]$. For $u \in [k-1]$ and $v \in [2, \ell]$, let

$$\mathcal{C}(u) = B' + \sum_{j=1}^{k-u} b_j + \sum_{i=0}^{u-1} c_i, \quad \mathcal{D}(v) = B' + \sum_{i=0}^{\ell-v} a_i + \sum_{j=0}^{v-1} b_j,$$

and $\mathcal{C}(k) = A' + \sum_{i=0}^{k-1} c_i$, $\mathcal{D}(1) = A' + \sum_{i=1}^{\ell-1} a_i$. Let $\mathcal{F}$ consists of all $\mathcal{C}(u)$ and $\mathcal{D}(v)$, then Claim 2 still holds. In fact, suppose there are $u_1, u_2 \in [k]$, $u_1 < u_2$ such that $\mathcal{C}(u_1) \cap \mathcal{C}(u_2) \neq \varnothing$ (the case when $\mathcal{D}(v_1) \cap \mathcal{D}(v_2) \neq \varnothing$ is simpler). Then there exist $y_1 \in B'$, $y_2 \in A'$ such that

$$y_1 + \sum_{j=k-u_2+1}^{k-u_1} b_j = y_2 + \sum_{i=u_1}^{u_2-1} c_i.$$

Let $e_0 = 0$, and $e_1, \ldots, e_{u_2-u_1-1} \in \mathcal{E}$ if $u_2 > u_1 + 1$. If $u_2 \leq \ell$, we have

$$s_0 = y_1 + \sum_{i \in [0, u_1-1] \cup [u_2, \ell]} a_i + \sum_{t=0}^{u_2-u_1-1} e_t - y_2 - \sum_{j \in [0, k-u_2] \cup [k-u_1+1, k-1]} b_j - \sum_{t=0}^{u_2-u_1-1} e_t.$$

If $u_1 \geq \ell + 1$, we get

$$s_0 = y_1 + \sum_{i \in [0, u_1-1-\ell] \cup [u_2-\ell, \ell]} a_i + \sum_{t=0}^{u_2-u_1-1} e_t - y_2 - \sum_{j \in [0, k-u_2] \cup [k-u_1+1, k-1]} b_j - \sum_{t=0}^{u_2-u_1-1} e_t.$$

If $u_1 \leq \ell$, $u_2 \geq \ell + 1$, and $u_2 - u_1 + 1 \leq \ell$,

$$s_0 = y_1 + \sum_{i \in [u_2-\ell, u_1-1]} a_i + \sum_{t=0}^{u_2-u_1-1} e_t - y_2 - \sum_{j \in [0, k-u_2] \cup [k-u_1+1, k-1]} b_j - \sum_{t=0}^{u_2-u_1-1} e_t.$$

If $u_1 \leq \ell$, $u_2 \geq \ell+1$, and $u_2 - u_1 \geq \ell$. Let $e_0 = 0$, $e_1, \dots, e_{\ell-1} \in \mathcal{E}$ if $\ell > 1$. Thus

$$s_0 = y_1 + \sum_{t=0}^{\ell-1} e_t - y_2 - \sum_{j \in [0,k-u_2] \cup [k-u_1+1,k-1]} b_j - \sum_{t=0}^{\ell-1} e_t - \sum_{i=u_1}^{u_2-1-\ell} a_i.$$

Note that $k \leq 2\ell+1$ implies $u_2 - 1 - \ell \leq \ell$.

In any case, since $y_1 \in B$, by (3.22), (3.23), and (3.24), there is $r \in [m']$ such that $s_0 + rd \in W$, which contradicts the assumption of s_0. Therefore,

$$(k+\ell-2)\overline{d}(B) + 2\overline{d}(A) \leq 1, \tag{3.25}$$

since $\overline{d}(A') = \overline{d}(A)$ and $\overline{d}(B') = \overline{d}(B)$.

We also modify the construction of $\mathcal{N}(t)$ in a similar way. For every $t \in [m']$, let $\mathcal{N}(t) = A' \setminus B + td$. Let $e_0 = 0$, and $e_1, \dots, e_{k-2} \in \mathcal{E}$ if $k \geq 3$. Let $A'' = A' \setminus (x + d \cdot [m'])$. Define

$$\mathcal{G} = \left\{ \mathcal{N}(t), A', A'' + \sum_{i=1}^{\ell-1} a_i - x - \sum_{j=0}^{k-2} e_j \;\middle|\; t \in [m'] \right\}$$

Then by using the similar argument, it is easy to see that Claim 3 still holds. We omit the details here. We have

$$(m-k-\ell+2)\overline{d}(A) - (m-k-\ell)\overline{d}(B) \leq 1,$$

since $\overline{d}(A'') = \overline{d}(A)$. Together with (3.25), finishes the proof. □

Next, we consider the analogue of Lemma 3.16 for restricted (k,ℓ)-sum-free sets. The structure here is slightly different from the (k,ℓ)-sum-free sets.

Lemma 3.19. *Let $\varepsilon > 0$ and let k, ℓ be positive integers with $\ell < k \leq 2\ell+1$. Then there is $D > 0$ such that the following holds. Let $A \subseteq \mathbb{N}^{>0}$ be a restricted (k,ℓ)-sum-free set, and $\overline{d}(A) > \frac{1}{k+\ell} + \varepsilon$. Then after removing at most $D(2k+\ell)$ elements from A, it is contained in a periodic restricted (k,ℓ)-sum-free set with period D.*

Proof. We pick $m > k + \ell$ such that

$$\frac{m-2}{(k+\ell)(m-k-\ell)+2(k+\ell-2)} < \frac{1}{k+\ell} + \varepsilon. \tag{3.26}$$

Let $L = L(\varepsilon, m)$ be as in Lemma 3.15, and let $D = L!$. We consider the partition of $\mathbb{N}$ into cosets:

$$\mathbb{N} = \bigcup_{x \in [D]} x + D \cdot \mathbb{N}.$$

For every $x \in [D]$, let $\mathbb{N}_x = x + D \cdot \mathbb{N}$, and $A_x = A \cap \mathbb{N}_x$. Let A' be a subset of A, obtained by removing A_x from A when $|A_x| < 2k + \ell$. Hence $\overline{d}(A') = \overline{d}(A)$. Next, we are going to show that A' is contained in a periodic restricted (k, ℓ)-sum-free set with period D. Suppose this is not the case. Let

$$C = \Big(\bigcup_{a \in A'} a + D \cdot \mathbb{Z}\Big) \cap \mathbb{N}^{>0}.$$

Thus C is not restricted (k, ℓ)-sum-free. This means, there are ℓ distinct elements $a_1, \ldots, a_\ell \in C$ and k distinct elements $b_1, \ldots, b_k \in C$, such that $\sum_{i=1}^{\ell} a_i = \sum_{j=1}^{k} b_j$. Let P be the "(k, ℓ)-sum-free" part of C, that for every $w \in P$, every $k-1$ distinct elements $y_1, \ldots, y_{k-1} \in C \backslash \{w\}$, and every ℓ distinct elements $z_1, \ldots, z_\ell \in C$, we have $w + \sum_{i=1}^{k-1} y_i \neq \sum_{j=1}^{\ell} z_j$. Let $e_0 = 0$, and let $\mathcal{E}$ consists of $k-1$ distinct elements $e_1, \ldots, e_{k-1} \in C$, such that $\mathcal{E}$ is disjoint from $\{a_i\}_{i=1}^{\ell}$ and $\{b_j\}_{j=1}^{k}$.

$$P' = P \setminus \Big(\bigcup_{i=1}^{\ell}\{a_i\} \cup \bigcup_{j=1}^{k}\{b_j\} \cup \mathcal{E}\Big).$$

Set $a_0 = b_0 = c_0 = 0$. Let $c_t = a_t$ when $t \in [\ell]$, and $c_t = a_{t-\ell}$ when $t \in [\ell+1, k-1]$. For every $u \in [k]$ and $v \in [\ell]$, let

$$\mathcal{M}(u) = P' + \sum_{j=0}^{k-u} b_j + \sum_{t=0}^{u-1} c_t, \qquad \mathcal{N}(v) = P' + \sum_{i=0}^{\ell-v} a_i + \sum_{t=0}^{v-1} b_t.$$

Let $\mathcal{F}$ be the collection of all $\mathcal{M}(u)$ and $\mathcal{N}(v)$. Then elements in $\mathcal{F}$ are pairwise disjoint.

Otherwise, suppose there are $u_1, u_2 \in [k]$, $u_1 < u_2$ such that $\mathcal{M}(u_1) \cap \mathcal{M}(u_2) \neq \varnothing$ (the case when $\mathcal{N}(v_1) \cap \mathcal{N}(v_2) \neq \varnothing$ is simpler). Thus, there are $y_1, y_2 \in P'$, such that

$$y_1 + \sum_{k-u_2+1}^{k-u_1} b_j = y_2 + \sum_{t=u_1}^{u_2-1} c_t.$$

Let $e_1, \ldots, e_{u_2-u_1-1} \in \mathcal{E}$ if $u_2 > u_1 + 1$. If $u_2 \leq \ell$, we have

$$\begin{aligned} 0 &= \sum_{i=1}^{\ell} a_i - \sum_{j=1}^{k} b_j \\ &= y_1 + \sum_{i \in [0, u_1-1] \cup [u_2, \ell]} a_i + \sum_{t=0}^{u_2-u_1-1} e_t - y_2 - \sum_{j \in [0, k-u_2] \cup [k-u_1+1, k]} b_j - \sum_{t=0}^{u_2-u_1-1} e_t. \end{aligned}$$

If $u_1 \geq \ell + 1$, we have

$$0 = y_1 + \sum_{i \in [0, u_1-1-\ell] \cup [u_2-\ell, \ell]} a_i + \sum_{t=0}^{u_2-u_1-1} e_t - y_2 - \sum_{j \in [0, k-u_2] \cup [k-u_1+1, k]} b_j - \sum_{t=0}^{u_2-u_1-1} e_t.$$

If $u_1 \leq \ell$, $u_2 \geq \ell + 1$, and $\ell \geq u_2 - u_1$, we get

$$0 = y_1 + \sum_{i=u_2-\ell}^{u_1-1} a_i + \sum_{t=0}^{u_2-u_1-1} e_t - y_2 - \sum_{j \in [0, k-u_2] \cup [k-u_1+1, k]} b_j - \sum_{t=0}^{u_2-u_1-1} e_t.$$

If $u_1 \leq \ell$, $u_2 \geq \ell + 1$, and $\ell < u_2 - u_1$. Let $e_1, \ldots, e_{\ell-1} \in \mathcal{E}$ if $\ell > 1$, we get

$$0 = y_1 + \sum_{t=0}^{\ell-1} e_t - y_2 - \sum_{j \in [0, k-u_2] \cup [k-u_1+1, k]} b_j - \sum_{i=u_1}^{u_2-1-\ell} a_i - \sum_{t=0}^{\ell-1} e_t.$$

In any case, we get a contradiction with the assumption of P' and the fact that $y_2 \in P'$. Therefore,

$$\overline{d}(P) \leq \frac{1}{k+\ell},$$

since $\overline{d}(P') = \overline{d}(P)$. This means, $\overline{d}(A' \backslash P) \geq \varepsilon$. By Lemma 3.15, $A' \backslash P$ contains a progression $x + d \cdot [m]$, and $d < L$. By the way we construct P, there are ℓ distinct elements $s_1, \ldots, s_\ell \in C$

and $k-1$ distinct elements $t_1, \dots, t_{k-1}$ in $C \setminus \{x+m\}$ such that

$$x + m = \sum_{i=1}^{\ell} s_i - \sum_{j=1}^{k-1} t_j.$$

By the way we construct A', for every $r \in [D]$, if $|A' \cap \mathbb{N}_r| > 0$, then $|A' \cap \mathbb{N}_r| \geq 2k + \ell$. Thus, there are ℓ distinct elements $\alpha_1, \dots, \alpha_\ell \in A'$ and $k-1$ distinct elements $\beta_1, \dots, \beta_{k-1} \in A'$, such that for every $i \in [\ell]$ and $j \in [k-1]$, we have that $\alpha_i \in s_i + D \cdot \mathbb{Z}$, and $\beta_j \in t_j + D \cdot \mathbb{Z}$. Let $s = \sum_{i=1}^{\ell} \alpha_i - \sum_{j=1}^{k-1} \beta_j$. Note that $|A' \cap \mathbb{N}_r| \geq 2k + \ell$ also implies that there is $r' \in [\ell]$, and $M \subseteq \mathbb{N}^{>0}$, $|M| \geq k$, such that

$$\alpha_{r'} + D \cdot M \subseteq A', \quad (\alpha_{r'} + D \cdot M) \cap \bigcup_{i=1}^{\ell} \{\alpha_i\} = \varnothing.$$

Thus if $s \cap \{\beta_j\}_{j=1}^{k-1} \neq \varnothing$, then by changing $\alpha_{r'}$ by $\alpha_{r'} + nD$ for some $n \in M$, one can make $s + nD \cap \{\beta_j\}_{j=1}^{k-1} = \varnothing$. Since $d \mid D$, we have $s \in x + d \cdot \mathbb{Z}$. By Lemma 3.14, we have that

$$\overline{d}(A) \leq \frac{m-2}{(k+\ell)(m-k-\ell) + 2(k+\ell-2)},$$

and this contradicts (3.26). $\square$

Let A be a restricted (k, ℓ)-sum-free set, and let A' be a subset of A obtained by removing finitely many elements from A. Observe that, if A' is contained in a periodic restricted (k, ℓ)-sum-free set, then A cannot contain a multiple of every natural number. Thus, using the same proof in Theorem 3.13, we conclude that $\widetilde{d}(A) \leq \frac{1}{k+\ell}$ if A is restricted (k, ℓ)-sum-free.

3.9 Concluding Remarks

In this chapter, we first study $\mathcal{M}_{(k,\ell)}(N)$. In particular, we prove that Conjecture 3.2 is true for infinitely many (k, ℓ). While solving Conjecture 3.2 might not be a realistic target at the moment, the following conjecture for the case when $k-\ell \geq 2$ might be feasible. This is because in this case, Lemma 3.10 implies that we have two different asymmetric maximal (k,ℓ)-sum-free open sets in $\mathbb{T}$, and the technique developed in this chapter might be useful.

Conjecture 3.20. *Let k,ℓ be positive integers and $k \geq \ell+2$. Then there is a function $\omega(N) \to \infty$ as $N \to \infty$, such that*

$$\mathcal{M}_{(k,\ell)}(N) \geq \frac{N}{k+\ell} + \omega(N).$$

We also study $\widehat{\mathcal{M}}_{(k,\ell)}(N)$ in Theorem 3.4. As we can see in the proofs in Section 3.8, when $k > 2\ell+1$, the current strategy failed to obtain disjoint sets $\mathcal{C}$ and $\mathcal{D}$ in the proof of Lemma 3.18, as well as disjoint sets $\mathcal{M}$ and $\mathcal{N}$ in the proof of Lemma 3.19. Although we think it is very likely that the conclusion in Theorem 3.4 holds for every k and ℓ, the case $k > 2\ell+1$ may require some new ingredients.

Conjecture 3.21. *For every positive integers k,ℓ with $k > 2\ell+1$,*

$$\widehat{\mathcal{M}}_{(k,\ell)}(N) = \Big(\frac{1}{k+\ell} + o(1)\Big)N.$$

A (k,ℓ)-sum-free set is a set forbidding a linear equation $\sum_{i=1}^{\ell} x_i = \sum_{j=1}^{k} y_j$. Another interesting direction is to consider the analogue problem on sets forbidding a system of linear equations. One of the most interesting problems along this line might be forbidding the projective cubes. Given a multiset $S=\{s_1,\ldots,s_d\}$, a *d-dimensional projective cube* generated by S is

$$\square^d(S) := \Big\{ \sum_{i\in I} s_i \;\Big|\; \varnothing \neq I \subseteq [d] \Big\}.$$

A set is *$\square^d$-free* if it does not contain any d-dimensional projective cubes as its subsets. Extremal properties of projective cubes have a vast literature, see e.g. [2, 61, 86, 126]. The problem on forbidding d-dimensional projective cubes can be viewed as a generalization of sum-free sets in another direction, since a sum-free set is also a $\square^2$-free set. Thus, the following problem is worthwhile to pursue.

Question 3.22. *Let $d \geq 3$ be an integer. Define*

$$\mathcal{M}_{\square^d}(N) := \inf_{\substack{A \subseteq \mathbb{N}^{>0} \\ |A|=N}} \max_{\substack{B \subseteq A \\ B \text{ is } \square^d\text{-free}}} |B|.$$

Determine $\mathcal{M}_{\square^d}(N)$.

Chapter 4

A closer look to the largest sum-free sets

Given A a set of N positive integers. In the last chapter, we have discussed an old conjecture in additive combinatorics:

Conjecture 4.1 (sum-free conjecture, combinatorial form). *There is a function $\omega(N) \to \infty$ as $N \to \infty$, such that*

$$\mathcal{M}_{(2,1)}(N) > \frac{N}{3} + \omega(N).$$

The conjecture is generally attacked in the literature by considering another stronger conjecture:

Conjecture 4.2 (sum-free conjecture, analytic form). *Let $\Omega = (1/3, 2/3) \subseteq \mathbb{T}$. Then when $N \to \infty$,*

$$\max_{x \in \mathbb{T}} \sum_{n \in A} (\mathbb{1}_\Omega - 1/3)(nx) \to \infty.$$

This analytic conjecture, if true, would also imply that a similar phenomenon occurs for $(2k, 4k)$-sum-free sets for every $k \geq 1$, though we do not know if these two conjectures are equivalent. In this chapter, we prove the latter result directly. The new ingredient of our proof is a structural analysis on the host set A, which might be of independent interest. This chapter is based on joint work with Wu [107].

4.1 Introduction

In [59], using a probabilistic argument, Erdős showed that $\mathcal{M}_{(2,1)}(N) \geq N/3$. This argument is actually not complicated: Let Ω be a maximal sum-free subset of $\mathbb{R}/\mathbb{Z}$, then $(\Omega+\Omega) \cap \Omega =$

$\varnothing$. By Kneser's inequality [115], we have an upper bound on $|\Omega|$ that $|\Omega| \le 1/3$. Now we fix a maximal sum-free set Ω of measure $1/3$ (e.g. $(1/3, 2/3)$). For any $x \in \mathbb{R}/\mathbb{Z}$, we let A_x be the set of integers n in A such that $nx \in \Omega$. Then clearly A_x is sum-free, and hence we have

$$\mathcal{M}_{(2,1)}(A) \ge |A_x| = \sum_{n \in A} \mathbb{1}_{\Omega}(nx),$$

where $\mathbb{1}_{\Omega}$ is the characteristic function of Ω. When x is chosen randomly from $\mathbb{R}/\mathbb{Z}$, the expected size of A_x is $N/3$, which implies that $|A_x| \ge N/3$ for some x. This is actually the motivation to formulate Conjecture 4.2, which could imply Conjecture 4.1. The lower bound estimate of $\max_{x \in \mathbb{R}/\mathbb{Z}} \sum_{n \in A} \mathbb{1}_{\Omega}(nx)$ is later improved to $(N+1)/3$ by Alon and Kleitman [3], and the best estimate up to date is obtained by Bourgain [28], where he showed that $\max_{x \in \mathbb{R}/\mathbb{Z}} \sum_{n \in A} \mathbb{1}_{\Omega}(nx) \ge (N+2)/3$.

For (k, ℓ)-sum-free sets, in general, we believe the following should be true, which is a generalization of Conjecture 4.2 to all (k, ℓ)-sum-free sets.

Conjecture 4.3. *There is a function $\omega(N) \to \infty$ as $N \to \infty$, such that for every set A of N positive integers, there exists a maximal (k, ℓ)-sum-free set $\Omega(k, \ell) \subseteq \mathbb{R}/\mathbb{Z}$, and we have*

$$\max_{x \in \mathbb{R}/\mathbb{Z}} \sum_{n \in A} \Big(\mathbb{1}_{\Omega(k,\ell)} - \frac{1}{k+\ell} \Big)(nx) = \omega(N).$$

Note that if Conjecture 4.2 holds, then this would imply that Conjecture 4.3 holds for $(2k, 4k)$-sum-free sets for all $k \ge 1$. Hence we believe that the $(2k, 4k)$-sum-free problem is one of the most interesting cases of Conjecture 4.3. In this chapter, we prove the $(2k, 4k)$-sum-free case without assuming Conjecture 4.2.

Theorem 4.4. *For every $k \ge 1$, there is a function $\omega(N) = \log N / \log\log N$, such that for every set A of N positive integers, there exists a maximal $(2k, 4k)$-sum-free set $\Omega(2k, 4k) \subseteq \mathbb{R}/\mathbb{Z}$, and we have*

$$\max_{x \in \mathbb{R}/\mathbb{Z}} \sum_{n \in A} \Big(\mathbb{1}_{\Omega(2k,4k)} - \frac{1}{6k} \Big)(nx) \gg \omega(N).$$

In particular, there is an absolute constant $c > 0$*, such that*

$$\mathcal{M}_{(2k,4k)}(N) \geq \frac{N}{6k} + c\,\omega(N).$$

We remark that Theorem 4.4 is not enough for the sum-free conjecture. Indeed, let $\Omega_1 = (1/6, 1/3)$ and $\Omega_2 = (2/3, 5/6)$ be two subsets of $\mathbb{R}/\mathbb{Z}$. Note that both Ω_1 and Ω_2 are $(2,4)$-sum-free in $\mathbb{R}/\mathbb{Z}$. By Theorem 4.4, one can find two elements $x_1, x_2 \in \mathbb{R}/\mathbb{Z}$ such that $A_{x_1} := \{n \in A : nx_1 \in \Omega_1\}$ has at least $N/6+\omega(N)$ elements, and $A_{x_2} := \{n \in A : nx_2 \in \Omega_2\}$ also has at least $N/6 + \omega(N)$ elements. Clearly, both A_{x_1} and A_{x_2} are $(2,4)$-sum-free. If $x_1 = x_2$, the above argument implies the sum-free conjecture since the union of Ω_1 and Ω_2 is sum-free in $\mathbb{R}/\mathbb{Z}$. However, it is possible that $x_1 \neq x_2$ and thus A_{x_1} and A_{x_2} may share many common elements.

The new ingredients used in proving Theorem 4.4 contain a structural analysis of the given set A. Recall that a Følner sequence in $(\mathbb{N}, \cdot)$ is a collection of sets of integers $\{F_n\}_{n=1}^{\infty}$, such that for every $a \in \mathbb{N}^{>0}$,

$$\lim_{n\to\infty} \frac{|F_n \triangle (a \cdot F_n)|}{|F_n|} = 0.$$

Thus, when A is close to a set in a Følner sequence, we expect that $|A \triangle (a \cdot A)|$ is small for appropriate a. Inspired by the structure Følner sequences (which is the only known constructive example whose largest (k,ℓ)-sum-free subsets have cardinality $N/(k+\ell)+o(N)$, for all (k,ℓ), see [56, 108]), we split our proof into two cases: when $|A \triangle (a \cdot A)|$ is small (close to having multiplicative structures), and when $|A \triangle (a \cdot A)|$ is large (close to having additive structures). We mainly consider the case $a = 3$ here since the Fourier coefficients appeared in the later proofs contain a multiplicative character mod 3. The first case is resolved by an application of Littlewood–Paley theorem, and the proof we given also works for sum-free sets. In the second case, since the main factors in the Fourier coefficients are not multiplicative, we carefully sieve out small prime factors, and apply a variant of the weak Littlewood conjecture. The nontrivial lower bound of $\omega(N)$ eventually comes from the largeness of $|A \triangle (3 \cdot A)|$. For convenience, we make the following definition.

Definition 4.5. We say a set $A \subseteq \mathbb{Z}^{>0}$ is a *geometric set*, if $|A \triangle (3 \cdot A)| \ll |A|^c$ for an absolute constant $c < 1$.

The chapter is organized as follows. In the next section, we deal with the case when $|A \triangle (3{\cdot}A)|$ is small (we actually prove a more general result there). In Section 4.3, we prove the case when $|A \triangle (3{\cdot}A)|$ is large, and finish the proof of Theorem 4.4.

Notation

Given a set A and a positive integer k, we use kA to denote the set $\{a_1 + \cdots + a_k : a_i \in A \text{ for } 1 \le i \le k\}$, and use $k \cdot A$ to denote the set $\{ka : a \in A\}$. For every $\theta \in \mathbb{R}/\mathbb{Z}$, we write $e(\theta) = e^{2\pi i\theta}$. We use the standard Vinogradov notation. That is, $f \ll g$ means $f = O(g)$, and $f \asymp g$ if $f \ll g$ and $f \gg g$.

4.2 When A is geometric

In this section, we study the size of the largest sum-free sets when the host set A is well-structured. Let $\Omega = (1/3, 2/3) \subseteq \mathbb{R}/\mathbb{Z}$, and it is easy to check that Ω is sum-free in $\mathbb{R}/\mathbb{Z}$. Define $\mathbb{1}_\Omega$ as the characteristic function of Ω, and let $f = \mathbb{1}_\Omega - 1/3$ be the balanced function of $\mathbb{1}_\Omega$. By orthogonality of characters we have

$$\widehat{f}(n) = \begin{cases} 0 & \text{if } n = 0, \\ \widehat{\mathbb{1}_\Omega}(n) & \text{otherwise.} \end{cases}$$

When $n > 0$,

$$\widehat{f}(n) = \int_{\mathbb{R}/\mathbb{Z}} \mathbb{1}_\Omega(x) e(-nx) \,\mathrm{d}\, x$$

$$= \frac{1}{2\pi i n}\Big(- e\Big(- \frac{2n}{3}\Big) + e\Big(- \frac{n}{3}\Big)\Big)$$
$$= \frac{1}{\pi n} e\Big(\frac{n}{2}\Big) \sin\Big(\frac{n\pi}{3}\Big).$$

Therefore we obtain

$$f(x) = \sum_{n\neq 0} \widehat{f}(n) e(nx) = \sum_{n\neq 0} \frac{1}{\pi n} e\Big(\frac{n}{2}\Big) \sin\Big(\frac{n\pi}{3}\Big) e(nx)$$
$$= -\frac{\sqrt{3}}{\pi} \sum_{n\geq 1} \frac{\chi(n)}{n} \cos(2\pi n x), \tag{4.1}$$

where $\chi(n)$ is a multiplicative character mod 3, that is

$$\chi(n) = \begin{cases} 1 & \text{when } n \equiv 1 \pmod 3, \\ -1 & \text{when } n \equiv 2 \pmod 3, \\ 0 & \text{otherwise.} \end{cases}$$

As we mentioned in the introduction,

$$\mathcal{M}_{(2,1)}(A) - \frac{N}{3} \geq \max_{x\in\mathbb{R}/\mathbb{Z}} \sum_{m\in A} f(mx). \tag{4.2}$$

In the rest of the section, we are going to estimate $\max_x \sum_{m\in A} f(mx)$ when A has certain algebraic constrains.

Let $\mathcal{P}_3$ be the collection of intervals $[3^k, 3^{k+1}) \cap \mathbb{N}$, where $k \geq 0$ is an integer. Let A be a set of N positive integers. We say that A is $(3, c)$**-lacunary**, if there is a subset $\mathcal{P} \subset \mathcal{P}_3$ with $|\mathcal{P}| \gg N^c$, such that each interval in $\mathcal{P}$ contains at least one element of A, and the intervals in $\mathcal{P}$ form a cover of A.

When the set A is $(3, c)$-lacunary, in some sense the distribution of A is not far away from a union of long geometric progressions, and we expect that approximately there is a square root cancellation for $\| \sum_{m\in A} e(mx)\|_{L^1(\mathbb{R}/\mathbb{Z})}$. To make this observation rigorous, we use the Littlewood–Paley theorem.

Theorem 4.6 (Littlewood–Paley)**.** *Let $g(x)$ be the trigonometric series*

$$g(x) = \sum_{n=1}^{\infty} a_n e^{inx}.$$

For the sequence $\{a_n\}$, we consider the following auxiliary truncated function Δ_k defined as

$$\Delta_k(x) = \sum_{n=n_{k-1}+1}^{n_k} a_n e^{inx},$$

where $n_0 = 0$, $n_1 = 1$, $n_{k+1}/n_k \geq \alpha > 1$. Then for any $1 < p < \infty$.

$$\left\| \Big(\sum_{k=1}^{\infty} |\Delta_k|^2 \Big)^{1/2} \right\|_{L^p(\mathbb{R}/\mathbb{Z})} \leq C_{p,\alpha} \|f\|_{L^p(\mathbb{R}/\mathbb{Z})}.$$

The proof of the Littlewood–Paley theorem can be found in [171, Chapter XV, Theorem 4.7], and the constant $C_{p,\alpha}$ was calculated in [27] when $\alpha = 2$.

The next lemma gives us a key estimate for lacunary sets.

Lemma 4.7. *Assume that $\{a_n\}_{n=1}^N$ is $(3, c)$-lacunary. Define $g(x)$ as*

$$g(x) = \sum_{n=1}^{N} e^{ia_n x}.$$

Then $\|g\|_{L^1(\mathbb{R}/\mathbb{Z})} \gg N^{c/3}$.

Proof. Let ϕ_k be the indicator function of the interval $[3^k, 3^{k+1})$. We denote by $\Delta_k(g)$ the Fourier truncation

$$\Delta_k(g)(x) = \sum_{n=1}^{N} \phi_k(a_n) e^{ia_n x}.$$

By the Littlewood–Paley theorem (Theorem 4.6), we have for any $1 < p < \infty$,

$$\|g\|_{L^p(\mathbb{R}/\mathbb{Z})} \geq C_p \Big\| \Big(\sum_{k=0}^{\infty} |\Delta_k g(x)|^2 \Big)^{1/2} \Big\|_{L^p(\mathbb{R}/\mathbb{Z})} = C_p \Big\| \Big(\sum_{k\in E} |\Delta_k g(x)|^2 \Big)^{1/2} \Big\|_{L^p(\mathbb{R}/\mathbb{Z})}. \tag{4.3}$$

Here the set E contains all the positive integers k satisfying $[3^k, 3^{k+1}) \in \mathcal{P}$, so

$$|E| = |\mathcal{P}| \gg N^c.$$

We bound the right hand side of equation (4.3) using Hölder's inequality so that

$$\Big\| \Big(\sum_{k\in E} |\Delta_k g(x)|^2 \Big)^{1/2} \Big\|_{L^p(\mathbb{R}/\mathbb{Z})} \geq |E|^{-1/2} \Big\| \Big(\sum_{k\in E} |\Delta_k g(x)| \Big) \Big\|_{L^p(\mathbb{R}/\mathbb{Z})},$$

which clearly implies

$$\Big\| \Big(\sum_{k\in E} |\Delta_k g(x)|^2 \Big)^{1/2} \Big\|_{L^p(\mathbb{R}/\mathbb{Z})} \geq |E|^{-1/2} \sum_{k\in E} \Big\| \Delta_k g(x) \Big\|_{L^1(\mathbb{R}/\mathbb{Z})}.$$

Using Hölder's inequality again, we get $\|\Delta_k g\|_2^2 \leq \|\Delta_k g\|_1 \|\Delta_k g\|_\infty$, and this implies

$$\|\Delta_k g\|_{L^1} \geq 1$$

uniformly in k. Therefore,

$$\Big\| \Big(\sum_{k\in E} |\Delta_k g(x)|^2 \Big)^{1/2} \Big\|_{L^p(\mathbb{R}/\mathbb{Z})} \geq |E|^{1/2} \gg N^{c/2}. \tag{4.4}$$

Since $\|\Delta_k g\|_p^p \leq \|\Delta_k g\|_1 \|\Delta_k g\|_\infty^{p-1}$, we can bound $\|g\|_{L^p}^p$ easily by

$$\|g\|_{L^1(\mathbb{R}/\mathbb{Z})} \geq N^{1-p} \|g\|_{L^p(\mathbb{R}/\mathbb{Z})}^p \tag{4.5}$$

Finally, we combine estimates (4.3), (4.4) and (4.5) to finish the proof of this lemma, by choosing $p = 1 + c/6$. $\square$

Now we are going to estimate the right hand side of (4.2) with assumption that A is $(3, c)$-lacunary for some constant $c > 0$. Let

$$F(x) = \sum_{m \in A} f(mx).$$

Since f is a balanced function, we have $\int_{\mathbb{R}/\mathbb{Z}} F = 0$, and this implies that

$$\max_{x \in \mathbb{R}/\mathbb{Z}} F(x) \geq \frac{1}{2} \|F\|_{L^1(\mathbb{R}/\mathbb{Z})}. \tag{4.6}$$

Let $P \asymp N^2$ be a prime, and let $\mathcal{M}$ be the collection of square-free integers generated by primes smaller than P. Let μ be the Möbius function, so by equation (4.1),

$$\begin{aligned}
\sum_{k \in \mathcal{M}} \frac{\mu(k)\chi(k)}{k} \sum_{m \in A} f(mkx) &= -\frac{\sqrt{3}}{\pi} \sum_{m \in A,\, n \geq 1} \frac{\chi(n)}{n} \cos(2\pi mnx) \sum_{k \in \mathcal{M},\, k | n} \mu(k) \\
&= -\frac{\sqrt{3}}{\pi} \sum_{\substack{m \in A \\ n \geq 1, n \in \mathcal{N}}} \frac{\chi(n)}{n} \cos(2\pi mnx),
\end{aligned}$$

where $\mathcal{N}$ is the set of integers n such that for every $p < P$, $\gcd(n, p) = 1$.

Therefore, by Minkowski's inequality we have

$$\begin{aligned}
\left\| \sum_{k \in \mathcal{M}} \frac{\mu(k)\chi(k)}{k} \sum_{m \in A} f(mkx) \right\|_{L^1(\mathbb{R}/\mathbb{Z})} &\gg \left\| \sum_{m \in A} \cos(2\pi mx) \right\|_{L^1(\mathbb{R}/\mathbb{Z})} \\
&\quad - \left\| \sum_{\substack{m \in A \\ n > 1, n \in \mathcal{N}}} \frac{\chi(n)}{n} \cos(2\pi mnx) \right\|_{L^1(\mathbb{R}/\mathbb{Z})}.
\end{aligned}$$

Via Cauchy–Schwarz inequality and Plancherel, the second term is bounded by

$$\Bigg\| \sum_{\substack{m\in A\\ n>1, n\in\mathcal{N}}} \frac{\chi(n)}{n}\cos(2\pi mnx)\Bigg\|_{L^1(\mathbb{R}/\mathbb{Z})} \le C|A|P^{-1/2}.$$

Note that $P \asymp N^2$. By Merten's estimate, we have

$$\begin{aligned}\Bigg\|\sum_{m\in A}\cos(2\pi mx)\Bigg\|_{L^1(\mathbb{R}/\mathbb{Z})} &\ll \sum_{k\in\mathcal{M}} \frac{|\mu(k)|}{k}\big\|F(x)\big\|_{L^1(\mathbb{R}/\mathbb{Z})} + O(1).\\ &\ll \prod_{p<P}\Big(1+\frac{1}{p}\Big)\big\|F(x)\big\|_{L^1(\mathbb{R}/\mathbb{Z})} \asymp \log N\big\|F(x)\big\|_{L^1(\mathbb{R}/\mathbb{Z})}.\end{aligned}$$

Since A is $(3,c)$-lacunary, we invoke Lemma 4.7 to get

$$\Bigg\|\sum_{m\in A}\cos(2\pi mx)\Bigg\|_{L^1(\mathbb{R}/\mathbb{Z})} \gg N^c,$$

which implies

$$\big\|F(x)\big\|_{L^1(\mathbb{R}/\mathbb{Z})} \gg \frac{N^{c/3}}{\log N}.$$

Finally, we use estimate (4.6) to conclude $\max_{x\in\mathbb{R}/\mathbb{Z}} F(x) \gg N^{c/4}$. □

4.3 When A is not geometric

In this section, we consider the case when the host set A is uniformly distributed, in the sense that $|A \triangle 3\cdot A| \gg N^c$ for some positive constant $c > 0$. We will focus on finding the largest $(2,4)$-sum-free in A. Let $\Omega_1 = (1/6, 1/3) \subseteq \mathbb{R}/\mathbb{Z}$, and let $\Omega_2 = (2/3, 5/6) \subseteq \mathbb{R}/\mathbb{Z}$. It is clear that both Ω_1 and Ω_2 are $(2,4)$-sum-free in $\mathbb{R}/\mathbb{Z}$. Let $\mathbb{1}_{\Omega_t}$ be the indicator function of Ω_t for $t = 1, 2$. Given $A \subseteq \mathbb{N}^{>0}$ of size N, let $\mathcal{M}_{(2,4)}(A)$ be the size of the maximum

$(2,4)$-sum-free subset of A. Again we have

$$\mathcal{M}_{(2,4)}(A) \geq \max_{x \in \mathbb{R}/\mathbb{Z}} \sum_{n \in A} \mathbb{1}_{\Omega_t}(nx),$$

for every $t = 1, 2$. We introduce the balanced function $f_t : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ defined by $f_t = \mathbb{1}_{\Omega_t} - \frac{1}{6}$. Hence,

$$\widehat{f_t}(n) = \begin{cases} 0 & \text{if } n = 0, \\ \widehat{\mathbb{1}_{\Omega_t}}(n) & \text{otherwise.} \end{cases}$$

When $n > 0$, the Fourier coefficient $\widehat{f_t}(n)$ is

$$\begin{aligned} \widehat{f_t}(n) &= \int_{\mathbb{T}} \mathbb{1}_{\Omega_t}(x) e(-nx) d\mu(x) \\ &= \frac{1}{2\pi i n}\Big(- e\big(- \frac{(t-1)n}{2} - \frac{n}{3}\big) + e\big(- \frac{(t-1)n}{2} - \frac{n}{6}\big)\Big) \\ &= \frac{1}{\pi n} e\Big(\frac{(2t-1)n}{4}\Big) \sin\Big(\frac{n\pi}{6}\Big). \end{aligned}$$

Hence, for every $t = 1, 2$ we have

$$f_t(x) = \sum_{n \neq 0} \widehat{f_t}(n) e(nx) = \sum_{n \neq 0} \frac{1}{\pi n} e\Big(\frac{(2t-1)n}{4}\Big) \sin\Big(\frac{n\pi}{6}\Big) e(nx).$$

We will prove that either $\|f_1\|_1 \gg \log N / \log\log N$ or $\|f_2\|_1 \gg \log N / \log\log N$. However, it seems hard to estimate $\|f_t\|_1$ directly. In order to get around this difficult, we consider their sum $f_1 + f_2$ and difference $f_1 - f_2$. Let $\Gamma(x) := f_1(x) + f_2(x)$ be the sum so that

$$\Gamma(x) = \frac{1}{2\pi} \sum_{n \geq 1} \frac{(-1)^n}{n} \sin\Big(\frac{n\pi}{3}\Big) \cos(4\pi n x). \tag{4.7}$$

Also, we let $\Lambda(x) = f_1(x) - f_2(x)$ be the difference and let

$$\gamma(n) = \begin{cases} 1 & \text{when } n \equiv 1 \pmod 4, \\ -1 & \text{when } n \equiv 3 \pmod 4, \\ 0 & \text{otherwise,} \end{cases}$$

so that we can express $\Lambda(x)$ as

$$\begin{aligned} \Lambda(x) &= \frac{2}{\pi} \sum_{n\geq 1} \frac{\gamma(n)}{n} \sin\left(\frac{n\pi}{6}\right) \sin(2\pi n x) \\ &= \frac{2}{\pi} \Bigg(\sum_{\substack{n\geq 1 \\ 3\nmid n, 2\nmid n}} \frac{1}{2n} \sin(2\pi n x) - \sum_{\substack{n\geq 1 \\ 3\mid n, 2\nmid n}} \frac{1}{n} \sin(2\pi n x) \Bigg). \end{aligned} \tag{4.8}$$

We first deal with the function $\Gamma(x)$. Recall that $\mathcal{N}_1$ is the set of positive integers m such that m only contains prime factor larger than $Q \asymp (\log N)^{100}$. We also define $\mathcal{N}_2$ be the set of square-free integers generated by primes that at most Q. Since $(-1)^n \sin(n\pi/3) = -\sqrt{3}\chi(n)$ where $\chi(n)$ is a multiplicative character mod 3, we can sieve out the small prime factors in (4.7) by

$$\sum_{t\in\mathcal{N}_2} \frac{\mu(t)\chi(t)}{t} \sum_{m\in A} \Gamma(mtx) = -\frac{\sqrt{3}}{2\pi} \sum_{m\in A} \Bigg(\cos(4\pi m x) + \sum_{n\in\mathcal{N}_1} \frac{\chi(n)}{n} \cos(4\pi n m x) \Bigg), \tag{4.9}$$

where μ is the Möbius function.

Next, we consider $\Lambda(x)$. Since the coefficients $\gamma(n)\sin(n\pi/6)$ is not multiplicative, $\Lambda(x)$ is more difficult to handle. As shown in equation (4.8), $\Lambda(x)$ can be partition into two parts according to the divisibility of the number 3. This motivates us to first sieve out those integers n that $3 \mid n$, by a restricted Möbius function defined only on integers divisible by 3. In this way, except for the first term, all other terms with significant contribution in the second part cancel out, while the first part remains unchanged. Then, we use another sieve for the first part in a similar fashion. It turns out that we can combine these two steps to one by using the Möbius function directly as our sieve. In fact,

$$\sum_{m\in\mathcal{N}_2}\frac{\mu(m)}{m}\Lambda(mx)=\frac{2}{\pi}\sum_{n\geq 1}\frac{1}{2n}\sin(2\pi nx)\sum_{m\in\mathcal{N}_2,m|n}\mu(m)\gamma\Big(\frac{n}{m}\Big)\sin\Big(\frac{n\pi}{6m}\Big).$$

The equation above and be further partitioned as

$$\begin{aligned}
&\sum_{m\in\mathcal{N}_2,m|n}\mu(m)\gamma\Big(\frac{n}{m}\Big)\sin\Big(\frac{n\pi}{6m}\Big)\\
&=\sum_{\substack{m\in\mathcal{N}_2,m|n\\ 3\nmid n}}\frac{1}{2}\mu(m)+\sum_{9\nmid n,3|n}\Bigg(\sum_{\substack{m\in\mathcal{N}_2,m|n\\ 3\nmid m}}-\mu(m)+\sum_{\substack{m\in\mathcal{N}_2,m|n\\ 3|m}}\frac{1}{2}\mu(m)\Bigg)+\sum_{\substack{9|n,m|n\\ m\in\mathcal{N}_2}}-\mu(m)\\
&=:\ I_1(n)+I_2(n)+I_3(n).
\end{aligned}$$

By the inclusive-exclusive principle, for $n\notin\mathcal{N}_1\cup 3\cdot\mathcal{N}_1$, $I_1(n)$ is always 0 unless $n=1$, and $I_3(n)$ is always 0. In $I_2(n)$,

$$\sum_{\substack{m\in\mathcal{N}_2,m|n\\ 3|m}}\frac{1}{2}\mu(m)=\sum_{\substack{m\in\mathcal{N}_2,m|n\\ 3\nmid m}}-\frac{1}{2}\mu\Big(\frac{m}{3}\Big),$$

which implies that $I_2(n)$ is 0 unless $n=3$. Therefore, we get

$$\begin{aligned}
\Lambda_1(x)&:=\sum_{t\in\mathcal{N}_2}\sum_{m\in A}\frac{\mu(t)}{t}\Lambda(tmx) \qquad (4.10)\\
&=\frac{2}{\pi}\sum_{m\in A}\Bigg(\frac{1}{2}\sin(2\pi mx)-\frac{1}{2}\sin(6\pi mx)+\sum_{n\in\mathcal{N}_1\cup 3\cdot\mathcal{N}_1}\frac{\eta(n)}{n}\sin(2\pi nmx)\Bigg),
\end{aligned}$$

where η is defined as

$$\eta(n)=\begin{cases}\frac{1}{2} & \text{when } n\in\mathcal{N}_1,\\ -\frac{3}{2} & \text{when } n\in 3\cdot\mathcal{N}_1.\end{cases}$$

Note that Λ_1 indeed has the expression

$$\Lambda_1(x) = \frac{1}{\pi} \sum_{m \in A} \left(\sin(2\pi mx) - \sin(6\pi mx) + \sum_{n \in \mathcal{N}} \frac{1}{n} \Big(\sin(2\pi nmx) - \sin(6\pi nmx) \Big) \right).$$

Let $B = A \triangle 3 \cdot A$, so by our assumption on the host set A, $|B| \gg N^c$. For any number $m \in A \cup (3 \cdot A)$, we define

$$\varepsilon(m) = \begin{cases} 1 & \text{when } m \in A, \\ -1 & \text{when } m \in 3 \cdot A. \end{cases}$$

We can simplify $\Lambda_1(x)$ as

$$\Lambda_1(x) = \frac{1}{\pi} \sum_{m \in B} \Big(\varepsilon(m) \sin(2\pi mx) + \sum_{n \in \mathcal{N}_1} \frac{\varepsilon(m)}{n} \sin(2\pi nmx) \Big). \tag{4.11}$$

Finally, we combine (4.9) and (4.11) to get

$$\begin{aligned} & -\frac{2\sqrt{3}\pi}{3} \sum_{t \in \mathcal{N}_2} \frac{\mu(t)\chi(t)}{t} \sum_{m \in A} \Gamma(mtx) + \frac{2\sqrt{3}\pi}{3} \sum_{t \in \mathcal{N}_2} \frac{\mu(t)\chi(t)}{t} \sum_{m \in A} \Gamma(3mtx) \\ & + i\pi \sum_{t \in \mathcal{N}_2} \frac{\mu(t)}{t} \sum_{m \in A} \Lambda(2tmx) \\ = & \sum_{m \in B} \left(\epsilon(m) \cos(4\pi mx) + \sum_{n \in \mathcal{N}_1} \frac{\chi(n)\varepsilon(m)}{n} \cos(4\pi nmx) \right) \\ & + i \sum_{m \in B} \left(\epsilon(m) \sin(4\pi mx) + \sum_{n \in \mathcal{N}_1} \frac{\varepsilon(m)}{n} \sin(4\pi nmx) \right) \\ = & \sum_{m \in B} e^{4\pi imx} \varepsilon(m) + \sum_{m \in B, n \in \mathcal{N}_1} \frac{\varepsilon(m)}{n} \Big((\chi(n)+1) e^{4\pi nmx} + (\chi(n)-1) e^{-4\pi nmx} \Big). \end{aligned} \tag{4.12}$$

Now we can employ Corollary 3.8 and the triangle inequality to (4.12), to obtain

$$2 \sum_{t \in \mathcal{N}_2} \frac{|\mu(t)|}{t} \|\Gamma\|_{L^1(\mathbb{R}/\mathbb{Z})} + \sum_{t \in \mathcal{N}_2} \frac{|\mu(t)|}{t} \|\Lambda\|_{L^1(\mathbb{R}/\mathbb{Z})} \gg \log N.$$

Merten's estimate tells us

$$\sum_{t\in\mathcal{N}_2}\frac{1}{t}\ll\prod_{p<Q}\Big(1+\frac{1}{p}\Big)\asymp\log\log N.$$

Hence we have

$$\max\left\{\left\|\Gamma\right\|_{L^1(\mathbb{R}/\mathbb{Z})},\left\|\Lambda\right\|_{L^1(\mathbb{R}/\mathbb{Z})}\right\}\gg\frac{\log N}{\log\log N}.$$

This implies that there is $t\in\{1,2\}$, such that $\|f_t\|\gg\log N/\log\log N$, and since f_t is balanced, we get $\max_{x\in\mathbb{R}/\mathbb{Z}} f_t(x)\gg\log N/\log\log N$.

With all tools in hand we are going to prove our main theorem.

Proof of Theorem 4.4. Fix $k\geq 1$. We first assume $|A\triangle 3\cdot A|\gg N^{1/2}$. By the result proved earlier in this section, we may assume that for $\Omega_1=(1/6,1/3)$, there is $x_0\in\mathbb{R}/\mathbb{Z}$ such that

$$\sum_{n\in A}\mathbb{1}_{\Omega_1}(nx_0)\gg\frac{N}{6}+c\frac{\log N}{\log\log N},$$

where $c>0$ is an absolute constant. Consider the continuous group homomorphism $\chi:\mathbb{R}/\mathbb{Z}\to\mathbb{R}/\mathbb{Z}$ with $\chi(x)=kx$ for every x. Then the Bohr set $\chi^{-1}(\Omega_1)$ is a union of k disjoint open intervals $I_1,\dots,I_k$ in $\mathbb{R}/\mathbb{Z}$, each of which has measure $1/6k$. It is also easy to see that I_t is $(2k,4k)$-sum-free for every $1\leq t\leq k$. Indeed, suppose that I_1 is not $(2k,4k)$-sum-free, then there are $6k$ elements $a_1,\dots,a_{2k},b_1,\dots,b_{4k}$ in I_1 such that $\sum_{j=1}^{2k}a_j=\sum_{j=1}^{4k}b_j$. We may assume $a_1\leq\cdots\leq a_{2k}$ and $b_1\leq\cdots\leq b_{4k}$. Define $\alpha_r=\frac{1}{k}\sum_{j=rk+1}^{(r+1)k}a_j$, and $\beta_s=\frac{1}{k}\sum_{j=sk+1}^{(s+1)k}b_j$ for all $r\in\{0,1\}$ and $s\in\{0,1,2,3\}$. Since I_1 is an interval, α_0,α_1 and $\beta_0,\dots,\beta_3$ all belong to I_1, and $\sum_{j=0}^{1}\alpha_j=\sum_{j=0}^{3}\beta_j$. Now, using the fact that χ is a group homomorphism, we have $\sum_{j=0}^{1}\chi(\alpha_j)=\sum_{j=0}^{3}\chi(\beta_j)$, and this contradicts the fact that Ω_1 is $(2,4)$-sum-free.

Let $x_1=x_0/k$. By pigeonhole principle, there is $t_0\in\{1,\dots,k\}$ such that

$$\sum_{n\in A}\mathbb{1}_{I_t}(nx_1)\gg\frac{N}{6k}+\frac{c}{k}\frac{\log N}{\log\log N},$$

this finishes the proof of the first case.

Now let us assume $|A \triangle 3 \cdot A| \ll N^{1/2}$. Then A is at least a $(3, 1/2)$-lacunary set. Let $\Omega = (1/3, 2/3)$. By the result proved in Section 4.2, there is $y_0 \in \mathbb{R}/\mathbb{Z}$, such that

$$\sum_{n \in A} \mathbb{1}_\Omega(ny_0) \gg \frac{N}{3} + cN^{1/8},$$

for some constant $c > 0$. Let $y_1 = y_0/2k$, then similarly there is an open interval $I \subseteq \mathbb{R}/\mathbb{Z}$, such that I has length $1/6k$, I is $(2k, 4k)$-sum-free, and

$$\sum_{n \in A} \mathbb{1}_I(ny_1) \gg \frac{N}{6k} + \frac{c}{2k}N^{1/8},$$

this finishes the proof. $\square$

Part II

The Nonabelian Groups

Chapter 5

Preliminaries to topological groups

In this chapter we provide background on topogical groups as needed in Part II of the thesis. This is based on the appendices of the joint paper with Tran and Zhang [106].

5.1 Some results about topological groups

This section gathers some facts about topological groups which is needed in the proof. We begin with the three isomorphism theorems of topological groups. Note that the third isomorphism theorem is almost the same as the familiar result for groups, whereas the first two isomorphism theorems require extra assumptions; see [25, Proposition III.2.24], [25, Proposition III.4.1], and [25, Proposition III.2.22] for details. For this fact, we do not need to assume that G is locally compact. The quotient G/H is equipped with the quotient topology (i.e., $X \subseteq G/H$ is open if and only if it inverse image under the quotient map is open).

Fact 5.1. *Suppose H is a closed normal subgroup of G. Then we have the following:*

1. (First isomorphism theorem) *Suppose $\phi : G \to Q$ is a continuous surjective group homomorphism with* $\ker \phi = H$*. Then the exact sequence of groups*

$$1 \to H \to G \to Q \to 1$$

is an exact sequence of topological groups if and only if ϕ is open; the former condition is equivalent to saying that Q is canonically isomorphic to G/H as topological groups.

2. (Second isomorphism theorem) *Suppose S is a closed subgroup of G and H is compact. Then $S/(S\cap H)$ is canonically isomorphic to the image of SH/H as topological groups. This is also equivalent to saying that we have the exact sequence of topological groups*

$$1 \to H \to SH \to S/(S\cap H) \to 1.$$

3. (Third isomorphism theorem) *Suppose $S \leq G$ is closed, and $H \leq S$. Then S/H is a closed subgroup of G/H. If $S \lhd G$ is normal, then S/H is a normal subgroup of G/H, and we have the exact sequence of topological groups*

$$1 \to S/H \to G/H \to G/S \to 1;$$

this is the same as saying that $(G/H)/(S/H)$ is canonically isomorphic to G/S as topological groups.

We also need the following simple property of locally compact groups [67, Theorem 6.7].

Fact 5.2. *Closed subgroups and quotients of a locally compact group by a closed normal subgroup are locally compact.*

The following lemma holds for all topological group.

Lemma 5.3. *Suppose $X, Y \subseteq G$, X is compact and Y is closed. Then XY is closed.*

Proof. Let a be in $G \setminus XY$. Then $X^{-1}a$ is compact and $X^{-1}a \cap Y = \varnothing$. For each point $x \in X^{-1}a$, we choose an open neighborhood of identity U_x such that $xU_x^2 \cap Y = \varnothing$. Then $(xU_x)_{x\in X^{-1}a}$ is an open cover of $X^{-1}a$. Using the fact that $X^{-1}a$ is compact, we get a subcover $(U_i)_{i=1}^k$. Set $U = \bigcap_{i=1}^k U_i$. It is easy to check that $X^{-1}aU \cap Y = \varnothing$. Then $aU \cap XY = \varnothing$, which implies that XY is closed as a can be chosen arbitrarily. $\square$

The next lemma records a simple fact on compact subgroups.

Lemma 5.4. *If H is a compact subgroup of G, then the quotient map $\pi : G \to G/H$ is a proper map (i.e., the inverse image of compact subsets are compact).*

Proof. Let Ω be a compact subset of G/H. In particular Ω is closed. Hence, $\pi^{-1}(\Omega)$ is closed, so it suffices to find a compact set containing $\pi^{-1}(\Omega)$. Since G is locally compact, we can find an open covering $(U_i)_{i\in I}$ of $\pi^{-1}(\Omega)$ such that U_i has compact closure $\overline{U_i}$ for each $i \in I$. Then $(\pi U_i)_{i\in I}$ is an open cover of Ω as π is open. Using the assumption that Ω is compact, we get a finite $I' \subseteq I$ such that $(\pi(U_i))_{i\in I'}$ is an open cover of Ω. Then $\bigcup_{i\in I'} \overline{U_i}H$ is a compact set containing $\pi^{-1}(\Omega)$. $\square$

5.2 Measures and the modular function

We say that a measure μ on the collection of Borel subsets of G is a *left Haar measure* if it satisfies the following properties:

1. (left-translation-invariant) $\mu(X) = \mu(aX)$ for all $a \in G$ and all measurable sets $X \subseteq G$.

2. (inner and outer regular) $\mu(X) = \sup \mu(K) = \inf \mu(U)$ with K ranging over compact subsets of X and U ranging over open subsets of G containing X.

3. (compactly finite) μ takes finite measure on compact subsets of G.

The notion of a *right Haar measure* is obtained by making the obvious modifications to the above definition. The following classical result by Haar makes the above notions enduring features of locally compact group [67, Theorem 2.20]:

Fact 5.5. *Up to multiplication by a positive constant, there is a unique left Haar measure on G. A similar statement holds for right Haar measure.*

Given a locally compact group G, and μ is a left Haar measure on G. For every $x \in G$, recall that $\Delta_G : x \mapsto \mu_x/\mu$ is the *modular function* of G, where μ_x is a left Haar measure on G defined by $\mu_x(A) = \mu(Ax)$, for every measurable set A. When the image of Δ_G is always

1, we say G is *unimodular*. In general, $\Delta_G(x)$ takes values in $\mathbb{R}^{>0}$. We use $(\mathbb{R}^{>0}, \times)$ to denote the multiplicative group of positive real number together with the usual Euclidean topology. The next fact records some basic properties of the modular function; See [67, Section 2.4].

Fact 5.6. *Let G be a locally compact group. Assuming μ is a left Haar measure and ν is a right Haar measure.*

1. *Suppose H is a normal closed subgroup of G, then $\Delta_H = \Delta_G$. In particular, if $H = \ker \Delta_G$, then H is unimodular.*

2. *The function $\Delta_G : G \to (\mathbb{R}^{>0}, \times)$ is a continuous homomorphism.*

3. *For every $x \in G$ and every measurable set A, we have $\mu(Ax) = \Delta_G(x)\mu(A)$, and $\nu(xA) = \Delta_G^{-1}(x)\nu(A)$.*

4. *There is a constant c such that $\int_G f \,\mathrm{d}\mu = c \int_G f \Delta_G \,\mathrm{d}\nu$ for every $f \in C_c(G)$.*

We use the following integral formula [67, Theorem 2.49] in our proofs.

Fact 5.7 (Quotient integral formula)**.** *Let G be a locally compact group, and let H be a closed normal subgroup of G. Given μ_G, μ_H left Haar measures on G and on H. Then there is a unique left Haar measure $\mu_{G/H}$ on G/H, such that for every $f \in C_c(G)$,*

$$\int_G f(x) \,\mathrm{d}\mu_G(x) = \int_{G/H} \int_H f(xh) \,\mathrm{d}\mu_H(h) \,\mathrm{d}\mu_{G/H}(x).$$

The following fact is a consequence of a result about Haar measure on closed subgroups and quotients [26, Proposition VII. 2.7.10].

Fact 5.8. *Suppose G is nonunimodular, and $\Delta_G : G \to (\mathbb{R}^{>0}, \times)$ is the modular function of G, then we have the following:*

1. *If $K \lhd G$ is a compact normal subgroup of G, $\Delta_{G/K}$ is the modular function of G/K, and $\pi : G \to G/K$ is the quotient map, then we have $\Delta_G = \Delta_{G/K} \circ \pi$.*

2. *If $H \lhd G$ is a closed unimodular group, and μ_H is a Haar measure on H. Suppose G/H is unimodular, and X is a compact subset of H. Then for every $g \in G$, $\mu_H(gXg^{-1}) = \Delta_G(g)\mu_H(X)$.*

5.3 Almost-Lie groups and the Gleason–Yamabe Theorem

In our proof we need the solution of Hilbert's 5th problem, which is known as the Gleason–Yamabe Theorem [73, 169], to reduce the problem into Lie groups. For convenience, we introduce the following terminology. A locally compact group G is an **almost-Lie group** if every open neighborhood U of the identity in G contains a compact $H \lhd G$ such that G/H is a Lie group.

Lemma 5.9. *Suppose G is an almost-Lie group. Then every open subgroup of G and every quotient of G by a closed normal subgroup is an almost-Lie group.*

Proof. We first show that every open subgroup of G is almost-Lie. Let S be an open subgroup of G, and U is an open neighborhood of identity in S. We need to find a compact subgroup K of S such that $K \subseteq U$ and S/K is a Lie group. Since U is also a neighborhood of identity in G, U contains a compact normal subgroup K of G such that G/K is a Lie group. Note that $K \lhd S$. As S is open, S/K is open in G/K and hence a Lie group as desired.

Next, suppose H is a closed normal subgroup of G, and $\pi : G \to G/H$ is the quotient map. If U is an open neighborhood of the identity in G/H, then $\pi^{-1}(U)$ is an open neighborhood of identity in G. Hence, we can get a normal compact subgroup K of G such that $K \subseteq \pi^{-1}(U)$ and that G/K is a Lie group. Then $\pi(K)$ is a compact subgroup of U. With $S = \pi^{-1}(\pi(K))$, we have $\pi(K) = S/H$. Since K is normal in G we have $\pi(K)$ is normal in G/H and thus S is normal in G. Whence by the third isomorphism theorem (Fact 5.1.3), we conclude that $(G/H)/\pi(K) \cong G/S$. By the third isomorphism theorem again, we have

$G/S \cong (G/K)/(S/K)$, thus G/S is a Lie group. $\square$

We use the following strong version of the Gleason–Yamabe Theorem.

Fact 5.10. *We have the following:*

1. (Gleason–Yamabe Theorem) *Suppose G is a locally compact group. Then there is an open subgroup of G which is an almost-Lie group.*

2. *An almost-Lie group G is a Lie group if and only if there is an open neighborhood U of the identity in G that contains no nontrivial compact subgroup of G.*

Fact 5.10.2 is not officially part of the Gleason–Yamabe Theorem. However, the forward direction is an easy fact about the no small subgroup property of Lie groups, and the and backward direction is a direct consequence of Fact 5.10.1.

5.4 Some results about Lie groups

In this section we gather some facts and lemmas about Lie groups and Lie algebras. Throughout the chapter, all the Lie groups are finite dimensional second countable *real* Lie groups.

Fact 5.11. *Closed subgroups and quotient groups of Lie groups are Lie groups.*

The *identity component* of a topological group G is the connected component containing the identity element. The identity component of a topological group G might not be open even if G is locally compact. For instance, there are nondiscrete totally disconnected locally compact groups. For these groups, the identity component only consists of the identity element, and it is not open because the topology is not discrete. Nevertheless, the following holds for Lie groups [91, Proposition 9.1.15].

Fact 5.12. *If G is a Lie group, then the identity component of G is open and is contained in every open subgroups of G.*

In Fact 5.1, we introduce the three isomorphism theorems of topological groups. When G is a Lie group, we can weaken the assumption required for the first two isomorphism theorems; see [24, Proposition 3.11.2, Proposition 3.31].

Fact 5.13. *Suppose G is a Lie group, and H is a closed normal subgroup of G. Then we have the following:*

1. (First isomorphism theorem for Lie groups) *If Q is a Lie group, $\phi : G \to Q$ is a surjective and continuous group homomorphism, and G has countably many connected components. Then Q is isomorphic as a topological group to G/H.*

2. (Second isomorphism theorem for Lie groups) *Suppose G is a finite dimensional Lie group, and S is a closed subgroup of G, and SH is a closed subgroup of G. Then $S/(S \cap H)$ is canonically isomorphic to the image of SH/H as Lie groups. This is also equivalent to saying that we have the exact sequence of Lie groups*

$$1 \to H \to SH \to S/(S \cap H) \to 1.$$

We also need the following fact about maximal compact subgroups consisting of Theorem 14.1.3 (iii) and Theorem 14.3.13 (i) (a) of [91]:

Fact 5.14. *Suppose G is a Lie group with finitely many connected components. Then we have the following:*

1. *All maximal compact subgroups of G are conjugate.*

2. *If $0 \to H \to G \xrightarrow{\pi} G/H \to 0$ is an exact sequence of connected Lie groups, and K is a maximal compact subgroup of G, then $K \cap H$ is a maximal compact subgroup of H, and $\pi(K)$ is a maximal compact subgroup of G/H.*

We also use the following simple classification results for Lie groups.

Fact 5.15. *Let G be a connected Lie group.*

1. *If G has dimension 1, then it is isomorphic to either $\mathbb{R}$ or $\mathbb{T}$ as topological groups.*

2. *If G is a solvable group with dimension d, and the maximal compact subgroups of G have dimension m. Then G is diffemorphic to $\mathbb{T}^m \times \mathbb{R}^{d-m}$. Moreover, if G is compact, then $G \cong \mathbb{T}^d$.*

We say that a topological group G is a covering group of a topological group G with covering homomorphism ρ if $\rho : G \to G'$ is a topological group homomorphism which is also a covering map. The following is a consequence of [91, Theorem 9.5.4]:

Fact 5.16. *Suppose that G and G' are connected Lie groups and that G is a covering group of G' with covering homomorphism ρ. Then $\ker \rho$ is a closed normal subgroup of the center $Z(G)$ of G.*

We end this section with a lemma about conjugate actions on compact sets in Lie groups.

Lemma 5.17. *For a Lie group G and a closed normal subgroup H, if a precompact $A \subseteq H$ such that the closure of A is in B and B is a relative open subset in H, then the following holds: When $g \in G$ is sufficiently close to id_G, we have $gAg^{-1} \subseteq B$.*

Proof. We prove the lemma by contradiction. Assuming there exist sequences $g_n \to \mathrm{id}$ and $\{h_n\} \subseteq A$ such that $g_n h_n g_n^{-1} \notin B$. Since A is precompact we may assume $h_n \to h \in \overline{A}$. But then $g_n h_n g_n^{-1} \to h \in \overline{A}$. This contradicts the fact that each $g_n h_n g_n^{-1}$ is in the closed set $H \setminus B$ that does not meet $\overline{A}$. Hence the assumption is false and the conclusion holds. □

5.5 Solvable and Semisimple Lie groups

From [91, Section 9.1], there is a functor $\mathbf{L}$ from the category of Lie groups to the category of Lie algebras that assigns to each Lie group G its Lie algebra $\mathbf{L}(G)$ and a to Lie group morphism $\phi : G \to H$ its tangent morphism $\mathbf{L}(\phi) : \mathbf{L}(G) \to \mathbf{L}(H)$ of Lie algebras. We will adopt a more colloquial language in this chapter, invoking this functor implicitly.

Fact 5.18. *Suppose G and H are Lie groups, and $\mathfrak{g}$ and $\mathfrak{h}$ are their Lie algebras. If H is a subgroup of G, then $\mathfrak{h}$ is a subalgebra of $\mathfrak{g}$. If H is a normal subgroup of G, then $\mathfrak{h}$ is an ideal in $\mathfrak{g}$, and $\mathfrak{g}/\mathfrak{h}$ is canonically isomorphic to the Lie algebra of G/H.*

Suppose $\mathfrak{g}$ is the Lie algebra of G. The exponential function $\exp : \mathfrak{g} \to G$ is defined as in [91, Section 9.2]. We will use the functoriality of the exponential function [91, Proposition 9.2.10]

Fact 5.19. *Suppose G and H are Lie groups, $\phi : G \to H$ is a homomorphism of Lie groups, $\mathfrak{g}$ and $\mathfrak{h}$ are the Lie algebras of G and H, $\alpha : \mathfrak{g} \to \mathfrak{h}$ is the tangent morphism of ϕ, and $\exp_G : \mathfrak{g} \to G$ and $\exp_H : \mathfrak{h} \to H$ are the exponential maps. Then $\exp_H \circ \alpha = \exp_G \circ \phi$. In other words, the following diagram commutes:*

$$\begin{tikzcd} G \arrow[r, "\phi"] & H \\ \mathfrak{g} \arrow[u, "\exp_G"] \arrow[r, "\alpha"] & \mathfrak{h} \arrow[u, "\exp_H"'] \end{tikzcd}$$

Suppose $\mathfrak{g}$ is a Lie algebra. The *derived Lie algebra* $[\mathfrak{g}, \mathfrak{g}]$ of $\mathfrak{g}$ is the subalgebra of $\mathfrak{g}$ generated by the Lie brackets of the pairs of elements of $\mathfrak{g}$. We say that $\mathfrak{g}$ is *solvable* if the derived sequence

$$\mathfrak{g} \geq [\mathfrak{g}, \mathfrak{g}] \geq [[\mathfrak{g}, \mathfrak{g}], [\mathfrak{g}, \mathfrak{g}]] \geq \dots$$

eventually arrive at the 0-algebra. A Lie group is *solvable* if its Lie algebra is solvable. The following is a consequence of [91, Proposition 5.4.3]:

Fact 5.20. *Every subalgebra and quotient algebra of a solvable Lie algebra is solvable. Hence, every closed subgroup and quotient group of a solvable Lie group is solvable.*

The following is another consequence of [91, Proposition 5.4.3]:

Fact 5.21. *Suppose $\mathfrak{g}$ is a Lie algebra. Then $\mathfrak{g}$ has a largest solvable subalgebra $\mathfrak{q}$. If G is a Lie group with Lie algebra $\mathfrak{g}$ and $\exp : \mathfrak{g} \to G$ is the exponential map, then $Q = \langle \exp(\mathfrak{q}) \rangle$ is the largest closed connected solvable subgroup of G. Hence, Q is a characteristic subgroup of G.*

The subalgebra $\mathfrak{q}$ as in Fact 5.21 is called the *radical* of $\mathfrak{g}$, and the subgroup Q as in Fact 5.21 is called the *radical* of G. A Lie algebra is *semisimple* if it has trivial radical. A lie group is *semisimple* if its Lie algebra is semisimple, or equivalently, if it has trivial radical. The following results follows from [91, Proposition 5.4.3]:

Fact 5.22. *Let G be a connected Lie group. Let Q be the radical of G. Then $S = G/Q$ is a semisimple Lie group.*

A Lie group is *simple* if its Lie algebra is simple. Note that a simple Lie group needs not to be simple as a group. We use the following fact for simple Lie groups.

Fact 5.23. *A connected Lie group G is a simple Lie group if and only if all its normal proper subgroups are discrete, and contained in $Z(G)$.*

Suppose $\mathfrak{g}$ is a finite dimensional Lie algebra. For $x \in \mathfrak{g}$, let $\mathrm{ad}x : \mathfrak{g} \to \mathfrak{g}, y \mapsto [x, y]$. Then ad is an endomorphism of $\mathfrak{g}$. The *Cartan–Killing form* of $\kappa_{\mathfrak{g}} : \mathfrak{g} \times \mathfrak{g} \to \mathbb{R}$ is given by

$$\kappa_{\mathfrak{g}}(x, y) = \mathrm{tr}(\mathrm{ad}x \ \mathrm{ad}y).$$

The Cartan–Killing form is invariant under an automorphism of $\mathfrak{g}$ as this corresponds to a change of basis. The following fact is from [91, Lemma 5.5.8]

Fact 5.24. *Suppose $\mathfrak{g}$ is a Lie algebra, $\kappa_{\mathfrak{g}}$ is the Cartan–Killing form of $\mathfrak{g}$, and $\mathfrak{h}$ is an ideal of $\mathfrak{g}$. Then the orthogonal space $\mathfrak{h}^{\perp}$ of $\mathfrak{h}$ with respect to $\kappa_{\mathfrak{g}}$ is also an ideal. If $\mathfrak{g}$ is semisimple, then $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{h}^{\perp}$ and $\kappa_{\mathfrak{g}} = \kappa_{\mathfrak{h}} \oplus \kappa_{\mathfrak{h}^{\perp}}$ where $\kappa_{\mathfrak{h}}$ and $\kappa_{\mathfrak{h}^{\perp}}$ are the Cartan–Killing form of $\mathfrak{h}$ and $\mathfrak{h}^{\perp}$.*

The following fact follows from [91, Lemma 5.5.13]. It is also a consequence of Fact 5.24 and the alternative characterization of semisimple Lie algebras as those whose Cartan–Killing form is nondegenerate.

Fact 5.25. *Every ideal and quotient algebra of a semisimple Lie algebra is semisimple. Hence, every normal subgroup and quotient group of a semisimple Lie group is semisimple.*

The first and second assertions in the following fact are immediate consequences of Facts 5.21, 5.20, 5.25

Fact 5.26. *If G is a connected semisimple Lie group, then its center $Z(G)$ is a finitely generated discrete group, the quotient map $\rho : G \to G/Z(G)$ is a covering map.*

The following fact is a consequence of [91, Proposition 9.5.2 and Theorem 9.5.4].

Fact 5.27. *If G and G' are connected Lie groups, $\rho : G \to G'$ is covering map, $Z(G)$ and $Z(G')$ are the centers of G and G'. Then we have $\ker \rho \leq Z(G)$ and $Z(G') = Z(G)/\ker \rho$.*

The first assertion in the following fact is known as Weyl's theorem on Lie groups with semisimple compact Lie algebra [91, Theorem 12.1.17].

Fact 5.28. *If G is a connected semisimple Lie group with compact Lie algebra, then G is compact and $Z(G)$ is finite.*

The following Fact is a consequence of Fact 5.28 and the result in [156]. This can also be proven directly using [91, Proposition 13.1.10 (ii)]; we thank Jinpeng An for pointing this out to us.

Fact 5.29. *If G is a simply connected simple Lie group, then the center $Z(G)$ of G has rank at most* 1.

Suppose $\mathfrak{g}$ is a finite dimensional Lie algebra with Cartan–Killing form $\kappa_{\mathfrak{g}}$. A Lie algebra automorphism τ of $\mathfrak{g}$ is a *Cartan involution* if $\tau^2 = \mathrm{id}_{\mathfrak{g}}$ and $(x, y) \mapsto -\kappa_{\mathfrak{g}}(x, \tau(y))$ is a positive definite bilinear form. The following fact is [91, Theorem 13.2.10]

Fact 5.30. *Let $\mathfrak{g}$ be a semisimple Lie algebra. Then $\mathfrak{g}$ has a Cartan involution τ.*

We refer the reader to [114, Section 6.4] for the full definition of Iwasawa decomposition; we will need the following fact which is a consequence of [114, Theorem 6.31, Theorem 6.46] and [91, Corollary 12.2.3].

Fact 5.31 (Iwasawa decomposition)**.** *Suppose G is a connected semisimple Lie group with Lie algebra $\mathfrak{g}$, τ is a Cartan's involution of $\mathfrak{g}$, $\mathfrak{k}$ the subalgebra of $\mathfrak{g}$ fixed by τ, and $\exp : \mathfrak{g} \to G$ is the exponential map. Then there is an Iwasawa decomposition $G = KAN$ such that the following holds:*

1. *the multiplication map*

$$\Phi : K \times A \times N \to G : (k, a, n) \mapsto kan$$

   is a diffeomorphism.

2. *$K = \exp(\mathfrak{k})$ is a connected closed subgroup of G, $Z(G) \subseteq K$, and K is a maximal compact subgroup of G if $Z(G)$ is finite.*

3. *A is an abelian closed subgroup of G, N is a nilpotent closed subgroup of G, and both A and N are simply connected.*

4. *$Q = AN$, we have that Q is a solvable closed subgroup of G, and $N \lhd Q$.*

The following fact is a consequence of the definition of Iwasawa decomposition in [114, Section 6.4].

Fact 5.32. *If G is a noncompact semisimple Lie group with Iwasawa decomposition $G = KAN$, then AN has dimension at least 2.*

Chapter 6

Minimal and nearly minimal measure expansions in connected unimodular groups

Let G be a connected unimodular group equipped with a (left and hence right) Haar measure μ_G, and suppose $A, B \subseteq G$ are nonempty and compact. An inequality by Kemperman [112] gives us

$$\mu_G(AB) \geq \min\{\mu_G(A) + \mu_G(B), \mu_G(G)\}.$$

Our first result determines the conditions for the equality to hold, providing a complete answer to a question asked by Kemperman in 1964. Our second result characterizes compact and connected G, A, and B that nearly realize equality, with quantitative bounds having the sharp exponent. This can be seen up-to-constant as a $(3k-4)$-theorem for this setting and confirms the connected case of conjectures by Griesmer [82] and by Tao [162]. As an application, we get a measure expansion gap result for connected compact simple Lie groups.

The tools developed in our proof include an analysis of the shape of minimal and nearly minimal expansion sets, a bridge from this to the properties of a certain pseudometric, and a construction of appropriate continuous group homomorphisms to either $\mathbb{R}$ or $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ from the pseudometric.

This chapter is based on joint work with Tran [105].

6.1 Introduction

6.1.1 Background

The Cauchy–Davenport theorem asserts that if X and Y are nonempty subsets of the group $\mathbb{Z}/p\mathbb{Z}$ of prime order p, then

$$|X+Y| \geq \min\{|X|+|Y|-1, p\},$$

where we set $X+Y := \{x+y : x \in X, y \in Y\}$. The condition for the equality to happen is essentially given by Vosper's theorem [166], which states that if

$$1 < |X|, |Y|, \text{ and } |X+Y| = |X|+|Y|-1 < p-1$$

then X and Y must be arithmetic progressions with the same common difference. When the equality nearly happens, one might expect that X and Y are instead contained in arithmetic progressions with slightly larger cardinalities. This was confirmed with a sharp exponent bound by Freiman [69] for small $|X|$ and $|Y|$. The optimal statement believed to be true for $X = Y$, known as the $(3k-4)$-conjecture for $\mathbb{Z}/p\mathbb{Z}$, remains wide open more than 60 years after the corresponding statement for $\mathbb{Z}$ was proven by Freiman [68]. In the mean time, we see many similar results obtained for other abelian groups; see e.g. [115, 111, 51, 79, 162, 82, 123]. In another direction, there has been much progress in the study of small expansions in the nonabelian settings; see [36], in particular, for the classification of approximate groups by Breuillard, Green, and Tao; see also e.g. [89, 32, 33, 102, 143, 19, 103]. These two trends together suggest that the theory of minimal and nearly minimal expansion can be extended to the nonabelian settings as well. In this chapter, we take a step towards realizing this intuition by considering an inequality by Kemperman, the continuous nonabelian counterpart of the Cauchy–Davenport theorem, and effectively determining the necessary and sufficient conditions for equality and near equality to happen.

Throughout, let G be a connected locally compact group, and μ_G a left Haar measure on G. We further assume that G is unimodular (i.e., the measure μ_G is invariant under right translation), so μ_G behaves like an appropriate notion of size. This assumption holds in many situations of interest (e.g, when G is compact, discrete, a nilpotent Lie group, a semisimple Lie group, etc). As usual, for $A, B \subseteq G$, we set $AB := \{ab : a \in A, b \in B\}$ and let A^n be the n-fold product of A for $n \in \mathbb{N}^{>0}$. In [112], Kemperman proved that if $A, B \subseteq G$ are nonempty and compact, then

$$\mu_G(AB) \geq \min\{\mu_G(A) + \mu_G(B), \mu_G(G)\}.$$

This generalizes earlier results for one-dimensional tori, n-dimensional tori, and abelian groups by Raikov [144], Macbeath [129], and Kneser [115].

The problem of determining when equality holds in the Kemperman inequality was proposed in the same paper [112]. After handling a number of easy cases, the problem can be reduced to classifying all connected and unimodular group G and pairs (A, B) of compact subsets of G such that

$$0 < \mu_G(A), \mu_G(B), \text{ and } \mu_G(AB) = \mu_G(A) + \mu_G(B) < \mu_G(G).$$

We call such (A, B) a **minimally expanding** pair on G. It is easy to see that, if I and J are closed intervals in $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, such that I and J have positive measures and the total of their measures is strictly smaller than $\mu_{\mathbb{T}}(\mathbb{T})$, then $I + J$ is an interval with length the total length of I and J. Hence, such (I, J) is a minimally expanding pair on $\mathbb{T}$. More generally, when G is a compact group, $\chi : G \to \mathbb{T}$ is a continuous surjective group homomorphism, I and J are as before,

$$A = \chi^{-1}(I) \text{ and } B = \chi^{-1}(J),$$

we can check by using the Fubini theorem that (A, B) is a minimally expanding pair. Note that an arithmetic progression on $\mathbb{Z}/p\mathbb{Z}$ is the inverse image under a group homomorphism $\phi : \mathbb{Z}/p\mathbb{Z} \to \mathbb{T}$ of an interval on $\mathbb{T}$, so this example is the counterpart of Vosper's classification. Another obvious example is when G is a noncompact group, $\chi : G \to \mathbb{R}$ is a continuous surjective group homomorphism with compact kernel, I and J are compact intervals in $\mathbb{R}$ of positive measures, $A = \chi^{-1}(I)$, and $B = \chi^{-1}(J)$. One might optimistically conjecture, in analogy with Vosper's theorem, that there are no other G, A, and B such that (A, B) is a minimally expanding pair on G.

In view of the earlier discussions, for compact $A, B \subseteq G$, we say that (A, B) is a **δ-nearly minimally expanding pair** on G if

$$0 < \mu_G(A), \mu_G(B), \text{ and } \mu_G(AB) < \mu_G(A) + \mu_G(B) + \delta \min\{\mu_G(A), \mu_G(B)\} < \mu_G(G).$$

The problem of determining when equality nearly holds in the Kemperman inequality can be then reasonably interpreted as classifying all connected and unimodular groups G and δ-nearly minimally expanding pairs (A, B) on G. In analogy with the discussion for the Cauchy–Davenport theorem, we hope for an answer along the following line: If G is compact, and (A, B) is a δ-nearly minimally expanding pair on G with small δ, then there is a continuous and surjective group homomorphism $\chi : G \to \mathbb{T}$, compact interval $I, J \subseteq T$, and small ε, such that

$$A \subseteq \chi^{-1}(I), B \subseteq \chi^{-1}(J), \mu_G(\chi^{-1}(I) \setminus A) < \varepsilon, \text{and } \mu_G(\chi^{-1}(I) \setminus B) < \varepsilon. \tag{6.1}$$

The optimistic conjecture for noncompact groups is similar, but with $\mathbb{T}$ replaced by $\mathbb{R}$ and an extra condition that χ has compact kernel.

Under the extra assumption that G is abelian, the optimistic conjectures for both classification problems were more or less confirmed before our work. In the same paper [115] mentioned earlier, Kneser solved the classification problem for equality with the answer we hope for. For the near equality problem, when $G = \mathbb{T}^d$, the desired classification was obtained by Bilu [16], and later improved by Candela and De Roton [39] for a special case when $d = 1$. When G is a general abelian group, a classification result was obtained by Tao [162] for compact G, and by Griesmer [82] when G is noncompact. Griesmer also proved more general results for disconnected groups [81, 82]. The results by Griesmer [81, 82] and by Tao [162] used nonstandard analysis methods, and do not provide how ε depends on δ in (6.1). A sharp exponent classification result (i.e., $\varepsilon = O(\delta)$) for compact abelian groups was obtained very recently by Christ and Iliopoulou [43]. Results with sharp exponent bounds are likely the best that one can achieve without solving the $(3k-4)$-conjecture for $\mathbb{Z}/p\mathbb{Z}$.

For nonabelian G, not much was known earlier than this paper. In closest proximity to what we are doing, Björklund considered in [17] a variation of Kemperman's inequality and the equality classification problem without assuming that G is connected while assuming additionally that G is compact, second countable, and has abelian identity component, and the sets A and B are "spread out" (i.e., far away from being subgroups). The only common case to this and our current setting with the connectedness assumption happens when G is abelian and connected. This is a case already covered by Kneser's classification result.

Toward showing that appropriate versions of the optimistic conjectures also hold for the nonabelian classification problem, there is an important new challenge: While the desired conclusions for the abelian setting are mainly about the structure of (A, B), the structure of G is also highly involved for the nonabelian setting. If $G = \mathrm{SO}_3(\mathbb{R})$, for example, one would not be able to find a minimally expanding pair according to the optimistic answers because there is no continuous surjective group homomorphism from $\mathrm{SO}_3(\mathbb{R})$ to $\mathbb{T}$. On the other hand, one can always find a continuous and surjective group homomorphism from a compact connected nontrivial abelian group to $\mathbb{T}$ and use this to construct minimally expanding pairs. For a noncompact abelian group G, the requirement that the kernel must be compact imposes some constraint on the group G, but this is still comparatively mild.

The above challenge connects our problem to the subject of small expansions in nonabelian groups, a fascinating topic that brings together ideas from different areas of mathematics. The phenomenon that expansion rate encodes structural information about the group can already be seen through the following famous theorem by Gromov [83] in geometric group theory: If G is a group generated by a finite set $X = X^{-1}$, and the cardinality of X^n grows polynomially as a function of n, then G must be virtually nilpotent. A more recent result by Breuillard indicates that some of the analysis goes through for locally compact groups [35]. Even more suggestive is the classifications of approximate groups in [36] mentioned earlier (see the definition in Section 6.6.3). In our proof, we will use the continuous version of the result in [36]; this was proven in the thesis of Carolino [40] and can also be deduced from the result in [36] using a result of Massicot–Wagner [131]. The ideas in the proof of these results can be traced back to the solution of Hilbert's Fifth problem by Montgomery–Zippin [137], Gleason [73], and Yamabe [169], which we will also use later on. Finally, let us mention that these stories are also closely tied to the study of definable groups in model theory. This is the natural habitat of the aforementioned result by Massicot–Wagner [131], and also of Hrushovski's Lie model theorem [102], a main ingredient for the proof of the main theorem in [36].

Before getting to the results, we briefly survey a number of works for nonabelian groups which are thematically relevant but use different techniques. When G is finite and $A, B \subseteq G$ are nonempty, generalizing the Cauchy–Davenport inequality, DeVos showed in [53] that $|AB| \geq \min\{|A| + |B| - |H|, |G|\}$ with H a proper subgroup of G with maximum cardinality. In the same paper, DeVos classifies all situations where equality can happen. In [18], Björklund and Fish studied an expansion problem with respect to upper Banach density in amenable nonabelian groups and obtained conclusions with similar flavor. Under model-theoretic assumptions, Terry, Conant, and Pillay [46, 45] obtained results that are surprisingly similar to ours.

It would also be interesting to study a different minimal and nearly minimal measure expansion problem where we fix a connected unimodular group G instead of letting G range over all connected unimodular group G as we are doing here. When G is $\mathbb{R}^n$, Kemperman inequality is a consequence of the Brunn–Minkowski inequality

$$\mu_G(AB)^{1/n} \geq \mu_G(A)^{1/n} + \mu_G(B)^{1/n}.$$

This inequality also holds for nilpotent G [133, 84, 160]. The equality holds in the Brunn–Minkowski inequality for $\mathbb{R}^n$ if and only if A and B are homothetic convex subsets of $\mathbb{R}^n$. This was a result by Brunn and Minkowski when A and B are further assumed to be convex, and a result by Lyusternik [128], Henstock and Macbeath [90] in the general case. A qualitative answer for the near equality Brunn–Minkowski problem for $\mathbb{R}^n$ is obtained by Christ [42], and a quantitative version is obtained by Figalli and Jerison [66]. We do not pursue this direction further here.

6.1.2 Statement of main results

Our first main result determines the conditions for equality to happen in the Kemperman inequality answering a question by Kemperman in [112]. Scenario (v) and (vi) in the theorem is a classification of the groups G and minimally expanding pairs (A, B) on G.

Theorem 6.1. *Let G be a connected unimodular group, and A, B be nonempty compact subsets of G. If*

$$\mu_G(AB) = \min(\mu_G(A) + \mu_G(B), \mu_G(G)).$$

then we have the following:

(i) $\mu_G(A) + \mu_G(B) = 0$ *implies* $\mu_G(AB) = 0$*;*

(ii) $\mu_G(A) + \mu_G(B) \geq \mu_G(G)$ *implies* $AB = G$*;*

(iii) $\mu_G(A) = 0$ *and* $0 < \mu_G(B) < \mu_G(G)$ *implies that there is a compact proper subgroup* H

of G such that $A \subseteq gH$ for some $g \in G$, and $B = HB$;

(iv) $0 < \mu_G(A) < \mu_G(G)$ *and* $\mu_G(B) = 0$ *imply that there is a compact proper subgroup* H *of* G *such that* $A = AH$, *and* $B \subseteq Hg$ *for some* $g \in G$;

(v) $0 < \min\{\mu_G(A), \mu_G(B), \mu_G(G) - \mu_G(A) - \mu_G(B)\}$, *and* G *is compact together imply that there is a surjective continuous group homomorphism* $\chi : G \to \mathbb{T}$ *and compact intervals* I *and* J *in* $\mathbb{T}$ *with* $I + J \neq \mathbb{T}$ *and* $\mu_{\mathbb{T}}(I), \mu_{\mathbb{T}}(J) > 0$ *such that* $A = \chi^{-1}(I)$ *and* $B = \chi^{-1}(J)$;

(vi) $0 < \min\{\mu_G(A), \mu_G(B)\}$, *and* G *is not compact together implies that there is a surjective continuous group homomorphism* $\chi : G \to \mathbb{R}$ *with compact kernel and compact intervals* I *and* J *in* $\mathbb{R}$ *with* $\mu_{\mathbb{R}}(I), \mu_{\mathbb{R}}(J) > 0$ *such that* $A = \chi^{-1}(I)$ *and* $B = \chi^{-1}(J)$.

Moreover, $\mu_G(AB) = \min(\mu_G(A) + \mu_G(B), \mu_G(G))$ *holds if and only if we are in exactly one of the implied scenarios in (i-vi).*

Next we obtain a classification of nearly minimally expanding pairs. This answers some questions by Griesmer [82] and confirms a conjecture by Tao [162, Conjecture 5.1], under the extra assumption of connectedness.

Theorem 6.2. *Let* G *be a connected compact group,* μ_G *be a normalized Haar measure on* G, *and* A, B *be compact subsets of* G *with positive measure. Set*

$$s = \min\{\mu_G(A), \mu_G(B), 1 - \mu_G(A) - \mu_G(B)\}.$$

Given $\varepsilon > 0$, *there is a constant* $K = K(s)$ *independent of* G, *such that if* $\delta < K\varepsilon$ *and*

$$\mu_G(AB) < \mu_G(A) + \mu_G(B) + \delta \min\{\mu_G(A), \mu_G(B)\}.$$

Then there is a surjective continuous group homomorphism $\chi : G \to \mathbb{T}$ *together with two* *compact intervals* $I, J \subseteq \mathbb{T}$ *with*

$$\mu_{\mathbb{T}}(I) - \mu_G(A) < \varepsilon\mu_G(A), \quad \mu_{\mathbb{T}}(J) - \mu_G(B) < \varepsilon\mu_G(B),$$

and $A \subseteq \chi^{-1}(I)$, $B \subseteq \chi^{-1}(J)$.

It worth noting that the linear dependence between ε and δ is the best possible up to a constant factor. As an application of our main result, we obtain a measure expansion gap result for sets in connected compact simple Lie groups.

Theorem 6.3 (Expansion gaps in compact simple Lie groups)**.** *There is a constant* $\eta > 0$ *such that the following holds. Let* $d > 0$ *be an integer. There is a constant* $C > 0$ *only depending on* d *such that if* G *is a connected compact simple Lie group of dimension* d*, and* A *is a compact set of* G *with* $0 < \mu_G(A) < C$*, then*

$$\mu_G(A^2) > (2 + \eta)\mu_G(A).$$

We can take $\eta > 10^{-10}$.

We did not try to optimise the constant η of Theorem 6.3 in this chapter.

One may compare Theorem 6.3 with expansion gaps for finite sets. The study of the latter problem was initiated by Helfgott [89] where he proved an expansion gap in $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$. Results on the expansions for finite sets are one of the main ingredients in proving many of spectral gap results. For example, the result by Helfgott is largely used in the proof by Bourgain and Gamburd [30, 31]. De Saxcé proved in [49] an expansion gap results in simple Lie groups, which is used in the later proof of spectral gap result [13, 34]. For more background in this direction we refer the reader to [37, 161].

6.1.3 Notation and convention

Throughout let m and n range over the set $\mathbb{N} = \{0, 1, \ldots\}$ of natural numbers. Let G be a locally compact group equipped with a left Haar measure μ_G. Let H range over closed subgroups of G equipped with a left Haar measure μ_H. We normalize μ_G (i.e., $\mu_G(G) = 1$) when G is compact, and do likewise when H is compact. We let G/H and $H\backslash G$ denote the left coset space and the right coset space of G with respect to H. Given a coset decomposition, say G/H, a fiber of a set $A \subseteq G$ refers to $A \cap xH$ for some $xH \in G/H$. We also use μ_H to denote the fiber lengths in the chapter, that is we sometimes write $\mu_H(A \cap xH)$ to denote $\mu_H(x^{-1}A \cap H)$. A constant in this chapter is always a positive real number, and by Lie group we mean a real Lie group with finite dimension.

6.2 Outline of the argument

In this section, we informally explain some of the major new ideas of the proofs. We decided to write a slightly longer outline as some of the later computations are rather technical.

6.2.1 Overview of the strategy

We will explain here the main steps of the proof of Theorem 6.1 and Theorem 6.2 illustrated by simple examples. The focus will be on the plausibility of the argument and how they fit together to resolve the main difficulties of the problem. More detailed discussion will be given in Sections 6.2.2, 6.2.3, and 6.2.4.

After handling a number of easy cases, the proofs of Theorem 6.1 and Theorem 6.2 require constructing appropriate continuous and surjective group homomorphisms into either $\mathbb{R}$ or $\mathbb{T}$ under the given data of a minimal or nearly minimal expanding pair (A, B) on G. The key difficulty of the problem is that many methods in the abelian setting (e-transform, fourier analytic, etc) have no obvious generalization to the nonabelian setting.

Instead, we will use a method in the direction of the solution of Hilbert's Fifth problem, namely, making use of pseudometrics on G (i.e., maps $d : G \times G \to \mathbb{R}$ satisfying all the properties of a metric except $d(g_1, g_2) = 0$ implying $g_1 = g_2$). With some caveats, our proof can be thought of as having three steps: obtaining a suitable pseudometric on G, constructing an appropriate group homomorphism from the pseudometric, and deduce an informative description about (A, B) from the pseudometric. As we want to construct a group homomorphism into $\mathbb{R}$ or $\mathbb{T}$ and not any other Lie groups, the new challenge is to develop properties that allow us to control the desired target group and to obtain a pseudometric on G satisfying these properties.

Let us first explain the second step to see that the pseudometric idea is plausible. Until the end of this section, we will focus on the case where G is compact with the other case treated similarly. If $\chi : G \to \mathbb{T}$ is a continuous and surjective group homomorphism, and $d_{\mathbb{T}}$ is the Euclidean metric on $\mathbb{T}$ (i.e. $d_{\mathbb{T}}(r + \mathbb{Z}, s + \mathbb{Z}) = \min\{|r - s|, |r - (s - 1)|\}$ for $0 \leq r \leq s < 1$), then setting $d(g_1, g_2) = d_{\mathbb{T}}(\chi(g_1), \chi(g_2))$, it is easy to see that d is a pseudometric on G with the "linear" property that

$$d(g_1, g_3) = |d(g_1, g_2) \pm d(g_2, g_3)|$$

for all $g_1, g_2, g_3 \in G$ such that $\max\{d(g_1, g_2), d(g_2, g_3), d(g_1, g_3)\} < 1/4$. Moreover, in the above situation we also have

$$\ker \chi = \{g \in G : d(g, \mathrm{id}_G) = 0\},$$

so the group homomorphism χ can be recovered from the pseudometric. In the second step of our proof, we will define the weaker properties *locally linear* and *locally almost linear*, and show that more or less under these conditions, we can obtain an appropriate homomorphism into $\mathbb{T}$. More details will be given in Section 6.2.3.

Now we would like to construct a locally (almost) linear pseudometric on G. There is a caveat: our primary technique discussed later only works for Lie groups. So we need an extra step in our strategy (also with an underlying pseudometric idea) where we use the Gleason–Yamabe Theorem to obtain a group homomorphism π from G to a Lie group G' with compact kernel H. This is relatively standard, but there are two important arrangements we need to make. We need to choose H carefully to ensure that it is connected and $\mu_{G'}(\pi(A)) + \mu_{G'}(\pi(B)) < \mu_G(G)$; in the third step, we will show that these conditions are enough to imply that $(\pi(A), \pi(B))$ essentially has nearly minimal expansion on G'. We also need to ensure that the dimension of G' is bounded to be able to to get our sharp exponent bound later on. This requires the application of the continuous version of the result in [36], which is proven in Carolino's thesis [40].

Now we focus on the case where G is a compact connected Lie group with dimension bounded from above by a constant, where the structure of G can be studied through its torus subgroups. Using a submodularity argument as in [162], we can carefully modify the original (A, B) to ensure that $\mu_G(A)$ and $\mu_G(B)$ are relatively small compared to $\mu_G(G)$ but still large compared to the error. The pseudometric we need is easy to define:

$$d(g_1, g_2) := \mu_G(A) - \mu_G(g_1 A \cap g_2 A).$$

The real challenge is to show that this is locally (almost) linear. This is achieved by gaining an understanding of the "shape" of the pair (A, B) at different levels. Keeping in mind that we already have the desired classification of minimal and nearly minimal expanding pairs for $\mathbb{T}$, we will choose a suitable one-dimension torus subgroup of G and use it as a tool to probe for information about (A, B).

Thinking of the one-dimensional torus subgroups of G as specifying the "directions" in G, we choose a one-dimensional torus subgroup $T \leq G$ such that for all $gT \in G/T$, the "length" $\mu_T(g^{-1}A \cap T)$ of the "left fiber" $gT \cap A$ of A satisfies

$$\mu_T(g^{-1}A \cap T) < \frac{1}{100},$$

and a similar condition holds for a "right fiber" $B \cap Tg$ of B with $Tg \in T\backslash G$. The existence of such T is by no mean obvious, and is saying that when (A, B) is nearly minimally expanding, A and B cannot be "Kakeya sets". More details will be given in the beginning of Section 6.2.2.

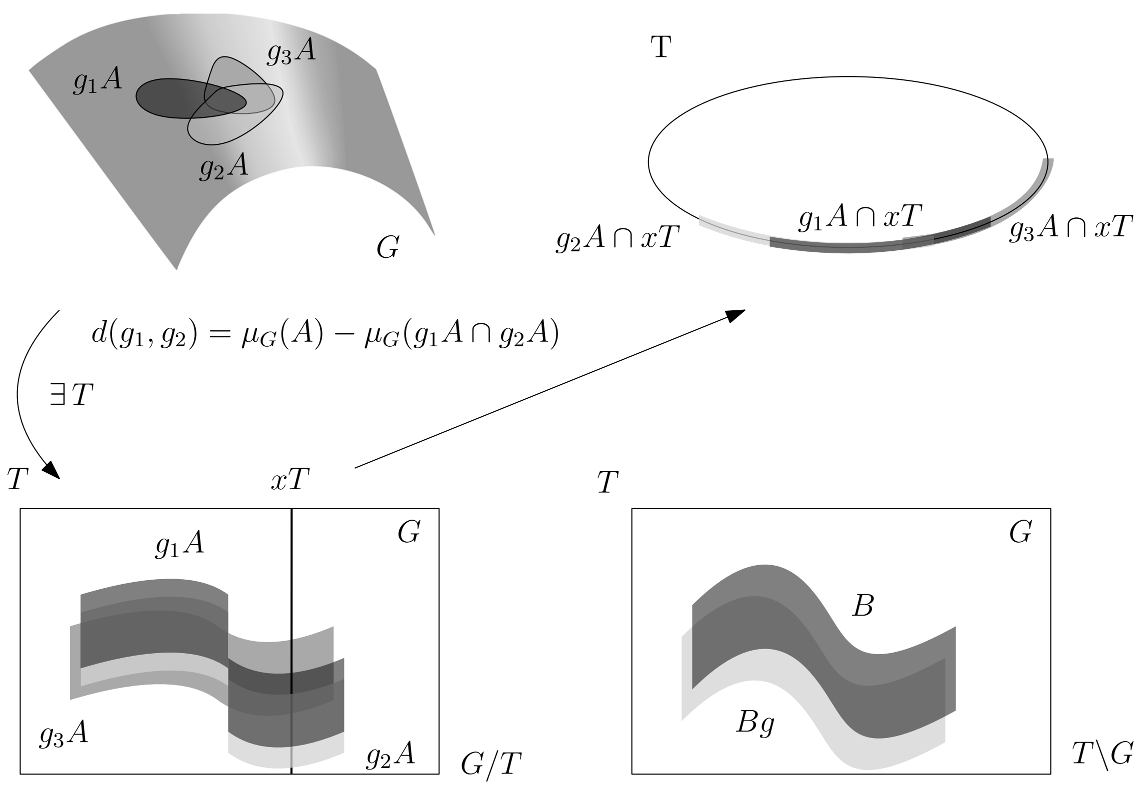


Figure 6.1: Intuition of the ideas in the proof.

We visualize G in two ways: a rectangle with the horizontal side representing G/T and each of the vertical section representing a left cosets of T, and a similar dual picture for $T\backslash G$; see the bottom right half of Figure 6.1. The main idea is to show that g_1A, g_2A, g_3A, and B geometrically look like in the picture with $g_1, g_2, g_3 \in G$ in a suitable neighborhood of id_G. (For instance, we want the "fibers" of A and B to be intervals of T, all the nonempty "left fibers" of $g_1 A$ to have similar "lengths", the "width" of A to be almost the same as that of B, the translated copy $g_1 A$ to only "move vertically" compared to $g_1 A$, and the "moving up" and "moving down" divisions in $g_2 A$ and $g_3 A$ to be almost the same. As a very small hint of why these are true, imagine that one of the fiber in B is too large compared to the rest. Then the product of A with that fiber has already much larger measure than $\mu_G(A)+\mu_G(B)$) One can see that the locally (almost) linearity follows from this picture. The key point is that, the global information about the pseudometric can be deduced via the geometrical shape from the information on a generic fiber xT, where the classification of nearly minimally expanding pairs is known. More detailed discussion is given in Section 6.2.2.

Finally, we discuss the third step of deducing the structure of A and B when we have the appropriate group homomorphism $\chi : G \to \mathbb{T}$. There is another minor caveat: we can modify the the group homomorphism obtain from the pseudometric to ensure that the kernel is connected and the image is small; this was also the same arrangement we made earlier when we use the Gleason–Yamabe Theorem. The following example, already containing the idea of the later proof, will illustrate to the reader that unless there are closed intervals $I, J \subseteq \mathbb{T}$ such that A and B are nearly $\chi^{-1}(I)$ and $\chi^{-1}(J)$, the number $\mu_G(AB)$ is much larger than $\mu_G(A) + \mu_G(B)$. Assume that $G = \mathbb{T}^2$, $\chi : \mathbb{T}^2 \to \mathbb{T}$ is the projection onto the second coordinate, and we identify $\mathbb{T}^2$ with its fundamental domain $[0, 1] \times [0, 1] \subseteq \mathbb{R}^2$. Set

$$A = (\mathbb{T} \times [0, 1/7]) \cup ([0, 1/3] \times [1/7, 2/7]).$$

Then, $\mu_G(A) = 1/7 + (1/3) \times (1/7)$. On the other hand,

$$A^2 = (\mathbb{T} \times [0, 3/7]) \cup ([0, 2/3] \times [3/7, 4/7]).$$

So $\mu_G(A^2) = 3/7 + (2/3) \times (1/7) > 2/7 + 2/3 \times (2/7) = 2\mu_G(A)$.

6.2.2 The first step: Obtaining a suitable pseudometric

Suppose G is a *connected* and *compact* Lie group, and (A, B) is a nearly minimally expanding pair on G with sufficiently small measure. We will show that there is a one-dimensional torus subgroup T of G such that the "length" $\mu_T(g^{-1}A \cap H)$ of each "left fiber" $A \cap gT$ of A is small, and a similar condition hold for "right fibers" of B. After that, we will show that g_1A, g_2A, g_3A satisfy the picture at the bottom left of Figure 1 with respect to T, from which we can deduce the local almost linearity of d.

Suppose A is a *Kakeya set*, i.e, it has a long left fiber $A \cap gT$ for every choice of "direction" T of G. Fix such a T. If a large proportion of right fibers of B are rather short, then using the Kemperman inequality for T and Fubini's theorem, we get $\mu_G((A \cap gT)B)$ is already much larger than $\mu_G(A) + \mu_G(B)$, a contradiction. So a large proportion of the right fibers of B in the direction T must be rather long. A reverse argument then shows that a large proportion of the left fibers of A are long. Thus, $\mu_G(AT)$ is not too large compared to $\mu_G(A)$.

From the discussion above, it suffices to show the contradiction when $\mu_G(AT)/\mu_G(A)$ is not too large for every T. We call a nonempty and compact subset of G a **toric K-nonexpander**, if it has this property for a given constant K. We will show in Section 6.8 a result with independent interest: Every nonempty compact subset of G with sufficiently small measure cannot be a toric K-nonexpander.

Let us present here a pseudo-argument, which nevertheless illustrate the idea. Assume A is a toric K-nonexpander. Obtain finitely many torus subgroups $T_1, \dots, T_n$ of G such that

$$G = T_1 \cdots T_n.$$

(Note that n depends on the dimension of G, which is bounded by the caveat in Step 1.) Let us pretend that using the assumption $\mu_G(AT_1) \leq K\mu_G(A)$ we can cover AT_1 with $(K + 1)$ right translations of A. It can be then shown that AT_1 is a toric $K(K + 1)$-nonexpander. Next, we further pretend that AT_1T_2 can be covered with $K(K + 1) + 1$ right translations of

AT_1 which can then be covered by $K(K+1)^2+(K+1)$ right translations of A. Continuing the procedure, we get $C(K)$ such that $AT_1\cdots T_n = G$ can be covered by $C(K)$ right translations of A. Thus, $\mu_G(A) > 1/C(K)$, contradicting the assumption that $\mu_G(A)$ is very small.

The pseudo-argument in the preceding paragraph does not work in most of the cases. In particular, one cannot deduce from $\mu_G(AT_1) < K\mu_G(A)$ that AT_1 can be covered by $(K+1)$ right translations of A. However, it does contain some truth, and we will be able to use a version of the Lovazs covering argument to approximate this pseudo-argument.

Now choosing a one-dimensional torus subgroup T of G such that for all $x \in G$ and $y \in G$, the fibers $xT \cap A$ and $B \cap Ty$ are both short. We will show that the set A and B have the shape as described in Figure 1. Without loss of generality we can arrange that the width $\mu_G(AT)$ of A in G/T is at most the width $\mu_G(TB)$ of B in $T\backslash G$. Choose uniformly at random $xT \in AT$, and applying the Kemperman inequality for T, we have

$$\begin{aligned}
\mu_G(AB) &\geq \mathbb{E}_{xT\in AT}\mu_G((A\cap xT)B)\\
&\geq \mathbb{E}_{xT\in AT}\mu_T(A\cap xT)\mu_{T\backslash G}(TB) + \mu_G(B)\\
&= \mu_G(A)\frac{\mu_{T\backslash G}(TB)}{\mu_{G/T}(AT)} + \mu_G(B)\\
&\geq \mu_G(A) + \mu_G(B);
\end{aligned}$$

As (A, B) is nearly minimally expanding, we have $\mu_G(AB)$ is nearly the same as $\mu_G(A) + \mu_G(B)$. The fourth line then gives us that $\mu_G(TB)$ is nearly the same as $\mu_G(AT)$. The second line now gives us that for each $xT \in AT$, the fiber $(xT \cap A)$ is nearly an interval up to an endomorphism of T. From the first line, $\mu_T(A \cap xT)$ is almost constant as xT ranges through AT.

We now discuss the relative position of g_1A, g_2A, and g_3A for g_1, g_2, and g_3 near id_G. Clearly, (g_1A, B), (g_2A, B), and (g_3A, B) are also nearly minimally expanding. Using a submodularity argument as in [162], we can show that $(g_1A \cap g_2A, B)$, $(g_1A \cup g_2A, B)$ are also nearly minimally expanding. A similar analysis applies to all these pairs. In particular, in order to have almost all vertical fiber of each pair to having the same height, we can only have one of the following two scenarios:

1. (Almost vertical movement) AT is close to g_1AT

2. (Almost horizontal movement) fibers in $(A \cup g_1A)T$ has almost the same length as those in AT

We note that (2) cannot happen because then $A \cup g_1 A$ will no longer have a similar width as B. A similar argument shows us that Figure 2 cannot happen. Hence, the "moving up" and "moving down" divisions in g_2A and g_3A must be the same as in Figure 1 in Section 6.2.1.

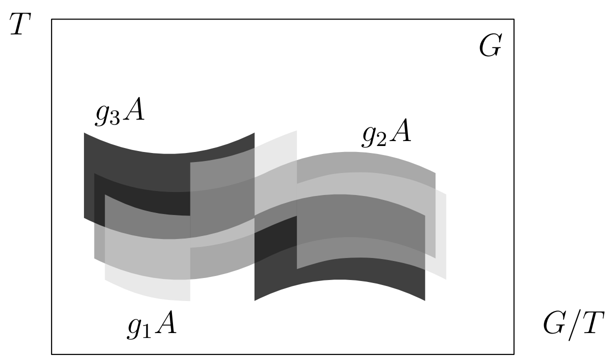


Figure 6.2: Different "moving up" and "moving down" divisions in g_2A and g_3A result in uneven fiber size.

We mention two subtler aspects of the geometry of minimally expanding pairs that we will not be able to get in details. First, we will also need it to show a condition called *path monotonicity* which is a necessary ingredient to show assumption (2) in Step 2 (Section 6.2.3)). Second, for the purpose of controlling the error in Step 2, we need to show a certain "convexity property" of

$$N_\lambda = \{g \in G : \mu_G(A) - \mu_G(A \cap gA) < \lambda\}.$$

This requires us to construct a "core" of A, which is related to the Sanders–Croot–Sisask theorem [151, 47] in additive combinatorics, and stabilizer theorems in model theory [102].

6.2.3 The second step: Constructing an appropriate homomorphism

We assume in this section that G is a *connected* and *compact* Lie group, d is a left invariant continuous pseudometric on G with the following properties:

1. (Local almost linearity) There is $\lambda \in \mathbb{R}^{>0}$ such that with

   $$N_\lambda := \{g \in G : d(\mathrm{id}_G, g) < \lambda\},$$

   there is $\varepsilon < 10^{-10}\mu_G(N_\lambda)$ such that for all $g_1, g_2, g_3 \in N_\lambda$

   $$d(g_1, g_3) \in |d(g_1, g_2) \pm d(g_2, g_3)| + I(\varepsilon),$$

   where $I(\varepsilon)$ is the interval $(-\varepsilon, \varepsilon) \subseteq \mathbb{R}$.

2. (Local monotonicity) With the same λ in (2), for all $g \in N_\lambda$,

   $$|d(\mathrm{id}_G, g^2) - 2d(\mathrm{id}_G, g)| \in I(\varepsilon).$$

We now sketch how to construct a continuous and surjective group homomorphism to $\mathbb{T}$ from these data. The condition presented here is a simplified but essentially equivalent condition as in Section 6.7. A crucial argument we will not be able to get into details here is to show that the local monotonicity condition can be deduced from a weaker property of path monotonicity obtained from the first step (Section 6.2.2).

When we are in the special case with $\varepsilon = 0$ in property (1), there is a relatively easy argument which also works for noncompact Lie groups. Set

$$\ker d = \{g \in G : d(\mathrm{id}_G, g) = 0\}.$$

Using the left invariance, continuity, and triangle inequality, one can show that $\ker d$ is a closed subgroup of G. Moreover, in this case, $G/\ker d$ must be isomorphic to $\mathbb{T}$, and the pseudometric d locally must agrees with a constant multiple of the pullback of the Euclidean metric. These are, perhaps, not too surprising as a Lie group equipped with a locally linear pseudometric is, intuitively, a very rigid object which locally looks like a straight line. In fact, property (2) is not needed as it is a consequence of property (1) in this case.

The general case is much harder as we no longer have the same type of rigidity. In particular, $\ker d$ might not be normal, and $G/\ker d$ might not be $\mathbb{T}$ even if $\ker d$ is normal. The reader familiar with the proof of Hilbert's Fifth problem would guess that we might try to slightly modify d to get a locally linear pseudometric d' and use the earlier strategy. This is still true at the conceptual level, but our actual argument is much more explicit allowing error control.

As an expository vehicle for the idea, let us still imagine that we somehow obtain a locally linear pseudometric d' only slightly differs from d. By the earlier argument, we get a group homomorphism $\chi' : G \to \mathbb{T}$ and $\alpha' \in \mathbb{R}$ such that $d'(g_1, g_2) = \alpha' \|\chi'(g_1) - \chi'(g_2)\|_{\mathbb{T}}$ near id_G. Therefore, for g near id_G we have

$$\chi'(g) = \pm(1/\alpha')d'(\mathrm{id}_G, g) + \mathbb{Z}.$$

Note that this gives us a way to determine $\chi'(g)$ for $g \in N_\lambda$ from d', but we need to know how to describe α' from d'. For $g_1, g_2, g_3 \in N_\lambda$, we say that g_2 is "between" g_1 and g_3 if $d'(g_1, g_3) = d'(g_1, g_2) + d'(g_2, g_3)$. We say that a product $g_1 \cdots g_n$ is "increasing" if $g_1 \cdots g_i$ is "between" $g_1 \cdots g_{i-1}$ and $g_1 \cdots g_{i+i}$. For an arbitrary element $g \in G$, we write it as an "increasing" product $g = g_1 \cdots g_n$ with $g_i \in N_\lambda$. Then, we can show that

$$\chi'(g) = \pm(1/\alpha') \sum_{i=0}^{n-1} d'(g_1 \ldots g_i, g_1 \ldots g_{i+1}) + \mathbb{Z}.$$

This gives us a way to determine $\chi'(g)$ for an arbitrary g if we can get α' from d'. Now, note that if we write id_G as an "increasing" product $1_G = g_1 \cdots g_n$ with $n > 0$ then $\sum_{i=0}^{n-1} d'(g_1 \dots g_i, g_1 \dots g_{i+1})$ will be a multiple of α'. Moreover, α' will be the minimum value of such sum. So we recover α' from d' and obtain a much more explicit way to describe the map χ'.

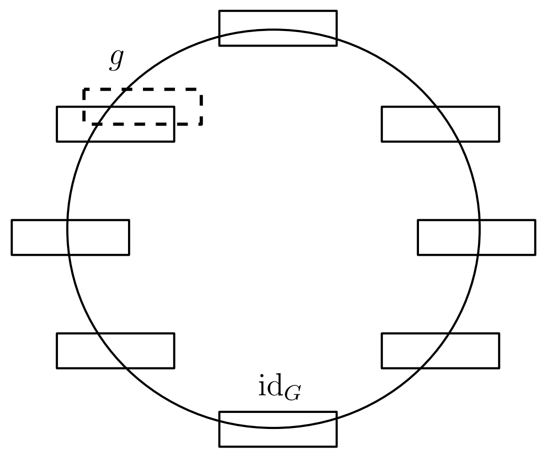


Figure 6.3: Any element g in G will be captured by one of the monitors.

Back to the situation where we only have d but not d', we can carry out the same strategy as above with d while being careful with the errors. There are a number of problems that arise. To define the "betweeness", property (2) plays an important technical role. Since the error propagate very fast, to get a linear error bound, we cannot write g as an "increasing" product $g = g_1 \cdots g_n$ with very large n. The upper bound in n comes from a lower bound on the size of N_λ, which essentially comes from the result on "core" in the first step. To get an analog of α' with the desired property, we need to choose a standard way of expressing id_G as an "increasing" product $\mathrm{id}_G = g_1 \cdots g_n$ and use $g_1, \dots, g_n$ to "monitor" the other elements in the group, as illustrated by Figure 6.3. This require us to develop the machinery of irreducible sequence and concatenation; see Section 6.7.3 for details. Finally, what we ends up with is a multi-valued almost homomorphism that is not continuous but still universally measurable. We will need to use a number of results from descriptive set theory and Riemannian geometry to extract from this the desired group homomorphism.

6.2.4 The third step: Deducing the structure of the pair

Suppose (A, B) is a nearly minimally expanding pair on G, H is a *connected*, *compact* and *normal* subgroup of G, $\pi : G \to G/H$ is the quotient map, and

$$\mu_{G/H}(\pi(A)) + \mu_{G/H}(\pi(B) < 1.$$

The goal of this step is to show the following *transfer to quotients* result: There is a nearly minimally expanding pair (A', B') on G/H such that $\mu_G(A \triangle \pi^{-1}(A'))$ and $\mu_G(B \triangle \pi^{-1}(B'))$ are both small.

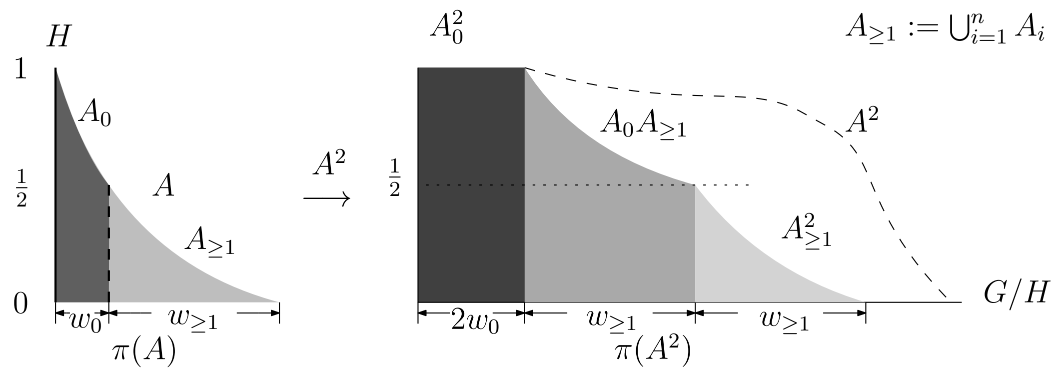


Figure 6.4: Lower bound for $\mu_G(A^2)$.

To illustrate the idea, we focus on the special case with $A = B$. Employing the geometric language in Section 6.2.1, we call $\mu_{G/H}(\pi(A))$ the *width* of A, for each g in G, we call $A \cap gH$ a *fiber* of A, and refer to $\mu_H(g^{-1}A \cap H)$ as its *length*. We consider a further special case assuming that A can be partitioned into $N + 1$ parts $A = \bigcup_{i=0}^{N} A_i$ such that the images under π of the A_i's are compact and pairwise disjoint, A_i has width w_i, the fibers in A_0 all have length $\geq 1/2$, the fibers in A_i all have the same length $l_i \leq 1/2$ for each $i \geq 1$, and $l_i \geq l_{i+1}$ for all $i < N$. This further special case is, in fact, quite representative as we can reduce the general problem to it using approximation techniques.

The proof of this step can be seen as the following "spillover" argument. Applying the Kemperman inequalities for H and G/H, we learn that all the fibers in A_0^2 has length 1, and the width of A_0^2 is at least $2w_0$. By Fubini's theorem, $\mu_G(A_0^2)$ is at least $2w_0$. Next, consider $A_0(A_0 \cup A_1)$. A similar argument gives us that all the fibers in $A_0(A_0 \cup A_1)$ has length at least $l_1 + 1/2$, and the width of $A_0(A_0 \cup A_1)$ in G/H is at least $2w_0 + w_1$. Note that $(l_1 + 1/2)(2w_0 + w_1)$ is a weak bound for $\mu_G(A_0(A_0 \cup A_1))$ since the fibers in A_0^2 are "exceptionally long". Taking all of these into account, a stronger lower bound for $\mu_G(A_0(A_0 \cup A_1))$ is

$$2w_0 + (l_1 + 1/2)w_1.$$

Iterate this procedure, a lower bound for $\mu_G(A_0 A) = \mu_G(A_0(A_0 \cup \ldots \cup A_N))$ is

$$2w_0 + \sum_{i=1}^{N}(l_i + 1/2)w_i.$$

Now every fiber in $(A_0 \cup A_1)A$ has length at least l_1, and the width of $(A_0 \cup A_1)A$ is at least $(w_0 + w_1) + w$. Using the same logic, a lower bound for $\mu_G((A_0 \cup A_1)A)$ is

$$2w_0 + \sum_{i=1}^{N}(l_i + 1/2)w_i + l_1 w_1.$$

Iterate the procedure, a lower bound for $\mu_G(A^2) = \mu_G((A_0 \cup \ldots \cup A_n)A)$ is

$$2w_0 + \sum_{i=1}^{N}(l_i + 1/2)w_i + \sum_{i=1}^{N} l_i w_i.$$

Note that $\mu_G(A) = l_0 w_0 + \ldots + l_N w_N$. Hence, $\mu_G(A^2)$ is nearly $2\mu_G(A)$ implies that we must nearly have $w_0 = 1$, $w_1 = \ldots = w_N = 0$, and $l_0 = 1$. From this, one can deduce the conclusion that we want for this step.

6.2.5 Structure of the chapter

The chapter is organized as follows. Section 6.3 includes some facts about Haar measures and unimodular groups, which will be used in the subsequent part of the chapter. Section 6.4

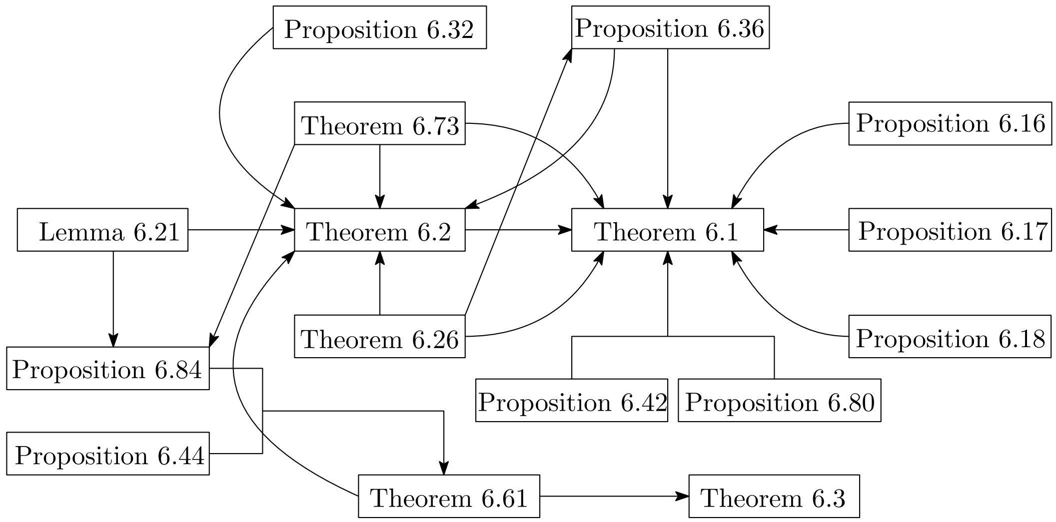
deals with the more immediate parts of Theorem 6.1 and hence sets up the stage for the main part of the argument. Section 6.5 allows us to arrange that in a minimally or a nearly minimally expanding pair (A, B), the sets A and B have small measure (Lemma 6.21). Sections 6.6, 6.7, and 6.8 contain main new technical ingredients of the proof, which will be put together in Section 6.9 to complete the proofs of Theorem 6.1, Theorem 6.2, and Theorem 6.3. Steps 1, 2, and 3 discussed in Sections 6.2.2, 6.2.3, and 6.2.4, corresponds to Sections 6.8, 6.7, and 6.6 respectively.

In Section 6.6.1, we proved the quotient domination theorem (Theorem 6.26), which allow us to transfer the problem into certain quotient groups. Section 6.6.2 gives an upper bound on the dimension of the Lie model (Proposition 6.32). Section 6.6.3 contains structural results assuming we have an appropriate homomorphism (Proposition 6.36). Together with the transfer theorem, we reduce the problem to a bounded dimension Lie group.

In Section 6.7.1, we showed that a locally linear pseudometric on G would induce a continuous surjective homomorphism to either $\mathbb{R}$ or $\mathbb{T}$, with compact kernel (Proposition 6.42). Sections 6.7.2 and 6.7.3 study the locally almost linear pseudometric in compact Lie groups. In particular, we proved that path monotonicity implies monotonicity (Proposition 6.44), and for almost monotone almost linear pseudometric, one can also find a homomorphism mapping to $\mathbb{T}$ (Theorem 6.61).

In Section 6.8.1, we bound the size of the toric expanders (Theorem 6.73). We construct the pseudometric from geometric properties of nearly minimal expansion sets in Sections 6.8.2 and 6.8.3. Section 6.8.2 provides a locally linear pseudometric from minimally expansion sets (Proposition 6.80). In Section 6.8.3, we construct a path monotone locally almost linear pseudometric (Proposition 6.84).

The dependency diagram of the chapter is as below.



6.3 Preliminaries

Throughout this section, we assume that G is a *connected* locally compact group (in particular, Hausdorff) equipped with a left Haar measure μ_G, and $A, B \subseteq G$ are nonempty.

6.3.1 Locally compact groups and Haar measures

Below are some basic facts about μ_G that we will use; see [50, Chapter 1] for details:

Fact 6.4. *Suppose μ_G is either a left or a right Haar measure on G. Then:*

(i) *If A is compact, then A is μ_G-measurable and $\mu_G(A) < \infty$.*

(ii) *If A is open, then A is μ_G-measurable and $\mu_G(A) > 0$.*

(iii) (Outer regularity) *If A is μ_G-measurable, then there is a decreasing sequence (U_n) of open subsets of G with $A \subseteq U_n$ for all n, and $\mu_G(A) = \lim_{n\to\infty} \mu_G(U_n)$.*

(iv) (Inner regularity) *If A is μ_G-measurable, then there is an increasing sequence (K_n) of compact subsets of A such that $\mu_G(A) = \lim_{n\to\infty} \mu_G(K_n)$.*

(v) (Measurability characterization) *If there is an increasing sequence (K_n) of compact subsets of A, and a decreasing sequence (U_n) of open subsets of G with $A \subseteq U_n$ for all n such that $\lim_{n\to\infty} \mu_G(K_n) = \lim_{n\to\infty} \mu_G(U_n)$, then A is measurable.*

(vi) (Uniqueness) *If μ'_G is another measure on G satisfying the properties (1-5), then there is $C \in \mathbb{R}^{>0}$ such that $\mu'_G = C\mu_G$.*

(vii) (Continuity of measure under symmetric difference) *Suppose $A \subseteq G$ is measurable, then the function $G \to \mathbb{R}, g \mapsto \mu_G(A \triangle gA)$ is continuous.*

We remark that the assumption that G is connected implies that every measurable set is σ-finite (i.e., countable union of sets with finite μ_G-measure). Without the connected assumption, we only have inner regularity for σ-finite sets. From Fact 6.4(vii), we get the following easy corollary:

Corollary 6.5. *Suppose A is μ_G-measurable and ε is a constant. Then $\mathrm{Stab}^{\varepsilon}_G(A)$ is closed in G, while $\mathrm{Stab}^{<\varepsilon}_G(A)$ is open in G. In particular, $\mathrm{Stab}^{0}_G(A)$ is a closed subgroup of G.*

We say that G is **unimodular** if μ_G (and hence every left Haar measure on G) is also a right Haar measure. The following is well known and can be easily verified:

Fact 6.6. *If G is unimodular, A is μ_G-measurable, then A^{-1} is also μ_G-measurable and $\mu_G(A) = \mu_G(A^{-1})$.*

We use the following isomorphism theorem of topological groups.

Fact 6.7. *Suppose G is a locally compact group, H is a closed normal subgroup of G. Then we have the following.*

(i) (First isomorphism theorem) *Suppose $\phi : G \to Q$ is a continuous surjective group homomorphism with $\ker \phi = H$. Then the exact sequence of groups*

$$1 \to H \to G \to Q \to 1$$

is an exact sequence of topological groups if and only if ϕ is open; the former condition is equivalent to saying that Q is canonically isomorphic to G/H as topological groups.

(ii) (Third isomorphism theorem) *Suppose $S \leq G$ is closed, and $H \leq S$. Then S/H is a closed subgroup of G/H. If $S \lhd G$ is normal, then S/H is a normal subgroup of G/H, and we have the exact sequence of topological groups*

$$1 \to S/H \to G/H \to G/S \to 1;$$

this is the same as saying that $(G/H)/(S/H)$ is canonically isomorphic to G/S as topological groups.

Suppose H is a closed subgroup of G. The following fact allows us to link Haar measures on G with the Haar measures on H for unimodular G and H:

Fact 6.8 (Quotient integral formula)**.** *Suppose H is a closed subgroup of G with a left Haar measure μ_H. If f is a continuous function on G with compact support, then*

$$xH \mapsto \int_H f(xh)\,\mathrm{d}\mu_H(x).$$

defines a function $f^H : G/H \to \mathbb{R}$ which is continuous and has compact support. If both G and H are unimodular, then there is unique invariant Radon measures $\mu_{G/H}$ on G/H such that for all continuous function $f : G \to \mathbb{R}$ with compact support, the following integral formula holds

$$\int_G f(x)\,\mathrm{d}\mu_G(x) = \int_{G/H}\int_H f(xh)\,\mathrm{d}\mu_H(h)\,\mathrm{d}\mu_{G/H}(xH).$$

A similar statement applies replacing the left homogeneous space G/H with the right homogeneous space $H\backslash G$.

We can extend Fact 6.8 to measurable functions on G, but the function f^H in the statement can be only be defined and is $\mu_{G/H}$-measurable μ_G-almost everywhere. So, in particular, this problem applies to indicator function $\mathbb{1}_A$ of a measurable set A. This causes problem in our later proof and prompts us to sometimes restrict our attention to a better behaved subcollection of measurable subsets of G. We say that a subset of G is **σ-compact** if it is a countable union of compact subsets of G.

Lemma 6.9. *We have the following:*

(i) *σ-compact sets are measurable.*

(ii) *the collection of σ-compact sets is closed under taking countable union, taking finite intersection, and taking product set.*

(iii) *For all μ_G-measurable A, we can find a σ-compact subset A' of A such that $\mu_G(A') = \mu_G(A)$.*

(iv) *Suppose G is unimodular, H is a closed subgroup of G with a left Haar measure μ_H, $A \subseteq G$ is σ-compact, and $\mathbb{1}_A$ is the indicator function of A. Then $aH \mapsto \mu_H(A \cap aH)$ defines a measurable function $\mathbb{1}_A^H : G/H \to \mathbb{R}$. If H is unimodular and, $\mu_{G/H}$ is the Radon measure given in Fact 6.8, then*

$$\mu_G(A) = \int_{G/H} \int_H \mu_H(A \cap aH) \,\mathrm{d}\mu_H(h) \,\mathrm{d}\mu_{G/H}(xH).$$

A similar statement applies replacing the left homogeneous space G/H with the right homogeneous space $H\backslash G$.

Proof. The verification of (i-iii) is straightforward. We now prove (iv). First consider the case where A is compact. By Baire's Theorem, $\mathbb{1}_A$ is the pointwise limit of a monotone nondecreasing sequence of continuous function of compact support. If $f : G \to \mathbb{R}$ is a continuous function of compact support, then the function

$$f^H : G/H \to R, aH \mapsto \int_H f(ax)dx$$

is continuous with compact support, and hence measurable; see [50, Lemma 1.5.1]. Noting that $\mu_H(A \cap aH) = \int_H \mathbb{1}_A(ax)dx$, and applying monotone convergence theorem, we get that

$\mathbb{1}_A^H$ is the pointwise limit of a monotone nondecreasing sequence of continuous function of compact support. Using monotone convergence theorem again, we get $\mathbb{1}_A^H$ is integrable, and hence measurable. Also by monotone convergence theorem, we get the quotient integral formula in the statement.

Finally, the general case where A is only σ-compact can be handled similarly, noting that $\mathbb{1}_A$ is then the pointwise limit of a monotone nondecreasing sequence of indicator functions of compact sets. $\square$

Suppose H is a closed subgroup of G. Then H is locally compact, but not necessarily unimodular. We use the following fact in order to apply induction arguments in the later proofs.

Fact 6.10. *Let G be a unimodular group. If H is a closed normal subgroup of G, then H is unimodular. Moreover, if H is compact, then G/H is unimodular.*

Given A, B subsets of some unimodular group G, each with finite positive measure, and suppose AB^{-1} is measurable. **Ruzsa's distance** is defined by

$$d(A,B) = \log \frac{\mu_G(AB^{-1})}{\mu_G(A)^{1/2}\mu_G(B)^{1/2}}.$$

The following fact is known as triangle inequality of Ruzsa's distance.

Fact 6.11. *Let C be a set of finite positive measure, and suppose AC^{-1}, CB^{-1} are measurable sets. Then $d(A,B) \leq d(A,C) + d(C,B)$.*

6.3.2 More on Kemperman's inequality and the inverse problem

We will need a version of Kemperman's inequality for arbitary sets. Recall that the *inner Haar measure* $\widetilde{\mu}_G$ associated to μ_G is given by

$$\widetilde{\mu}_G(A) = \sup\{\mu_G(K) : K \subseteq A \text{ is compact.}\}$$

The following is well known and can be easily verified:

Fact 6.12. *Suppose $\widetilde{\mu}_G$ is the inner Haar measure associated to μ_G. Then we have the following:*

(i) (Agreement with μ_G) *If A is measurable, then $\widetilde{\mu}_G(A) = \mu(A)$.*

(ii) (Inner regularity) *There is σ-compact $A' \subseteq A$ such that*

$$\widetilde{\mu}_G(A) = \widetilde{\mu}_G(A') = \mu_G(A).$$

(iii) (Superadditivity) *If A and B are disjoint, then*

$$\widetilde{\mu}_G(A \cup B) \geq \widetilde{\mu}_G(A) + \widetilde{\mu}_G(B).$$

(iv) (Left invariance) *For all $g \in G$, $\widetilde{\mu}_G(gA) = \widetilde{\mu}(A)$.*

(v) (Right invariance) *If G is unimodular, then for all $g \in G$, $\widetilde{\mu}_G(Ag) = \widetilde{\mu}(A)$.*

It is easy to see that we can replace the assumption that A and B are compact in Kemperman's inequality in the introduction with the weaker assumption that A and B are σ-compact. Together with the inner regularity of $\widetilde{\mu}_G$ (Fact 6.12.2), this give us the first part of the following Fact 6.13. The second part of Fact 6.13 follows from the fact that taking product sets preserves compactness, σ-compactness, and analyticity. Note that taking product sets in general does not preserve measurability, so we still need inner measure in this case.

Fact 6.13 (Generalized Kemperman inequality for connected groups)**.** *Suppose $\tilde{\mu}_G$ is the inner Haar measure on G, and $A, B \subseteq G$ are nonempty. Then*

$$\tilde{\mu}_G(AB) \geq \min\{\widetilde{\mu}_G(A) + \widetilde{\mu}_G(B), \widetilde{\mu}_G(G)\}.$$

Moreover, if A and B are compact, σ-compact, or analytic, then we can replace $\tilde{\mu}_G$ with μ_G.

The remaining parts of Theorem 6.1 consist of classifying the minimally expanding pairs (A, B) and show that they match the description in situations (iii) and (iv) of Theorem 6.1. For compact group, our strategy is to reduce the problem to the known situations of one dimensional tori. Hence, we need the following special case of Kneser's classification result, and the sharp dependence between ε and δ is essentially due to Bilu [16].

Fact 6.14 (Inverse theorem for $\mathbb{T}^d$)**.** *Let A, B be compact subsets of $\mathbb{T}^d$. For every $\tau > 0$, there is a constant $c = c(\tau)$ such that if*

$$\tau^{-1}\mu_{\mathbb{T}^d}(A) \leq \mu_{\mathbb{T}^d}(B) \leq \mu_{\mathbb{T}^d}(A) \leq c,$$

then either $\mu_{\mathbb{T}^d}(A + B) \geq \mu_{\mathbb{T}^d}(A) + 2\mu_{\mathbb{T}^d}(B)$, or there are compact intervals I, J in $\mathbb{T}$ with $\mu_{\mathbb{T}}(I) = \mu_{\mathbb{T}^d}(A+B) - \mu_{\mathbb{T}^d}(B)$ and $\mu_{\mathbb{T}}(J) = \mu_{\mathbb{T}^d}(A+B) - \mu_{\mathbb{T}^d}(A)$, and a continuous surjective group homomorphism $\chi : \mathbb{T}^d \to \mathbb{T}$, such that $A \subseteq \chi^{-1}(I)$ and $B \subseteq \chi^{-1}(J)$.

For noncompact group, we reduce the problem to to the known situation of additive group of real numbers. The following result can be seen as the stability theorem of the Brunn–Minkowski inequality in $\mathbb{R}^d$ when $d = 1$.

Fact 6.15 (Inverse theorem for $\mathbb{R}$)**.** *Let A, B be compact subsets in $\mathbb{R}$ with $\mu_G(A) \geq \mu_G(B)$, and let $\mu_{\mathbb{R}}$ be the Lebesgue measure in $\mathbb{R}$. Suppose we have*

$$\mu_{\mathbb{R}}(A + B) < \mu_{\mathbb{R}}(A) + 2\mu_{\mathbb{R}}(B).$$

Then there are compact intervals $I, J \subseteq \mathbb{R}$ with $\mu_{\mathbb{R}}(I) = \mu_{\mathbb{R}}(A + B) - \mu_{\mathbb{R}}(B)$ and $\mu_{\mathbb{R}}(J) = \mu_{\mathbb{R}}(A + B) - \mu_{\mathbb{R}}(A)$, such that $A \subseteq I$ and $B \subseteq J$.

6.4 Reduction to nearly minimal expansion pairs

To set the stage for the later discussion, we would like to separate the core part of Theorem 6.1 from the more immediate parts. Throughout G is a connected unimodular group, μ_G is a Haar measure on G, and A and B are nonempty compact subsets of G.

Proposition 6.16. *Suppose one of the situation listed in Theorem 6.1 holds, then* $\mu_G(AB) = \min(\mu_G(A) + \mu_G(B), \mu_G(G))$.

Proof. We will only consider situation (iii) because (i) and (ii) are immediate and (iv) can be showed in a same way as (iii). Suppose we are in situation (iii) of Theorem 6.1. As χ is a group homomorphism, we have $AB = \chi^{-1}(I + J)$. Note that by quotient integral formula, we have $\mu_G(A) = \mu_{\mathbb{T}}(I)$, $\mu_G(B) = \mu_{\mathbb{T}}(J)$, $\mu_G(AB) = \mu_{\mathbb{T}}(I + J)$. The desired conclusion follows from the easy that $\mu_{\mathbb{T}}(I + J) = \mu_{\mathbb{T}}(I) + \mu_{\mathbb{T}}(J)$. $\square$

The following lemma clarifies the second statement in situation (ii) of Theorem 6.1.

Proposition 6.17. *Suppose* $\mu_G(A) + \mu_G(B) \geq \mu_G(G)$. *Then* $AB = G$.

Proof. Suppose g is an arbitrary element of G. It suffices to show that $A^{-1}g$ and B has nonempty intersection. As G is unimodular, $\mu_G(A) = \mu_G(A^{-1})$ by Fact 6.6. Hence $\mu_G(A^{-1}g) + \mu_G(B) = \mu_G(G)$. If $\mu_G(A^{-1}g \cap B) > 0$, then we are done. Otherwise, we have have $\mu_G(A^{-1}g \cap B) = 0$, and so $\mu_G(A^{-1}g \cup B) = \mu_G(G)$ by inclusion exclusion principle. As A and B are compact, $A^{-1}g \cup B$ is also compact, and the complement of $A^{-1}g \cup B$ is open. Since nonempty open sets has positive measure, $\mu_G(A^{-1}g \cup B) = \mu_G(G)$ implies $A^{-1}g \cup B = G$. Now, since G is connected, we must have $A^{-1}g \cap B$ must be nonempty. $\square$

Now we clarify the situation in (iii) of Theorem 6.1, situation (iii) can be proved in the same way.

Proposition 6.18. *Let G be a connected unimodular group, and A, B be nonempty compact subsets of G. Suppose $\mu_G(A) = 0$, $0 < \mu_G(B) < \mu_G(G)$, and $\mu_G(AB) = \min(\mu_G(A) + \mu_G(B), \mu_G(G))$. Then there is a compact subgroup H of G such that $A \subseteq H$, and $B = HB$.*

Proof. Without loss of generality, we can assume that A and B both contain id_G. Let H be the smallest closed subgroup containing A. It suffices to show that $HB = B$. Indeed, H is then a closed subset of B which implies that H is compact.

From Corollary 6.5, $\mathrm{Stab}_0(B) = \{g \in G \mid \mu_G(B \triangle gB) = 0\}$ is a closed subgroup of G. As $\mu_G(AB) = \mu_G(B)$ and id_G is in A, we must have $A \subseteq \mathrm{Stab}_0(B)$. By the assumption that H is the smallest closed subgroup containing A, one must have $H \leq \mathrm{Stab}_0(B)$.

We first consider the special case where $\mu_G(U \cap B) > 0$ for every $b \in B$ and open neighborhood U of b. As μ_G is both left and right invariant, this assumption also implies that $\mu_G(U \cap gB) > 0$ for all $g \in G$, $b \in gB$, and open neighborhood U of b. Suppose b is in $HB \setminus B$. Since $HB = \bigcup_{h \in H} hB$, we obtain $h \in H$ such that b is in hB. Set $U = G \setminus B$. Then U is an open neighborhood of b. From the earlier discussion, we then have $\mu_G(U \cap hB) > 0$. As hB is a subset of HB, it implies that $\mu_G(HB \setminus B) > 0$ which is a contradiction.

It remains to reduce the general situation to the above special case. Set

$$B_0 = \{b \in B : \text{ There is an open neighborhood } U_b \text{ of } b \text{ with } \mu_G(U_b \cap B) = 0\}.$$

If b is in B_0, and $b' \in U_b \cap B$, U_b also witnesses that b' is in B_0. Hence, $B \setminus B_0$ is a closed subset of B, which implies that $B \setminus B_0$ is compact. Now we show that $\mu_G(B_0) = 0$. Suppose B' is a compact subset of B_0. Then from the definition of B_0, we can obtain an open covering $(U_i)_{i \in I}$ of B' such that $\mu_G(U_i \cap B) = 0$ for all $i \in I$. As B' is compact, we get from $(U_i)_{i \in I}$ a finite subcovering of B'. Hence, $\mu_G(B') = 0$. By inner regularity of Haar measure, $\mu_G(B_0) = 0$. Replacing B with $B \setminus B_0$, we reduce the situation to the above special case. □

6.5 Reduction to sets with small measure

Throughout this section, G is a connected *compact* group, μ_G is the normalized Haar measure on G, and $A, B \subseteq G$ are σ-compact sets with positive measure. We will show that if (A, B)

is nearly minimally expanding in G, then we can σ-compact A' and B' each with smaller measure such that the pair (A', B') is also nearly minimally expanding. The similar approach used in this section is introduced by Tao [162] and used to obtain an inverse theorem in the abelian setting. We first prove the following easy fact, which will be used several times later in the paper.

Let $f, g : G \to \mathbb{C}$ be functions. For every $x \in G$, we define the **convolution** of f and g to be

$$f * g(x) = \int_G f(y) g(y^{-1}x) \,\mathrm{d}\mu_G(y).$$

Note that $f * g$ is not commutative, but associative by Fubini's Theorem.

Lemma 6.19. *Let t be any real numbers such that $\mu_G(A)^2 \leq t \leq \mu_G(A)$. Then there are $x, y \in G$ such that $\mu_G(A \cap (xA)) = \mu_G(A \cap (Ay)) = t$.*

Proof. Consider the maps:

$$\pi_1 : x \mapsto \mathbb{1}_A * \mathbb{1}_{A^{-1}}(x) = \mu_G(A \cap (xA)), \text{ and } \pi_2 : y \mapsto \mathbb{1}_{A^{-1}} * \mathbb{1}_A(y) = \mu_G(A \cap (Ay)).$$

By Fact 6.4, both π_1 and π_2 are continuous functions, and equals to $\mu_G(A)$ when $x = y = \mathrm{id}_G$. By Fubini's theorem

$$\mathbb{E}\,(\mathbb{1}_A * \mathbb{1}_{A^{-1}}) = \mu_G(A)^2 = \mathbb{E}\,(\mathbb{1}_{A^{-1}} * \mathbb{1}_A).$$

Then the lemma follows from the intermediate value theorem, and the fact that G is connected. □

Recall that $\mathfrak{d}_G(A, B) = \mu_G(AB) - \mu_G(A) - \mu_G(B)$ is the discrepancy of A and B on G. The following property is sometimes refered to as **submodularity** in the literature. Note that this is not related to modular functions in locally compact groups or the notion of modularity in model theory.

Lemma 6.20. *Let $\gamma_1, \gamma_2 > 0$, and A, B_1, B_2 are σ-compact subsets of G. Suppose that $\mathfrak{d}_G(A, B_1) \leq \gamma_1$, $\mathfrak{d}_G(A, B_2) \leq \gamma_2$, and*

$$\mu_G(B_1 \cap B_2) > 0, \qquad \textit{and} \;\; \mu_G(A) + \mu_G(B_1 \cup B_2) \leq 1.$$

Then both $\mathfrak{d}_G(A, B_1 \cap B_2)$ and $\mathfrak{d}_G(A, B_1 \cup B_2)$ are at most $\gamma_1 + \gamma_2$.

Proof. Observe that for every $x \in G$ we have

$$\mathbb{1}_{AB_1}(x) + \mathbb{1}_{AB_2}(x) \geq \mathbb{1}_{A(B_1 \cap B_2)}(x) + \mathbb{1}_{A(B_1 \cup B_2)}(x),$$

which implies

$$\mu_G(AB_1) + \mu_G(AB_2) \geq \mu_G(A(B_1 \cap B_2)) + \mu_G(A(B_1 \cup B_2)). \tag{6.2}$$

By the fact that $\mathfrak{d}_G(A, B_1) \leq \gamma_1$ and $\mathfrak{d}_G(A, B_2) \leq \gamma_2$, we obtain

$$\mu_G(AB_1) \leq \mu_G(A) + \mu_G(B_1) + \gamma_1, \text{ and } \mu_G(AB_2) \leq \mu_G(A) + \mu_G(B_2) + \gamma_2.$$

Therefore, by equation (6.2) we have

$$\begin{aligned} &\mu_G(A(B_1 \cap B_2)) + \mu_G(A(B_1 \cup B_2)) \\ \leq\, &2\mu_G(A) + \mu_G(B_1 \cap B_2) + \mu_G(B_1 \cup B_2) + \gamma_1 + \gamma_2. \end{aligned}$$

On the other hand, as $\mu_G(B_1 \cap B_2) > 0$ and $\mu_G(A) + \mu_G(B_1 \cup B_2) \leq 1$, and using Kemperman's inequality, we have

$$\mu_G(A(B_1 \cap B_2)) \geq \mu_G(A) + \mu_G(B_1 \cap B_2),$$

and

$$\mu_G(A(B_1 \cup B_2)) \geq \mu_G(A) + \mu_G(B_1 \cup B_2).$$

This implies

$$\mu_G(A(B_1 \cap B_2)) \le \mu_G(A) + \mu_G(B_1 \cap B_2) + \gamma_1 + \gamma_2,$$

and

$$\mu_G(A(B_1 \cup B_2)) \le \mu_G(A) + \mu_G(B_1 \cup B_2) + \gamma_1 + \gamma_2.$$

Thus we have $\mathfrak{d}_G(A, B_1 \cap B_2), \mathfrak{d}_G(A, B_1 \cup B_2) \le \gamma_1 + \gamma_2$. $\square$

The following lemma is the main result of this section, it says if G admits a small expansion pair, one can another find pair of sets with sufficiently small measures, and still has small expansion.

Lemma 6.21. *Let $d_1, d_2 \in (0, 1/4)$ be positive real numbers, and let*

$$m = \min\{\mu_G(A), \mu_G(B), 1 - \mu_G(A) - \mu_G(B)\}.$$

Suppose $\mathfrak{d}_G(A, B) \le \gamma$. Then there are σ-compact sets $A', B' \subseteq G$ satisfying

(i) *$\mu_G(A') = d_1$ and $\mu_G(B') = d_2$,*

(ii) *$\mathfrak{d}_G(A', B)$, $\mathfrak{d}_G(A, B')$, and $\mathfrak{d}_G(A', B')$ are at most $O_{d_1,d_2}(\gamma/m)$.*

Proof. Without loss of generality we assume $\mu_G(A) > d_1$ and $\mu_G(A) \ge \mu_G(B)$. The case when $\mu_G(A)$ is less than d_1 can be proved in a similar way by replacing taking intersections by taking unions. Observe that for every $g \in G$, both $\mathfrak{d}_G(gA, B)$ and $\mathfrak{d}_G(A, Bg)$ are still upper bounded by γ. By Lemma 6.19, for every t with $\mu_G(A)^2 \le t \le \mu_G(A)$, there is $g \in G$ such that $\mu_G(A \cap gA) = t$. Assuming that in each step, we can choose g such that $\mu_G(A \cap gA) = \mu_G(A)^2$, and replace A by $A \cap gA$. Hence after $O(\log\log 1/d_1)$ steps, the measure of A will achieve d_1. The issue of this simple argument is that we may have $\mu_G(A \cup gA) + \mu_G(B) > 1$ when $\mu_G(A) \ge 1/3$, so that we cannot apply Lemma 6.20. Thus in the first few steps, we will choose g such that $\mu_G(A \cup gA)$ is not too large.

We first consider the case when $\mu_G(A) \geq 1/3$, and $\mu_G(A) - m \geq \mu_G(A)^2$. We are going to choose $g \in G$ such that

$$2\mu_G(A) - \mu_G(A \cap gA) + \mu_G(B) = \mu_G(A \cup gA) + \mu_G(B) \leq 1, \tag{6.3}$$

and $\mu_G(A \cap gA) \geq \max\{d_1, \mu_G(A) - m\}$. Such g exists by Lemma 6.19. Let $A_1 = A \cap gA$, then $\mu_G(A_1) \leq \mu_G(A) - m, \mu_G(A)^2$. By Lemma 6.20, $\mathfrak{d}_G(A_1, B) \leq 2\gamma$. Next we choose $g_1 \in G$ satisfying (6.3) with A replaced by A_1, and $\mu_G(A_1 \cap g_1 A_1) \geq \min\{d_1, \mu_G(A_1)^2\}$. Let $A_2 = A_1 \cap g_1 A_1$, then $\mu_G(A_2) \leq \max\{\mu_G(A_1) - 2m, \mu_G(A_1)^2\}$, and $\mathfrak{d}_G(A_2, B) \leq 4\gamma$. Repeat this procedure for t_1 steps until either $\mu_G(A_{t_1}) = d_1$, or $\mu_G(A_t) - 2^{t-1}m \leq \mu_G(A_t)^2$. In either case we have $t_1 \leq \log(1/3m)$.

Next, if $\mu_G(A_{t_1}) > d_1$, we choose g_{t_1} in G such that $\mu_G(A_{t_1} \cap g_{t_1} A_{t_1}) = \mu_G(A_{t_1})^2$. By the way we define t_1, we have $\mu_G(A_{t_1} \cup g_{t_1} A_{t_1}) + \mu_G(B) \leq 1$. Set $A_{t_1+1} = A_{t_1} \cap g_{t_1} A_{t_1}$. Repeat this procedure for t_2 steps until $\mu_G(A_{t_1+t_2}) = d_1$. We have

$$t_2 \leq \log \frac{\log d_1}{\log \mu_G(A_{t_1})} \leq \log\log \frac{1}{d_1},$$

and $\mathfrak{d}_G(A_{t_1+t_2}, B) \leq 2^{t_1+t_2}\gamma = O_{d_1}(\gamma/m)$. We then apply the same procedures for B to arrange B having measure d_2.

If we have $\mu_G(A) < 1/3$ at the beginning, we are able to choose g such that $\mu_G(A \cap gA) = \mu_G(A)^2$ and $\mu_G(A \cup gA) + \mu_G(B) \leq 1$. Hence it only requires at most $\log\log(1/d_1)$ steps to make A having measure d_1. $\square$

6.6 Geometry of minimal and nearly minimal expansion pairs I

This section studies the shape of a nearly minimally expanding pairs relative to a connected compact normal subgroup of the ambient topological group such that the images of the pair under the quotient map have small measure. In Section 6.1, we obtain results that will allow us to reduce the Theorem 6.1 and Theorem 6.2 to analogous result about a simpler quotient group. Section 6.2 applies Section 6.1 to reduce Theorem 6.1 and Theorem 6.2 to the case of Lie groups and also prove a coarse version of these results. Section 6.3 applies Section 6.1 to further reduce Theorem 6.1 and Theorem 6.2 to the problem of constructing suitable group homomorphism into either $\mathbb{T}$ or $\mathbb{R}$.

Throughout this section, G is a connected unimodular locally compact group with Haar measure μ_G, and A and B are σ-compact subsets of G with positive μ_G-measure. We will assume familiarity with the preliminary Section 3.1 on locally compact group and Haar measuure.

6.6.1 Preservation of minimal expansion under quotient

In this section, H is a *connected compact* normal subgroup of G, so H and G/H are unimodular by Fact 6.10. Let μ_H, and $\mu_{G/H}$ be the Haar measure on G, H, and G/H, and let $\widetilde{\mu}_G$ and $\widetilde{\mu}_{G/H}$ be the inner Haar measures on G and G/H. We also let

Suppose r and s are in $\mathbb{R}$, the sets $A_{(r,s]}$ and $\pi A_{(r,s]}$ are given by

$$A_{(r,s]} := \{a \in A : \mu_H(A \cap aH) \in (r,s]\}$$

and

$$\pi A_{(r,s]} := \{aH \in G/H : \mu_H(A \cap aH) \in (r,s]\}.$$

In particular, $\pi A_{(r,s]}$ is the image of $A_{(r,s]}$ under the map π. We have a number of immediate observations. We define $B_{(r',s']}$ and $\pi B_{(r',s']}$ likewise for $r', s' \in \mathbb{R}$.

Lemma 6.22. *Let r, s, r', s' be in $\mathbb{R}^{>0}$. For all $aH \in \pi A_{(r,s]}$, $bH \in \pi B_{(r',s']}$, the sets $A_{(r,s]} \cap aH$, $B_{(r',s']} \cap bH$ are nonempty σ-compact. For all subintervals $(r,s]$ of $(0,1]$, $A_{(r,s]}$ is μ_G-measurable and $\pi A_{(r,s]}$ is $\mu_{G/H}$-measurable.*

Proof. The first assertion is immediate from the definition. Let $\mathbb{1}_A$ be the indicator function of A. Then the function

$$\begin{aligned} \mathbb{1}_A^H : G/H &\to R \\ aH &\mapsto \mu_H(A \cap aH) \end{aligned}$$

is well-defined and measurable by Lemma 6.9. As $\pi A_{(r,s]} = (\mathbb{1}_A^H)^{-1}(r,s]$ and $A_{(r,s]} = A \cap \pi^{-1}(\pi A_{(r,s]})$, we get the second assertion. $\square$

Note that $\pi A_{(r,s]} \pi B_{(r',s']}$ is not necessarily $\mu_{G/H}$-measurable, so Lemma 6.23(ii) does requires the inner measure $\widetilde{\mu}_{G/H}$.

Lemma 6.23. *We have the following:*

(i) *For every $aH \in \pi A$ and $bH \in \pi B$,*

$$\mu_H\big((A \cap aH)(B \cap bH)\big) \geq \min\{\mu_H(A \cap aH) + \mu_H(B \cap bH), 1\}.$$

(ii) *If $A_{(r,s]}$ and $B_{(r',s']}$ are nonempty, then*

$$\widetilde{\mu}_{G/H}(\pi A_{(r,s]} \pi B_{(r',s']}) \geq \min\{\mu_{G/H}(\pi A_{(r,s]}) + \mu_{G/H}(\pi B_{(r',s']}), \mu_{G/H}(G/H)\}.$$

Proof. Note that both H and G/H are connected. So (i) is a consequence of the Kemperman inequality for H and (ii) is a consequence of the generalized Kemperman inequality for G/H

(Fact 6.13). $\square$

As the functions we are dealing with are not differentiable, we will need Riemann–Stieltjes integral which we will now recall. Consider a closed interval $[a, b]$ of $\mathbb{R}$, and functions $f : [a, b] \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$. A *partition* P of $[a, b]$ is a sequence $(x_i)_{i=0}^n$ of real numbers with $x_0 = a$, $x_n = b$, and $x_i < x_{i+1}$ for $i \in \{0, \dots, n-1\}$. For such P, its *norm* $\|P\|$ is defined as $\max_{i=0}^{n-1} |x_{i+1} - x_i|$, and a corresponding *partial sum* is given by $S(P, f, g) = \sum_{i=0}^n f(c_{i+1})(g(x_{i+1}) - g(x_i))$ with $c_{i+1} \in [x_i, x_i + 1]$. We then define

$$\int_a^b f(x)\, \mathrm{d}g(x) := \lim_{\|P\| \to 0} S(P, f, g)$$

if this limit exists where we let P range over all the partition of $[a, b]$ and $S(P, f, g)$ ranges over all the corresponding partial sums of P. The next fact records some basic properties of the integral.

Fact 6.24. *Let $[a, b]$, $f(x)$, and $g(x)$ be as above. Then we have:*

(i) (Integrability) *If $f(x)$ is continuous on I, and $g(x)$ is monotone and bounded on $[a, b]$, then $f(x)\, \mathrm{d}g(x)$ is Riemann–Stieltjes integrable on $[a, b]$.*

(ii) (Integration by parts) *If $f(x)\, \mathrm{d}g(x)$ is Riemann–Stieltjes integrable on interval $[a, b]$, then $g(x)\, \mathrm{d}f(x)$ is also Riemann–Stieltjes integrable on $[a, b]$, and*

$$\int_a^b f(x)\, \mathrm{d}g(x) = f(b)g(b) - f(a)g(a) - \int_a^b g(x)\, \mathrm{d}f(x).$$

The next lemma uses "spillover" estimate, which gives us a lower bound estimate on $\mu_G(AB)$ when the projection of A and B are not too large.

Lemma 6.25. *Suppose $\mu_{G/H}(\pi A) + \mu_{G/H}(\pi B) < 1$. Set $\alpha = \sup_{a \in A} \mu_H(A \cap aH)$, $\beta =$*

$\sup_{b\in B}\mu_H(B\cap bH)$, *and* $\gamma = \max\{1, \alpha+\beta\}$. *Then*

$$\mu_G(AB) \geq \frac{\alpha+\beta}{\gamma}\left(\mu_{G/H}(\pi A_{(\alpha/\gamma,\alpha]}) + \mu_{G/H}(\pi B_{(\beta/\gamma,\beta]})\right) + \frac{\alpha+\beta}{\alpha}\mu_G(A_{(0,\alpha/\gamma]}) + \frac{\alpha+\beta}{\beta}\mu_G(B_{(0,\beta/\gamma]}).$$

Proof. For $x \in (0,1]$, set $C_x = AB \cap \pi^{-1}(\pi A_{(x\alpha,\alpha]}\pi B_{(x\beta,\beta]})$. One first note that

$$\mu_G(AB) \geq \widetilde{\mu}_G(C_0).$$

By Fact 6.24(1), $\mathrm{d}\widetilde{\mu}_G(C_x)$ is Riemann–Stieltjes integrable on any closed subinterval of $[0,1]$. Hence,

$$\widetilde{\mu}_G(C_0) = \widetilde{\mu}_G(C_{1/\gamma}) - \int_0^{\frac{1}{\gamma}} \mathrm{d}\widetilde{\mu}_G(C_x).$$

Lemma 6.22 and Lemma 6.23(1) give us that

$$\widetilde{\mu}_G(C_{1/\gamma}) \geq \widetilde{\mu}_{G/H}(\pi A_{(\alpha/\gamma,\alpha]}\pi B_{(\beta/\gamma,\beta]}).$$

Likewise, for $x, y \in \mathbb{R}^{>0}$ with $x < y \leq 1/\gamma$, $\widetilde{\mu}_G(C_x) - \widetilde{\mu}_G(C_y)$ is at least

$$r(\alpha+\beta)\left(\widetilde{\mu}_{G/H}(\pi A_{(x\alpha,\alpha]}\pi B_{(x\beta,\beta]}) - \widetilde{\mu}_{G/H}(\pi A_{(y\alpha,\alpha]}\pi B_{(y\beta,\beta]})\right).$$

Therefore,

$$\widetilde{\mu}_G(C_0) \geq \widetilde{\mu}_{G/H}(\pi A_{(\alpha/\gamma,\alpha]}\pi B_{(\beta/\gamma,\beta]}) - \int_0^{\frac{1}{\gamma}} (\alpha+\beta)x\,\mathrm{d}\widetilde{\mu}_{G/H}(\pi A_{(x\alpha,\alpha]}\pi B_{(x\beta,\beta]}).$$

Using integral by parts (Fact 6.24.2), we get

$$\widetilde{\mu}_G(C_0) \geq \int_0^{\frac{1}{\gamma}} \widetilde{\mu}_{G/H}(\pi A_{(x\alpha,\alpha]}\pi B_{(x\beta,\beta]})\,\mathrm{d}(\alpha+\beta)x.$$

Applying Lemma 6.23.2 and using the assumption that $\mu_{G/H}(\pi A)+\mu_{G/H}(\pi B) < 1$, we have

$$\widetilde{\mu}_G(C_0) \geq \int_0^{\frac{1}{\gamma}} (\mu_{G/H}(\pi A_{(x\alpha,\alpha]}) + \mu_{G/H}(\pi B_{(x\beta,\beta]}))\, \mathrm{d}(\alpha+\beta)x.$$

Using integral by parts (Fact 6.24.2), we arrive at

$$\begin{aligned}\widetilde{\mu}_G(C_0) \geq{}& \frac{\alpha+\beta}{\gamma}\left(\mu_{G/H}(\pi A_{(\alpha/\gamma,\alpha]}) + \mu_{G/H}(\pi B_{(\beta/\gamma,\beta]})\right)\\ &- \int_0^{\frac{1}{\gamma}} (\alpha+\beta)x\, \mathrm{d}(\mu_{G/H}(\pi A_{(x\alpha,\alpha]}) + \mu_{G/H}(\pi B_{(x\beta,\beta]})).\end{aligned}$$

As $\mathrm{d}(\mu_{G/H}(\pi A_{(x\alpha,\alpha]}) + \mu_{G/H}(\pi B_{(x\beta,\beta]})) = -\,\mathrm{d}(\mu_{G/H}(\pi A_{(0,x\alpha]}) + \mu_{G/H}(\pi B_{(0,x\beta]}))$,

$$\begin{aligned}\widetilde{\mu}_G(C_0) \geq{}& \frac{\alpha+\beta}{\gamma}\left(\mu_{G/H}(\pi A_{(\alpha/\gamma,\alpha]}) + \mu_{G/H}(\pi B_{(\beta/\gamma,\beta]})\right)\\ &+ \int_0^{\frac{1}{\gamma}} (\alpha+\beta)x\, \mathrm{d}(\mu_{G/H}(\pi A_{(0,x\alpha]}) + \mu_{G/H}(\pi B_{(0,x\beta]})).\end{aligned}$$

Finally, recall that

$$\int_0^{1/\gamma} x\alpha\, \mathrm{d}\mu_{G/H}(\pi A_{(0,x\alpha]}) = \mu_G(A_{(0,\alpha/\gamma]}) \text{ and } \int_0^{1/\gamma} \beta x\, \mathrm{d}\mu_{G/H}(\pi B_{(0,x\beta]}) = \mu_G(B_{(0,\beta/\gamma]}).$$

Thus, we arrived at the desired conclusion. □

The next result in the main result in this subsection. It says if the projections of A and B are not too large, the small expansion properties will be kept in the quotient group.

Theorem 6.26 (Quotient domination)**.** *Suppose $\mu_{G/H}(\pi A)+\mu_{G/H}(\pi B) < \mu_{G/H}(G/H)$ and $\mathfrak{d}_G(A,B) < \min\{\mu_G(A), \mu_G(B)\}$. Then there are σ-compact $A', B' \subseteq G/H$ such that*

$$\mathfrak{d}_{G/H}(A',B') < 7\mathfrak{d}_G(A,B)$$

and $\max\{\mu_G(A \triangle \pi^{-1}A'), \mu_G(B\triangle \pi^{-1}B')\} < 3\mathfrak{d}_G(A,B)$.

Proof. Let α and β be as in Lemma 6.25. We first show that $\alpha+\beta \geq 1$. Suppose to the contrary that $\alpha+\beta<1$. Then Lemma 6.25 gives us

$$\mu_G(AB) \geq \frac{\alpha+\beta}{\alpha}\mu_G(A) + \frac{\alpha+\beta}{\beta}\mu_G(B)$$

It follows that $\mu_G(AB) > \mu_G(A) + \mu_G(B) + \min\{\mu_G(A), \mu_G(B)\}$, a contradiction.

Now we have $\alpha+\beta \geq 1$. Hence, Lemma 6.25 yields

$$\begin{aligned}\mu_G(AB) \geq\ & \mu_{G/H}(\pi A_{(\alpha/(\alpha+\beta),\alpha]}) + \mu_{G/H}(\pi B_{(\beta/(\alpha+\beta),\beta]}) \\ & + \frac{\alpha+\beta}{\alpha}\mu_G(A_{(0,\alpha/\gamma]}) + \frac{\alpha+\beta}{\beta}\mu_G(B_{(0,\beta/(\alpha+\beta)]}).\end{aligned}$$

Choose σ-compact $A' \subseteq \pi A_{(\alpha/(\alpha+\beta),\alpha]}$ and $B' \subseteq \pi B_{(\beta/(\alpha+\beta),\beta]})$ σ-compact such that

$$\mu_{G/H}(A') = \mu_{G/H}(\pi A_{(\alpha/(\alpha+\beta),\alpha]}) \text{ and } \mu_{G/H}(B') = \mu_{G/H}(\pi B_{(\beta/(\alpha+\beta),\beta]}).$$

We will verify that A' and B' satisfy the desired conclusion.

Since $\mu_{G/H}(A') \geq (1/\alpha)\mu_G(A_{(\alpha/(\alpha+\beta),\alpha]})$, $\mu_{G/H}(B') \geq (1/\beta)\mu_G(B_{(\beta/(\alpha+\beta),\beta]})$ and $\alpha+\beta>1$, we have

$$\mu_G(AB) \geq \frac{1}{\alpha}\mu_G(A) + \frac{1}{\beta}\mu_G(B).$$

From $\mu_G(AB) - \mu_G(A) - \mu_G(B) = \mathfrak{d}_G(A,B) \leq \min\{\mu_G(A), \mu_G(B)\}$, we deduce that $\alpha, \beta \geq 1/2$.

By our assumption $\mu_G(AB) < \mu_G(A) + \mu_G(B) + \mathfrak{d}_G(A,B)$. Hence,

$$\begin{aligned}\mathfrak{d}_G(A,B) \geq\ & \mu_{G/H}(A') - \mu_G(A_{(\alpha/(\alpha+\beta),\alpha]}) + \mu_{G/H}(B') - \mu_G(B_{(\beta/(\alpha+\beta),\beta]}) \\ & + \frac{\beta}{\alpha}\mu_G(A_{(0,\alpha/\gamma]}) + \frac{\alpha}{\beta}\mu_G(B_{(0,\beta/(\alpha+\beta)]}).\end{aligned}$$

Therefore, $\mu_{G/H}(A') - \mu_G(A_{(\alpha/(\alpha+\beta),\alpha]})$ and $(\beta/\alpha)\mu_G(A_{(0,\alpha/\gamma]})$ are at most $\mathfrak{d}_G(A,B)$. Noting also that $\beta/\alpha \leq 1/2$, we get $\mu_G(A \triangle \pi^{-1}(A') \leq 3\mathfrak{d}_G(A,B)$. A similar argument yield $\mu_G(B \triangle \pi^{-1}(B') \leq 3\mathfrak{d}_G(A,B)$.

Finally, note that $\pi^{-1}(A'B')$ is equal to $A_{(\alpha/(\alpha+\beta),\alpha]}B_{(\beta/(\alpha+\beta),\beta]}$, which is a subset of AB. Combining with $\mu_G(AB) < \mu_G(A) + \mu_G(B) + \mathfrak{d}_G(A,B)$, we get

$$\mu_{G/H}(A'B') \leq \mu_G(A) + \mu_G(B) + \mathfrak{d}_G(A,B) \leq \mu_{G/H}(A') + \mu_{G/H}(B') + 7\mathfrak{d}_G(A,B),$$

which completes the proof. □

The next corollary of the proof of Theorem 6.26 gives a complementary result when without the asssumption that $\mu_{G/H}(\pi A) + \mu_{G/H}(\pi B) < \mu_{G/H}(G/H)$.

Corollary 6.27. *Suppose G is noncompact and $\mathfrak{d}_G(A,B) = 0$. Then there are σ-compact $A', B' \subseteq G/H$ such that $\mathfrak{d}_{G/H}(A',B') = 0$, $\mu_G(A \triangle \pi^{-1}A') = 0$, and $\mu_G(B \triangle \pi^{-1}B') = 0$.*

Proof. Choose an increasing sequence (A_n) of compact subsets of A and an increasing sequence (B_n) of compact subsets of B such that $A = \bigcup_{n=0}^{\infty} A_n$ and $B = \bigcup_{i=0}^{\infty} B_n$. Then $\lim_{n\to\infty} \mathfrak{d}_G(A_n, B_n) = 0$. For each n, A_n and B_n are compact, so πA_n and πB_n are also compact and has finite measure. Let A'_n and B'_n be defined for A_n and B_n as in the proof of Theorem 6.26. Then for n sufficiently large, we have

$$\mu_G(\pi^{-1}A'_n \triangle A_n) < 3\mathfrak{d}_G(A_n,B_n) \text{ and } \mu_G(\pi^{-1}B'_n \triangle B_n) < 3\mathfrak{d}_G(A_n,B_n)$$

and

$$\mu_{G/H}(A'_nB'_n) < \mu_{G/H}(A'_n) + \mu_{G/H}(B'_n) + 5\mathfrak{d}_G(A_n,B_n).$$

Moreover, we can arrange that the sequences (A'_n) and (B'_n) are increasing. Take $A' = \bigcup_{n=1}^{\infty} A'_n$ and $B' = \bigcup_{n=1}^{\infty} B'_n$. By taking $n \to \infty$, we have

$$\mu_G(\pi^{-1}A' \triangle A) = 0 \text{ and } \mu_G(\pi^{-1}B' \triangle B) = 0.$$

and $\mathfrak{d}_{G/H}(A', B') = 0$ as desired. $\square$

6.6.2 Coarse versions of the main theorems

For the given G, there might be no continuous surjective group homomorphism to either $\mathbb{T}$ or $\mathbb{R}$ (e.g. $G = \mathrm{SO}_3(\mathbb{R})$). However, the famous theorem below by Gleason [73] and Yamabe [169] allows us to naturally obtain continuous and surjective group homomorphism to a Lie group. Using together with Corollary 6.27, this allows us to reduce the noncompact case of Theorem 6.1 to that of Lie group. The connectedness of H is not often stated as part of the result, but can be arranged by replacing H with its identity component.

Fact 6.28 (Gleason–Yamabe Theorem)**.** *For any connected locally compact group G and any neighborhood U of the identity in G, there is a connected compact normal subgroup $H \subseteq U$ of G such that G/H is Lie group.*

With some further effort, we can also arrange that $\mu_{G/H}(\pi A) + \mu_{G/H}(\pi B) < \mu_{G/H}(G/H)$ as necessary to apply Theorem 6.26. However, when $\mathfrak{d}_G(A, B) > 0$, we will need a dimension control on the Lie group we obtained from the Gleason–Yamabe Theorem. For that, we need Fact 6.30, which can be thought of as a refinement of the Gleason–Yamabe theorem coming from arithmetic combinatorics and model theory.

Recall that an open precompact set $S \subseteq G$ is a K-**approximate** group if $\mathrm{id}_G \in S$, $S^{-1} = S$, and $S^2 \subseteq XS$ for some finite set X of cardinality K. The next theorem by Tao [159] allows us to extract an approximate group from a nearly minimally expanding pair; as stated in [159], this theorem is only applicable when A, B are open, but the proof also goes through without this assumption.

Fact 6.29 (Approximate groups from small expansion)**.** *Suppose K is a constant and $\mu_G(AB) < K\mu_G^{1/2}(A)\mu_G^{1/2}(B)$, then there is an open precompact $O(K^{O(1)})$-approximate group S with $\mu_G(S) = O(K^{O(1)})\mu_G^{1/2}(A)\mu_G^{1/2}(B)$ and a finite set X or cardinality $O(K^{O(1)})$ such that $A \subseteq XS$ and $B \subseteq SX$.*

The study of continuous approximate groups by Carolino [40] is able to find a Lie model, and to control the dimension of the Lie model. This can be seen as a finer version of the Gleason–Yamabe theorem.

Fact 6.30 (Lie model from approximate groups)**.** *Suppose K is a constant and S is an open precompact K-approximate group on G. Then there is a connected compact normal subgroup H of G, such that $H \subseteq S^4$ and G/H is a Lie group of dimension $O_K(1)$.*

The next lemma is the main result in this subsection. Using this, we can pass the problem to connected Lie groups with bounded dimensions.

Lemma 6.31 (Lie model from small expansions)**.** *If $\mu_G(AB) \leq K\mu_G^{1/2}(A)\mu_G^{1/2}(B)$ and then there is a connected compact subgroup H of G such that G/H is a Lie group of dimension $O_K(1)$ and, with $\pi : G \to G/H$ the quotient map, πA and πB have μ_G-measure $O(K^{O(1)})\mu_G^{1/2}(A)\mu_G^{1/2}(B)$.*

Proof. By Fact 6.29, there is an open K-approximate group S, with

$$\mu_G(S) = O(K^{O(1)})\mu_G^{1/2}(A)\mu_G^{1/2}(B)$$

such that A can be covered by $O(K^{O(1)})$ right translation of S, and B can be covered by $O(K^{O(1)})$ left translation of S. By Fact 6.30, there is a closed connected normal subgroup H in S^4, such that G/H is a Lie group of dimension at most $O_K(1)$. Let π be the quotient map. Since $H \subseteq S^4$, we have

$$\mu_{G/H}(\pi(S)) = \mu_G(SH) \leq \mu_G(S^5) = O(K^{O(1)})\mu_G^{1/2}(A)\mu_G^{1/2}(B).$$

Note that $\pi(A)$ can be covered by $O(K^{O(1)})$ right translations of $\pi(S)$, and $\pi(B)$ can be covered by $O(K^{O(1)})$ left translations of $\pi(S)$. Hence, we get the desired conclusion. $\square$

The following proposition tells us that small measure expansion phenomenon can always be reduced to the same phenomenon on a Lie group with small dimension.

Proposition 6.32 (Coarse versions of the main theorems)**.** *There is a constant τ such that if either G is noncompact and $\mu_G(A) = \mu_G(B)$ or G is compact and $\mu_G(A) = \mu_G(B) < \tau$, and*

$$\mathfrak{d}_G(A, B) < \min\{\mu_G(A), \mu_G(B)\},$$

then there is a connected compact normal subgroup H of G, and σ-compact subsets A', B' of G/H satisfying:

(i) *G/H is a Lie group of dimension $O(1)$;*

(ii) *With $\pi : G \to G/H$ the quotient map,*

$$\mu_G(A \triangle \pi^{-1}A') < 3\mathfrak{d}_G(A, B) \text{ and } \mu_G(B \triangle \pi^{-1}B') < 3\mathfrak{d}_G(A, B);$$

(iii) *$\mathfrak{d}_{G/H}(A', B') < 7\mathfrak{d}_G(A, B)$.*

Proof. From the assumption we have $\mu_G(AB) < 3\mu_G^{1/2}(A)\mu_G^{1/2}B$. Obtain H as in Lemma 6.31, and when $\mu_G(A), \mu_G(B)$ are small enough, we have $\mu_{G/H}(\pi A)+\mu_{G/H}(\pi B) < 1$. Then G/H is a Lie group of dimension $O(1)$. By applying Theorem 6.26, we get the desired conclusion. □

6.6.3 Structure control on the nearly minimal expansion sets

The following useful lemma is a corollary of Theorem 6.26, which will be used at various points in the later proofs. It tells us the character given in the parallel Bohr sets is essentially unique.

Lemma 6.33 (Stability of characters)**.** *Suppose G is compact, $\chi : G \to \mathbb{T}$ is a continuous surjective group homomorphism, $J \subseteq \mathbb{T}$ is a compact interval, and η is a constant. Suppose we have*

(i) *$\mathfrak{d}_G(A, B) < \min\{\mu_G(A), \mu_G(B)\}$;*

(ii) $\mu_{\mathbb{T}}(J) = \mu_G(B)$ *and* $\mu_G(B \triangle \chi^{-1}(J)) = \eta \mathfrak{d}_G(A, B)$*;*

Then there is a compact interval I *in* $\mathbb{T}$ *such that* $\mu_{\mathbb{T}}(I) = \mu_G(A)$ *and*

$$\mu_G(A \triangle \chi^{-1}(I)) \le (45 + 2\eta)\mathfrak{d}_G(A, B).$$

Proof. Let $K = \ker(\chi)$. Note that χ is an open map, and hence by Fact 6.7 we have $G/K \cong \mathbb{T}$. By Theorem 6.26, there are sets A' and B' in $\mathbb{T}$ such that $\mathfrak{d}_{\mathbb{T}}(A', B') \le 7\mathfrak{d}_G(A, B)$, and

$$\mu_G(A \triangle \chi^{-1}(A')) = 3\mathfrak{d}_G(A, B), \quad \mu_G(B \triangle \chi^{-1}(B')) = 3\mathfrak{d}_G(A, B).$$

Then $\mu_{\mathbb{T}}(B' \triangle J) \le (\eta + 3)\mathfrak{d}_G(A, B)$. Without loss of generality, we assume $\mu_G(B) = \kappa$, and $J = [0, \kappa]$. As A' is measurable, for any $\varepsilon > 0$, there is a finite union of intervals $K \subseteq A'$ in $\mathbb{T}$ such that $\mu_{\mathbb{T}}(K) \ge \mu_{\mathbb{T}}(A') - \varepsilon$. Fix $\varepsilon = \mathfrak{d}_G(A, B)$. For every $x \in [0, \kappa]$, define

$$f : \mathbb{R} \to \mathbb{R}, x \mapsto \mu_{\mathbb{T}}(K + [0, x]).$$

Note that f is continuous and piecewise linear. Thus by the fundamental theorem of calculus

$$\begin{aligned}\int_0^\kappa f'(x)\,\mathrm{d}x &= f(\kappa) - f(0) \\ &\le \mu_{\mathbb{T}}(A') + \kappa + (7 + 3 + \eta)\mathfrak{d}_G(A, B) - \mu_{\mathbb{T}}(A') + \mathfrak{d}_G(A, B) \\ &\le \kappa + (11 + \eta)\mathfrak{d}_G(A, B).\end{aligned}$$

Note that $f' \ge 1$ and only taking values in $\mathbb{Z}$. Hence, there is $\Omega \subseteq [0, \kappa]$ with $\mu_{\mathbb{R}}(\Omega) \ge \kappa - (11 + \eta)\mathfrak{d}_G(A, B)$, such that $f'(x) = 1$ for $x \in \Omega$. Clearly, there is $x_0 \le (11 + \eta)\mathfrak{d}_G(A, B)$, and $f'(x_0) = 1$. This implies $K + [0, x_0]$ is an interval. Again we have

$$\begin{aligned}\mu_{\mathbb{T}}(K + [0, x_0]) &= f(\kappa) - \int_{x_0}^\kappa f'(x)\,\mathrm{d}x \\ &\le \mu_{\mathbb{T}}(A') + (21 + \eta)\mathfrak{d}_G(A, B).\end{aligned}$$

Thus, there is an interval $I \subseteq \mathbb{T}$ with $\mu_{\mathbb{T}}(I) = \mu_G(A)$, and $\mu_G(A \bigtriangleup \chi^{-1}(I)) \leq (2(21+\eta)+3)\mathfrak{d}_G(A,B) = (45+2\eta)\mathfrak{d}_G(A,B)$. $\square$

The next lemma shows that, if the symmetric difference of a set A and an interval is small, then A is also contained in a interval of bounded length.

Lemma 6.34. *Suppose G is compact, A, B are σ-compact subsets of G, and the discrepancy $\mathfrak{d}_G(A,B) < \varepsilon$. Let $\chi : G \to \mathbb{T}$ be a continuous surjective group homomorphism, and I, J compact intervals in $\mathbb{T}$, with $\mu_{\mathbb{T}}(I) = \mu_G(A), \mu_{\mathbb{T}}(J) = \mu_G(B)$, and*

$$\mu_G(A \bigtriangleup \chi^{-1}(I)) < \varepsilon, \quad \mu_G(B \bigtriangleup \chi^{-1}(J)) < \varepsilon.$$

Then there are intervals $I', J' \subseteq \mathbb{T}$, such that $A \subseteq \chi^{-1}(I')$, $B \subseteq \chi^{-1}(J')$, and

$$\mu_{\mathbb{T}}(I') - \mu_G(A) < 10\varepsilon, \quad \mu_{\mathbb{T}}(J') - \mu_G(B) < 10\varepsilon.$$

Proof. Suppose there is $g \in A$ and $g \notin \chi^{-1}(I)$, and the distance between $\chi(g)$ and the nearest element in I' is strictly greater than $5\varepsilon\mu_G(A)$ in $\mathbb{T}$. Thus

$$\mu_{\mathbb{T}}(\chi(g)\chi(B) \setminus I\chi(B)) \geq 5\varepsilon - \mu_{\mathbb{T}}(J \setminus \chi(B)) \geq 4\varepsilon.$$

and this implies that $\mu_G(gB \setminus \chi^{-1}(I)\chi^{-1}(J)) \geq 3\varepsilon$. Therefore,

$$\begin{aligned} \mu_G(AB) &\geq \mu_G\big((\chi^{-1}(I) \cap A)(\chi^{-1}(J) \cap B)\big) + \mu_G(gB \setminus \chi^{-1}(I)\chi^{-1}(J)) \\ &\geq \mu_G(A) + \mu_G(B) - 2\varepsilon + 3\varepsilon, \end{aligned}$$

and this contradicts the fact that $\mathfrak{d}_G(A,B) < \varepsilon$. Hence there are intervals I', J' in $\mathbb{T}$ such that $A \subseteq \chi^{-1}(I')$ and $B \subseteq \chi^{-1}(J')$, and

$$\mu_G(\chi^{-1}(I') \setminus A) < 10\varepsilon \quad \mu_{\mathbb{T}}(\chi^{-1}(J') \setminus B) < 10\varepsilon,$$

as desired. $\square$

The stability lemma, together with Theorem 6.26, will be enough to derive a different proof of a theorem by Tao [162], with a sharp exponent bound. As we mentioned in the introduction, the same result with a sharp exponent bound was also obtained by Christ and Iliopoulou [43] recently, via a different approach.

Theorem 6.35 (Theorem 6.2 for compact abelian groups)**.** *Let G be a connected compact abelian group, and A, B be compact subsets of G with positive measure. Set*

$$s = \min\{\mu_G(A), \mu_G(B), 1 - \mu_G(A) - \mu_G(B)\}.$$

Given $0 < \varepsilon < 1$, there is a constant $K = K(s)$ does not depends on G, such that if $\delta < K\varepsilon$ and

$$\mu_G(A + B) < \mu_G(A) + \mu_G(B) + \delta \min\{\mu_G(A), \mu_G(B)\}.$$

Then there is a surjective continuous group homomorphism $\chi : G \to \mathbb{T}$ together with two compact intervals $I, J \in \mathbb{T}$ with

$$\mu_{\mathbb{T}}(I) - \mu_G(A) < \varepsilon\mu_G(A), \quad \mu_{\mathbb{T}}(J) - \mu_G(B) < \varepsilon\mu_G(B),$$

and $A \subseteq \chi^{-1}(I)$, $B \subseteq \chi^{-1}(J)$.

Proof. We first assume that $\mathfrak{d}_G(A, B)$ is sufficiently small, and we will compute the bound on $\mathfrak{d}_G(A, B)$ later. As G is abelian, by Proposition 6.32, there is a quotient map $\pi : G \to \mathbb{T}^d$, and $A', B' \subseteq \mathbb{T}^d$, such that

$$\mu_G(A \triangle \pi^{-1}A') < 3\mathfrak{d}_G(A, B) \text{ and } \mu_G(B \triangle \pi^{-1}B') < 3\mathfrak{d}_G(A, B)$$

and $\mathfrak{d}_{G/H}(A', B') < 7\mathfrak{d}_G(A, B)$. Let $c = c(\tau)$ be as in Fact 6.14, and by Lemma 6.21, there is a constant L depending only on s and c, and sets $A'', B'' \subseteq \mathbb{T}^d$ with $\mu_{\mathbb{T}^d}(A'') = \mu_{\mathbb{T}^d}(B'') = c$

such that

$$\max\{\mathfrak{d}_G(A'', B'), \mathfrak{d}_G(A'', B''), \mathfrak{d}_G(A', B'')\} < L\mathfrak{d}_G(A, B).$$

By Fact 6.14, there are intervals $I', J' \subseteq \mathbb{T}$ with $\mu_{\mathbb{T}}(I') = \mu_{\mathbb{T}}(J') = c$, and a continuous surjective group homomorphism $\rho : \mathbb{T}^d \to \mathbb{T}$, such that

$$\mu_{\mathbb{T}^d}(A'' \triangle \rho^{-1}(I')) < L\mathfrak{d}_G(A, B) \quad \text{and} \quad \mu_{\mathbb{T}^d}(B'' \triangle \rho^{-1}(J')) < L\mathfrak{d}_G(A, B).$$

By Lemma 6.33, there are intervals $I', J' \subseteq \mathbb{T}$ with

$$\mu_{\mathbb{T}^d}(A' \triangle \rho^{-1}(I')) < (45 + 2L)\mathfrak{d}_G(A, B) \text{ and } \mu_{\mathbb{T}^d}(B' \triangle \rho^{-1}(J')) < (45 + 2L)\mathfrak{d}_G(A, B).$$

Let $\chi = \pi \circ \rho$. Hence we have

$$\mu_G(A \triangle \chi^{-1}(I')) < (48 + 2L)\mathfrak{d}_G(A, B) \text{ and } \mu_G(B \triangle \chi^{-1}(J')) < (48 + 2L)\mathfrak{d}_G(A, B).$$

Using Lemma 6.34, there are intervals $I, J \subseteq \mathbb{T}$, such that $A \subseteq \chi^{-1}(I)$, $B \subseteq \chi^{-1}(J)$, and

$$\mu_{\mathbb{T}}(I) - \mu_G(A) < (480 + 20L)\mathfrak{d}_G(A, B),$$
$$\mu_{\mathbb{T}}(J) - \mu_G(B) < (480 + 20L)\mathfrak{d}_G(A, B).$$

Now, we fix

$$K := \min\left\{\frac{1}{480 + 20L}, \frac{c}{L}\right\},$$

and $\delta < K\varepsilon$, where $\mathfrak{d}_G(A, B) = \delta \min\{\mu_G(A), \mu_G(B)\}$. Clearly, we will have

$$\mu_{\mathbb{T}}(I) - \mu_G(A) < \varepsilon \min\{\mu_G(A), \mu_G(B)\},$$
$$\mu_{\mathbb{T}}(J) - \mu_G(B) < \varepsilon \min\{\mu_G(A), \mu_G(B)\}.$$

Note that in the above argument, we apply Fact 6.14 on A'', B'', and this would require that $L\mathfrak{d}_G(A,B) < c$. By the way we choose K, we have

$$L\mathfrak{d}_G(A,B) = L\delta \min\{\mu_G(A), \mu_G(B)\} < c,$$

as desired. $\square$

The following theorem shows that, once we have a certain group homomorphism to tori, we will get a good structural control on the (nearly) minimal expansion sets.

Proposition 6.36 (Toric domination from a given homomorphism)**.** *Suppose A, B have $\mathfrak{d}_G(A,B) < \min\{\mu_G(A), \mu_G(B)\}$, and $\chi : G \to \mathbb{T}$ is a continuous surjective group homomorphism such that $\mu_{\mathbb{T}}(\chi(A)) + \mu_{\mathbb{T}}(\chi(B)) < 1/5$. Then there is a continuous and surjective group homomorphism $\rho : G \to \mathbb{T}$, a constant K_0 only depending on $\min\{\mu_G(A), \mu_G(B)\}$, and compact intervals $I, J \subseteq \mathbb{T}$ with $\mu_{\mathbb{T}}(I) = \mu_G(A)$ and $\mu_{\mathbb{T}}(J) = \mu_G(B)$, such that*

$$\mu_G(A \triangle \rho^{-1}(I)) < K_0\mathfrak{d}_G(A,B), \quad \textit{and} \quad \mu_G(B \triangle \rho^{-1}(J)) < K_0\mathfrak{d}_G(A,B).$$

Proof. By Theorem 6.26, there are $A', B' \subseteq \mathbb{T}$, such that

$$\mu_G(A \triangle \chi^{-1}(A')) < 3\mathfrak{d}_G(A,B) \text{ and } \mu_G(B \triangle \chi^{-1}(B')) < 3\mathfrak{d}_G(A,B), \tag{6.4}$$

and $\mathfrak{d}_{\mathbb{T}}(A', B') < 7\mathfrak{d}_G(A,B)$. By Theorem 6.35, there are continuous surjective group homomorphism $\eta : \mathbb{T} \to \mathbb{T}$, a constant L depending only on $\min\{\mu_G(A), \mu_G(B)\}$, and compact intervals $I, J \subseteq \mathbb{T}$ such that $\mu_{\mathbb{T}}(I) = \mu_{\mathbb{T}}(A')$, $\mu_{\mathbb{T}}(J) = \mu_{\mathbb{T}}(B')$, and

$$\mu_{\mathbb{T}}(A' \triangle \eta^{-1}(I)) < L\mathfrak{d}_G(A,B) \text{ and } \mu_{\mathbb{T}}(B' \triangle \eta^{-1}(J)) < L\mathfrak{d}_G(A,B). \tag{6.5}$$

Set $\rho = \eta \circ \chi$. The conclusion follows from (6.4) and (6.5) with $K_0 = L + 3$. $\square$

In light of Proposition 6.36, in the rest of the paper, we will be focusing on finding the desired group homomorphism mapping G to tori.

6.7 Pseudometrics and group homomorphisms into tori

Proposition 6.32 and Proposition 6.36 reduce the proof of Theorem 6.1 and Theorem 6.2 to problems of constructing certain group homomorphisms from a Lie group into $\mathbb{T}$ or $\mathbb{R}$. In this section, we show these problems can be reduced further to problems of constructing pseudometrics with certain properties on the ambient group. Section 7.1 shows that a linear pseudometric suffices, and Section 7.2 and Section 7.3 does so when the pseudometric is almost linear and almost monotone.

Throughout, G is a *connected* and *unimodular Lie group* with Haar measure μ_G. Recall that a **pseudometric** on a set X is a function $d : X \times X \to \mathbb{R}$ satisfying the following three properties:

1. (Reflexive) $d(a, a) = 0$ for all $a \in X$,

2. (Symmetry) $d(a, b) = d(b, a)$ for all $a, b \in X$,

3. (Triangle inequality) $d(a, c) \leq d(a, b) + d(b, c) \in X$.

Hence, a pseudometric on X is a metric if for all $a, b \in X$, we have $d(a, b) = 0$ implies $a = b$. If d is a pseudometric on G, for an element $g \in G$, we set $\|g\|_d = d(\mathrm{id}_G, g)$.

6.7.1 Linear pseudometrics

Suppose d is a pseudometric on G. We say that d is **left-invariant** if for all $g, g_1, g_2 \in G$, we have $d(gg_1, gg_2) = d(g_1, g_2)$. left-invariant pseudometrics arise naturally from measurable sets in a group; the pseudometric we will construct in Section 8 is of this form.

Proposition 6.37. *Suppose A is a measurable subset of G. For g_1 and g_2 in G, define*

$$d(g_1, g_2) = \mu_G(A) - \mu_G(g_1 A \cap g_2 A).$$

Then d is a continuous left-invariant pseudometric on G.

Proof. We first verify the triangle inequality. Let g_1, g_2, and g_3 be in G, we need to show that

$$\mu_G(A) - \mu_G(g_1 A \cap g_3 A) \leq \mu_G(A) - \mu_G(g_1 A \cap g_2 A) + \mu_G(A) - \mu_G(g_2 A \cap g_3 A). \tag{6.6}$$

As $\mu_G(A) = \mu_G(g_2 A)$, we have $\mu_G(A) - \mu_G(g_1 A \cap g_2 A) = \mu_G(g_2 A \setminus g_1 A)$, and $\mu_G(A) - \mu_G(g_2 A \cap g_3 A) = \mu_G(g_2 A \setminus g_3 A)$. Hence, (6.6) is equivalent to

$$\mu_G(g_2 A) - \mu_G(g_2 A \setminus g_1 A) - \mu_G(g_2 A \setminus g_3 A) \leq \mu(g_1 A \cap g_3 A).$$

Note that the left hand side is at most $\mu_G(g_1 A \cap g_2 A \cap g_3 A)$, which is less than the right hand side. Hence, we get the desired conclusion. The continuity of d follows from Fact 6.4(vii), and the remaining parts are straightforward. $\square$

Another natural source of left-invariant pseudometrics is group homomorphims onto metric groups. Suppose $\widetilde{d}$ is a continuous left-invariant metric on a group H and $\pi : G \to H$ is a group homomorphism, then for every g_1, g_2 in G, one can naturally define a pseudometric $d(g_1, g_2) = \widetilde{d}(\pi(g_1), \pi(g_2))$. It is easy to see that such d is a continuous left-invariant pseudometric, and $\{g \in G : \|g\|_d = 0\} = \ker(\pi)$ is a normal subgroup of G. The latter part of this statement is no longer true for an arbitrary continuous left-invariant pseudometric, but we still have the following:

Lemma 6.38. *Suppose d is a continuous left-invariant pseudometric on G. Then the set $\{g \in G : \|g\|_d = 0\}$ is the underlying set of a closed subgroup of G.*

Proof. Suppose g_1 and g_2 are elements in G such that $\|g_1\|_d = \|g_2\|_d = 0$. Then

$$d(\mathrm{id}_G, g_1g_2) \leq d(\mathrm{id}_G, g_1) + d(g_1, g_1g_2) = d(\mathrm{id}_G, g_1) + d(\mathrm{id}_G, g_2) = 0.$$

Now, suppose (g_n) is a sequence of elements in G converging to g with $\|g_n\|_d = 0$. Then $\|g\|_d = 0$ by continuity, we get the desired conclusions. □

In many situations, a left-invariant pseudometric allows us to construct surjective continuous group homomorphism into metric groups. The following lemma tells us precisely when this happens. We omit the proof as the result is motivationally relevant but will not be used later on.

Lemma 6.39. *Let d be a continuous left-invariant pseudometric on G. The following are equivalent*

(i) *The set $\{g \in G : \|g\|_d = 0\}$ is the underlying set of a closed normal subgroup of G.*

(ii) *There is a continuous surjective group homomorphism $\pi : G \to H$, and $\widetilde{d}$ is a left-invariant metric on H. Then*

$$d(g_1, g_2) = \widetilde{d}(\pi g_1, \pi g_2).$$

Moreover, when (ii) happens, $\{g \in G : \|g\|_d = 0\} = \ker \pi$, hence H and $\widetilde{d}$ if exist are uniquely determined up to isomorphism.

The group $\mathbb{R}$ and $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ are naturally equipped with the metrics $d_{\mathbb{R}}$ and $d_{\mathbb{T}}$ induced by the Euclidean norms, and these metrics interact in a very special way with the additive structures. Hence one would expect that if there is a group homomorphism from G to either $\mathbb{R}$ or $\mathbb{T}$, then G can be equipped with a pseudometric which interacts nontrivially with addition.

Let d be a left-invariant pseudometric on G. The **radius** ρ of d is defined to be $\sup\{\|g\|_d : g \in G\}$; this is also $\sup\{d(g_1, g_2) : g_1, g_2 \in G\}$ by left invariance. We say that d is **locally linear** if it satisfies the following properties:

1. d is continuous and left-invariant;

2. for all g_1, g_2, and g_3 with $d(g_1, g_2) + d(g_2, g_3) < \rho$, we have either

$$d(g_1, g_3) = d(g_1, g_2) + d(g_2, g_3), \text{ or } d(g_1, g_3) = |d(g_1, g_2) - d(g_2, g_3)|. \tag{6.7}$$

A pseudometric d is **monotone** if for all $g \in G$ such that $\|g\|_d < \rho/2$, we have

$$\|g^2\|_d = 2\|g\|_d.$$

To investigate the property of this notion further, we need the following fact about the adjoint representations of Lie groups [91, Proposition 9.2.21].

Fact 6.40. *Let $\mathfrak{g}$ be the Lie algebra of G, and let $\mathrm{Ad} : G \to \mathrm{Aut}(\mathfrak{g})$ be the adjoint representation. Then $\ker(\mathrm{Ad})$ is the center of G.*

The following result is the first time we need G to be a Lie group instead of just a locally compact group.

Proposition 6.41. *If d is a locally linear pseudometric on G, then d is monotone.*

Proof. We first prove an auxiliary statement.

Claim 6. Suppose $s : G \to G, g \mapsto g^2$ is the squaring map. Then there is no open $U \subseteq G$ and proper closed subgroup H of G such that $s(U) \subseteq H$.

Proof of Claim. Consider the case where G is a connected component of a linear algebraic subgroup of $\mathrm{GL}_n(\mathbb{R})$. Let J_s be the Jacobian of the function s. Then the set

$$\{g \in G : \det J_s(g) = 0\}$$

has the form $G \cap Z$ where Z is a solution set of a system of polynomial equations. It is not possible to have $G \cap Z = G$, as s is a local diffeomorphism at id_G. Hence, $G \cap Z$ must be of strictly lower dimension than G. By the inverse function theorem, $s|_{G \setminus Z}$ is open. Hence $s(U)$ is not contained in a subgroup of G with smaller dimension.

We also note a stronger conclusion for abelian Lie group: If V is an open subset of a not necessarily connected abelian Lie group A, then the image of A under $a \mapsto a^2$ is not contained in a closed subset of A with smaller dimension. Indeed, A is isomorphic as a topological group to $D \times \mathbb{T}^m \times \mathbb{R}^m$, with D a discrete group. If

$$U \subseteq D \times \mathbb{T}^m \times \mathbb{R}^m,$$

then it is easy to see that $\{a^2 : a \in V\}$ contains a subset of $D \times \mathbb{T}^m \times \mathbb{R}^m$, and is therefore not a subset of a closed subset of A with smaller dimension.

Finally, we consider the general case. Suppose to the contrary that $s(U) \subseteq H$ with H a proper closed subgroup of G. Let $Z(G)$ be the center of G, $G' = G/Z(G)$, $\pi : G \to G'$ be the quotient map, $U' = \pi(U)$, and

$$s' : G' \to G', g' \mapsto (g')^2.$$

Then U' is an open subset of G', which is isomorphic as a topological group to a connected component of an algebraic group by Fact 6.40. By the earlier case, $s'(U')$ is not contained in any proper closed subgroup of G', so we must have $\pi(H) = G'$. In particular, this implies $\dim(H \cap Z(G)) < \dim Z(G)$, and $HZ(G) = G$. Choose $h \in H$ such that $hZ(G) \cap U$ is nonempty. Then

$$s(hZ(G) \cap U) = \{h^2 a^2 : a \in Z(G) \cap h^{-1}U\}.$$

As $s(hZ(G) \cap U) \subseteq H$, we must have $\{a^2 : a \in Z(G) \cap h^{-1}U\}$ is a subset of $H \cap Z(G)$. Using the case for abelian Lie groups, this is a contradiction, because $H \cap Z(G)$ is a closed subset of $Z(G)$ with smaller dimension. $\bowtie$

We now get back to the problem of showing that d is monotone. As d is invariant, $d(\mathrm{id}_G, g) = d(g, g^2)$ for all $g \in G$. From local linearity of d, for all $g \in G$ with $\|g\|_d < \rho/2$, we either have

$$\|g^2\|_d = 2\|g\|_d \quad \text{or} \quad \|g^2\|_d = 0.$$

It suffices to rule out the possibility that $0 < \|g\|_d < \rho/4$, and $\|g^2\|_d = 0$.

As d is continuous, there is an open neighborhood W of g such that for all $g' \in W$, we have $\|g'\|_d > 0$ and $\|(g')^2\|_d = 0$. From Lemma 6.38, the set $\{g \in G : \|g\|_d = 0\}$ is a closed subgroup of G. As d is nontrivial and G is a connected Lie group, $\{g \in G : \|g\|_d = 0\}$ must be a Lie group with smaller dimension. Therefore, we only need to show that if W is an open subset of G, then $s(W)$ is not contained in a closed subgroup of G with smaller dimension, where $s : G \to G$ is the squaring map, and this is guaranteed by the earlier claim. $\square$

The next result confirms our earlier intuition: locally linear pseudometric in G will induced a homomorphism mapping to either $\mathbb{T}$ or $\mathbb{R}$.

Proposition 6.42. *Suppose d is a locally linear pseudometric with radius $\rho > 0$. Then $\ker d$ is a normal subgroup of G, $G/\ker d$ is isomorphic to $\mathbb{T}$ if G is compact, and $G/\ker d$ is isomorphic to $\mathbb{R}$ if G is noncompact.*

Proof. We first prove that $\ker d$ is a normal subgroup of G. Suppose $\|g\|_d = 0$ and $h \in G$ satisfies $\|h\|_d < \rho/4$. We have

$$\begin{aligned} d(h, hgh^{-1}) &= d(\mathrm{id}_G, gh^{-1}) \\ &\qquad = |d(\mathrm{id}_G, g) \pm d(g, gh^{-1})| = d(\mathrm{id}_G, h^{-1}) = d(\mathrm{id}_G, h). \end{aligned}$$

Hence, $d(\mathrm{id}_G, hgh^{-1}) = |d(\mathrm{id}_G, h) \pm d(h, hgh^{-1})|$ is either 0 or $2d(\mathrm{id}_G, h)$. Assume first that

$\|hgh^{-1}\|_d = 0$ for every such h when $\|g\|_d = 0$. Let

$$U := \{h : \|h\|_d < \rho/4\}.$$

By the continuity of d, U is open. Hence for every h in G, h can be written as a finite product of elements in U. By induction, we conclude that for every $h \in G$, $\|hgh^{-1}\|_d = 0$ given $\|g\|_d = 0$, and this implies that $\ker d$ is normal in G.

Suppose $\|hgh^{-1}\|_d = 2\|h\|_d$. By Proposition 6.41, d is monotone. Hence, we have

$$\|hg^2h^{-1}\|_d = 4\|h\|_d.$$

On the other hand, as $\|g\|_d = 0$, repeating the argument above, we get $\|hg^2h^{-1}\|_d$ is either 0 or $2\|h\|_d$. Hence, $\|h\|_d = 0$, and so $\|hgh^{-1}\|_d = 0$.

We now show that $G' = G/\ker d$ has dimension 1. Let d' be the pseudometric on G' induced by d. Choose $g \in G'$ in the neighborhood of $\mathrm{id}_{G'}$ such that g is in the image of the exponential map and $\|g\|_{d'} < \rho/4$. If g' is another element in the neighborhood of $\mathrm{id}_{G'}$ which is in the image of the exponential map and $\|g'\|_{d'} < \rho/4$. Without loss of generality, we may assume $\|g'\|_{d'} \leq \|g\|_{d'}$. Suppose $g' = \exp(X)$. Then by the monotonicity, there is $k \geq 1$ such that $\|(g')^k\|_{d'} \geq \|g\|_{d'}$. By the continuity of the exponential map, there is $t \in (0, 1]$ such that

$$\|g\|_{d'} = \|\exp(tkX)\|_{d'}.$$

This implies that g and g' are on the same one parameter subgroup, which is the desired conclusion. $\square$

6.7.2 Almost linear pseudometrics: relative sign and total weight functions

In this section, we will introduce a weakening of the notion of a locally linear pseudometric and define the relative sign function and total weight function associate to it. When d is a pseudometric arising from a measurable subset A as in Proposition 6.37, these roughly give the "direction" and the "distance" that an element of the group translates A.

Throughout this section, d is a pseudometric on G with radius $\rho > 0$, and γ is a constant with $0 < \gamma < 10^{-8}\rho$. For a constant λ, we write $I(\lambda)$ for the interval $(-\lambda, \lambda)$ in either $\mathbb{R}$ or $\mathbb{T}$, and we write $N(\lambda)$ for $\{g \in G : \|g\|_d \in I(\lambda)\}$. By Fact 6.4(vii), $N(\lambda)$ is an open set, and hence measurable. We say that d is **γ-linear** if it satisfies the following conditions:

1. d is continuous and left-invariant;

2. for all $g_1, g_2, g_3 \in G$ with $d(g_1, g_2) + d(g_2, g_3) < \rho - \gamma$, we have either

$$d(g_1, g_3) \in d(g_1, g_2) + d(g_2, g_3) + I(\gamma),$$

   or

$$d(g_1, g_3) \in |d(g_1, g_2) - d(g_2, g_3)| + I(\gamma).$$

Given $\alpha \leq \rho$, let $N(\alpha) = \{g \in G : \|g\|_d \leq \alpha\}$. We say that d is **γ-monotone** if for all $g \in N(\rho/2 - \gamma)$, we have

$$\|g^2\|_d \in 2\|g\|_d + I(\gamma).$$

The next lemma says that under the γ-linearity condition, the group G essentially has only one "direction": if there are three elements have the same distance to id_G, then at least two of them are very close to each other.

Lemma 6.43. *Suppose d is a γ-linear pseudometric on G. If $g, g_1, g_2 \in G$ such that*

$$\|g\|_d = \|g_1\|_d = \|g_2\|_d \in I(\rho/4 - \gamma) \setminus I(2\gamma),$$

and $d(g_1, g_2) \in 2\|g\|_d + I(\gamma)$. Then either $d(g, g_1) \in I(\gamma)$ or $d(g, g_2) \in I(\gamma)$.

Proof. Suppose both $d(g, g_1)$ and $d(g, g_2)$ are not in $I(\gamma)$. By γ-linearity of d, we have

$$d(g, g_1) \in |d(\mathrm{id}_G, g) \pm d(\mathrm{id}_G, g_1)| + I(\gamma),$$

and so $d(g, g_1) \in 2\|g\|_d + I(\gamma)$. Similarly, we have $d(g, g_2) 2\|g\|_d + I(\gamma)$.

Suppose first that $d(g_1, g_2) \in d(g, g_1) + d(g, g_2) + I(\gamma)$, then

$$d(g_1, g_2) \in 4\|g\|_d + I(3\gamma).$$

On the other hand, by γ-linearity we have $d(g_1, g_2) \leq 2\|g\|_d + \gamma$. Hence, we have $\|g\|_d \in I(2\gamma)$, a contradiction.

The other two possibilities are $d(g_1, g_2) + d(g, g_2) \in d(g, g_1) + I(\gamma)$ or $d(g_1, g_2) + d(g, g_1) \in d(g, g_2) + I(\gamma)$, but similar calculations also lead to contradictions. $\square$

Proposition 6.44 below is a partial replacement for Proposition 6.41 for linear pseudometric. The fact that we do not automatically have monotonicity is a reason that the later Section 8.3 is much harder than Section 8.2.

Proposition 6.44 (Path monotonicity implies global monotonicity)**.** *Let $\mathfrak{g}$ be the Lie algebra of G,* $\exp : \mathfrak{g} \to G$ *the exponential map, and d a γ-linear pseudometric on G. Suppose for each X in $\mathfrak{g}$, we have one of the following two possibilities:*

(i) $\|\exp(tX)\|_d < \gamma$ *for all $t \in \mathbb{R}$;*

(ii) *there is* $t_0 \in \mathbb{R}^{>0}$ *with* $\|\exp(t_0X)\|_d \in I(\rho/2-\gamma) \setminus I(\rho/4)$,

$$\|\exp(2t_0X)\|_d = 2\|\exp(t_0X)\|_d + I(\gamma), \tag{6.8}$$

and

$$\|\exp(tX)\|_d + \|\exp((t_0-t)X)\|_d \in \|\exp(t_0X)\|_d + I(\gamma) \tag{6.9}$$

for all $t \in [0, t_0]$.

Then d *is* (9γ)*-monotone.*

Proof. Fix an element g of G with $\|g\|_d \in I(\rho/2 - 16\gamma)$. Our job is to show that $\|g^2\|_d \in 2\|g\|_d + I(9\gamma)$. Since G is compact and connected, the exponential map exp is surjective. We get $X \in \mathfrak{g}$ such that $g \in \{\exp(tX) : t \in \mathbb{R}\}$. If we are in scenario (i), then $\|g\|_d < \gamma$, hence $\|g^2\|_d \in 2\|g\|_d + I(3\gamma)$. Therefore, it remains to deal with the case where we have an t_0 as in (ii).

Set $g_0 = \exp(t_0X)$. We consider first the special case where $\|g\|_d < \|g_0\|_d - 2\gamma$. As d is continuous, there is $t_1 \in [0, t_0]$ such that with $g_1 = \exp(t_1X)$, we have $\|g_1\|_d = \|g\|_d$. Let $t_2 = -t_1$, and $g_2 = \exp(t_2X) = g_1^{-1}$. Since d is invariant,

$$\|g_2\|_d = d(g_1^{-1}, \mathrm{id}_G) = d(\mathrm{id}_G, g_1) = \|g_1\|_d.$$

Hence, $\|g_1\|_d = \|g_2\|_d = \|g\|_d$. If $\|g\|_d < 2\gamma$, then $\|g^2\|_d \in 2\|g\|_d + I(5\gamma)$ and we are done. Thus we suppose $\|g\|_d \geq 2\gamma$. Then, by Lemma 6.43, either $d(g, g_1) < \gamma$, or $d(g, g_2) < \gamma$.

Since these two cases are similar, we assume that $d(g, g_1) < \gamma$. By γ-linearity, $\|g_1^2\|_d$ is in either $2\|g_1\|_d + I(\gamma)$ or $I(\gamma)$. Using $\|g_0^2\|_d \in 2\|g_0\|_d + I(\gamma)$ and the assumption that $\|g\|_d < \|g_0\|_d - 2\gamma$, in either case, we have

$$\|g_1^2\|_d < \|g_0^2\|_d - 2\gamma. \tag{6.10}$$

Since $g_0^{-1}g_1 = g_1g_0^{-1}$, and by γ-linearity of d, we get

$$d(g_1^2, g_0^2) = d(\mathrm{id}_G, g_1^{-2}g_0^2) = d(\mathrm{id}_G, (g_1^{-1}g_0)^2) \in \{0, 2d(g_1, g_0)\} + I(\gamma). \tag{6.11}$$

By (6.9), we have $\|g_1\|_d + d(g_1, g_0) \in \|g_0\|_d + I(\gamma)$. Recalling that $\|g_1\|_d = \|g\|_d > 2\gamma$, and from (6.8) and (6.11), we have

$$d(g_1^2, g_0^2) < 2\|g_0\|_d - 3\gamma = \|g_0^2\| - 2\gamma. \tag{6.12}$$

By (6.10), (6.12), and the γ-linearity of d, we have

$$\|g_1^2\|_d \in \|g_0^2\|_d - d(g_1^2, g_0^2) + I(\gamma).$$

Therefore by (6.9) and (6.11), we have either

$$\|g_1^2\|_d \in 2\|g_1\|_d + I(5\gamma) \quad \text{or} \quad \|g_1^2\|_d \in 2\|g_0\|_d + I(3\gamma).$$

As $\|g_1\|_d^2 \le 2\|g_1\| + \gamma < 2\|g_0\| - 5\gamma$, we must have $\|g_1^2\| \in 2\|g_1\| + I(5\gamma)$. Now, since $\|g_1^{-1}g\|_d = d(g_1, g) < \gamma$, again by the γ-linearity we conclude that

$$d(g_1^2, g^2) = \|(g_1^{-1}g)^2\|_d < 3\gamma.$$

Thus, $\|g^2\|_d \in 2\|g\|_d + I(9\gamma)$.

Finally, we consider the other special case where $\|g_0\|_d + 2\gamma < \|g\|_d < \rho/2 - 16\gamma$. For $g_1 = \exp(t_1X)$ with $t_1 \in [0, t_0]$, we have $\|g_1^2\|_d \in 2\|g_1\| + I(8\gamma)$ by a similar argument as above. Using continuity, we can choose t_1 such that $\|g_1^2\|_d = \|g\|_d$, and let $g_2 = g_1^{-1}$. The argument goes in exactly the same way with the role of g_1 replaced by g_1^2 and the role of g_2 replaced by g_2^2. $\square$

Suppose d is γ-linear. We define $s(g_1, g_2)$ to be the **relative sign** for $g_1, g_2 \in G$ satisfying $\|g_1\|_d + \|g_2\|_d < \rho - \gamma$ by

$$s(g_1, g_2) = \begin{cases} 0 & \text{if } \min\{\|g_1\|_d, \|g_2\|_d\} \le 4\gamma, \\ 1 & \text{if } \min\{\|g_1\|_d, \|g_2\|_d\} > 4\gamma \text{ and } \|g_1 g_2\|_d \in \|g_1\|_d + \|g_2\|_d + I(\gamma). \\ -1 & \text{if } \min\{\|g_1\|_d, \|g_2\|_d\} > 4\gamma \text{ and } |g_1 g_2|_d \in \big|\|g_1\|_d - \|g_2\|_d\big| + I(\gamma). \end{cases}$$

Note that this is well defined because when $\min\{\|g_1\|_d, \|g_2\|_d\} \ge 4\gamma$ in the above definition, the differences between $\big|\|g_1\|_d - \|g_2\|_d\big|$ and $\|g_1\|_d + \|g_2\|_d$ is at least 6γ. The following lemma gives us tools to relate signs between different elements.

Proposition 6.45. *Suppose d is γ-linear and γ-monotone. Then for g_1, g_2, and g_3 in $N(\rho/4 - \gamma) \setminus N(4\gamma)$, we have the following*

(i) $s(g_1, g_1^{-1}) = -1$ *and* $s(g_1, g_1) = 1$.

(ii) $s(g_1, g_2) = s(g_2, g_1)$.

(iii) $s(g_1, g_2) = s(g_1^{-1}, g_2^{-1}) = -s(g_1^{-1}, g_2) = -s(g_1, g_2^{-1})$.

(iv) $s(g_1, g_2)s(g_2, g_3)s(g_3, g_1) = 1$.

(v) *If* $\|g_1\|_d \le \|g_2\|_d$, *and* $g_1 g_2$ *is in* $N(\rho/4 - \gamma) \setminus N(4\gamma)$, *then*

$$s(g_0, g_1 g_2) = s(g_0, g_2 g_1) = s(g_0, g_2).$$

Proof. As g_1, g_2, and g_3 are in $N(\rho/4-\gamma) \setminus N(4\gamma)$, one has $s(g_i, g_j) \neq 0$ for all $i, j \in \{1, 2, 3\}$. The first part of (i) is immediate from the fact that $\|\mathrm{id}_G\|_d = 0$, and the second part of (i) follows from the γ-monotonicity and the definition of the relative sign.

We now prove (ii). Suppose to the contrary that $s(g_1, g_2) = -s(g_2, g_1)$. Without loss of generality, assume $s(g_1, g_2) = 1$. Then $\|g_1 g_2 g_1 g_2\|_d$ is in $2\|g_1 g_2\|_d + I(\gamma)$, which is a subset of $2\|g_1\|_d + 2\|g_2\|_d + I(3\gamma)$. On the other hand, as $s(g_2, g_1) = -1$, we have

$$\|g_1 g_2 g_1 g_2\|_d \in \big|\|g_1\|_d \pm (\|g_2\|_d - \|g_1\|_d) \pm \|g_2\|_d\big| + I(3\gamma).$$

This contradicts the assumption that g_1 and g_2 are not in $N(4\gamma)$.

Next, we prove the first and third equality in (iii). Note that $\|g\|_d = \|g^{-1}\|_d$ for all $g \in G$ as d is symmetric and invariant. Hence, $\|g_1 g_2\|_d = \|g_2^{-1} g_1^{-1}\|_d$. This implies that $s(g_1, g_2) = s(g_2^{-1}, g_1^{-1})$. Combining with (ii), we get the first equality in (iii). The third equality in (iii) is a consequence of the first equality in (iii).

Now, consider the second equality in (iii). Suppose $s(g_1^{-1}, g_2^{-1}) = s(g_1^{-1}, g_2)$. Then from (ii) and the first equality of (iii), we get $s(g_2, g_1) = s(g_1^{-1}, g_2)$. Hence, either

$$\|g_2 g_1 g_1^{-1} g_2\|_d \in 2\left(\|g_1\|_d + \|g_2\|_d\right) + I(3\gamma)$$

or

$$\|g_2 g_1 g_1^{-1} g_2\|_d \in 2\big|\|g_1\|_d - \|g_2\|_d\big| + I(3\gamma).$$

On the other hand, $\|g_2 g_1 g_1^{-1} g_2\|_d = \|g_2^2\|_d$, which is in $2\|g_2\|_d + I(\gamma)$. We get a contradiction with the fact that g_1 and g_2 are not in $N(4\gamma)$.

We now prove (iv). Without loss of generality, assume $\|g_1\|_d \leq \|g_2\|_d \leq \|g_3\|_d$. Using (iii) to replace g_3 with g_3^{-1} if necessary, we can further assume that $s(g_2, g_3) = 1$. We need to show that $s(g_1, g_2) = s(g_1, g_3)$. Suppose to the contrary. Then from (iii), we get $s(g_1, g_2) = s(g_1^{-1}, g_3)$. Using (iii) to replacing g_1 with g_1^{-1} if necessary, we can assume that $s(g_1, g_2) = s(g_1^{-1}, g_3) = 1$. Using (ii), we get $s(g_2, g_1) = 1$. Hence, either

$$\|g_2 g_1 g_1^{-1} g_3\|_d \in 2\|g_1\|_d + \|g_2\|_d + \|g_3\|_d + I(3\gamma)$$

or

$$\|g_2 g_1 g_1^{-1} g_3\|_d \in \|g_3\|_d - \|g_2\|_d + I(3\gamma).$$

On the other hand, $\|g_2 g_1 g_1^{-1} g_3\|_d = \|g_2 g_3\|_d$ is in $\|g_2\|_d + \|g_3\|_d + I(\gamma)$. Hence, we get a contradiction to the fact that g_1, g_2, and g_3 are not in $N(4\gamma)$.

Finally, we prove (v). Using (iv), it suffices to show $s(g_1g_2, g_2) = s(g_2g_1, g_2) = 1$. We will only show the former, as the proof for the latter is similar. Suppose to the contrary that $s(g_1g_2, g_2) = -1$. Then $\|g_1 g_2^2\|_d$ is in $\big|\|g_1g_2\|_d - \|g_2\|_d\big| + I(\gamma)$, which is a subset of $\|g_1\|_d + I(2\gamma)$. On the other hand, $\|g_1g_2^2\|_d$ is also in $\big|\|g_1\|_d - \|g_2^2\|_d\big| + I(\gamma)$ which is a subset of $2\|g_2\|_d - \|g_1\|_d + I(2\gamma)$. Hence, we get a contradiction with the assumption that g_1 and g_2 are not in $N(4\gamma)$. $\square$

The notion of relative sign corrects the ambiguity in calculating distance, as can be seen in the next result.

Lemma 6.46. *Suppose d is γ-monotone γ-linear, and g_1 and g_2 are in $N(\rho/16 - \gamma)$ with $\|g_1\|_d \le \|g_2\|_d$. Then we have the following*

(i) *Both $\|g_1g_2\|_d$ and $\|g_2g_1\|_d$ are in $s(g_1, g_2)\|g_1\|_d + \|g_2\|_d + I(5\gamma)$.*

(ii) *If g_0 is in $N(\rho/4) \setminus N(4\gamma)$, then both $s(g_0, g_1g_2)\|g_1g_2\|_d$ and $s(g_0, g_2g_1)\|g_2g_1\|_d$ are in*

$$s(g_0, g_1)\|g_1\|_d + s(g_0, g_2)\|g_2\|_d + I(25\gamma).$$

Proof. We first prove (i). When $g_1, g_2 \notin N(4\gamma)$, the statement for $\|g_1g_2\|_d$ is immediate from the definition of the relative sign, and the statement for $\|g_2g_1\|_d$ is a consequence of Proposition 6.45(ii). Now suppose $\|g_1\|_d < 4\gamma$. From the γ-linearity, we have

$$\|g_2\|_d - \|g_1\|_d - \gamma < \|g_1g_2\|_d < \|g_1\|_d + \|g_2\|_d + \gamma.$$

We deal with the case where $\|g_2\|_d < 4\gamma$ similarly.

We now prove (ii). Fix g_0 in $N(\rho/4-\gamma)\setminus N(4\gamma)$. We will consider two cases, when g_1 is not in $N(4\gamma)$ and when g_1 is in $N(4\gamma)$. Suppose we are in the first case, that is $g_1 \notin N(4\gamma)$. As $\|g_1\|_d \le \|g_2\|_d$, we also have $g_2 \notin N(4\gamma)$. If both g_1g_2 and g_2g_1 are not in $N(4\gamma)$, then the desired conclusion is a consequence of (i) and Proposition 6.45(iv, v). Within the first case, it remains to deal with the situations where g_1g_2 is in $N(4\gamma)$ or g_2g_1 is in $N(4\gamma)$.

Since these two situations are similar, we may assume g_1g_2 is in $N(4\gamma)$. From (i), we have $s(g_1,g_2)=-1$ and $\|g_2\|_d-\|g_1\|_d$ is at most 5γ. Therefore, $\|g_2g_1\|_d$ is in $I(6\gamma)$. By Proposition 6.45(iv), we have $s(g_0,g_1)=-s(g_0,g_2)$, and so

$$s(g_0,g_1)\|g_1\|_d+s(g_0,g_2)\|g_2\|_d \in I(6\gamma).$$

Since both $s(g_0,g_1g_2)\|g_1g_2\|_d$ and $s(g_0,g_2g_1)\|g_1g_2\|_d$ are in $I(6\gamma)$, they are both in

$$s(g_0,g_1)\|g_1\|_d+s(g_0,g_2)\|g_2\|_d+I(12\gamma)$$

giving us the desired conclusion.

Continuing from the previous paragraph, we consider the second case when g_1 is in $N(4\gamma)$. If g_2 is in $N(16\gamma)$, then both $\|g_1g_2\|_d$ and $\|g_2g_1\|_d$ are in $I(25\gamma)$ by (i), and the desired conclusion follows. Now suppose g_2 is not in $N(16\gamma)$. Then from (i) and the fact that $g_1\in N(4\gamma)$, we get g_1g_2 and g_2g_1 are both not in $N(4\gamma)$. Note that $s(g_1g_2,g_2^{-1})=-1$, because otherwise we get

$$\|g_1\|_d \ge \|g_1g_2\|_d+\|g_2^{-1}\|_d-5\gamma > 4\gamma.$$

A similar argument gives $s(g_2^{-1},g_2g_1)=-1$. Hence, $s(g_1g_2,g_2)=s(g_2g_1,g_2)=1$. By Proposition 6.45(v), we get

$$s(g_0,g_2)=s(g_0,g_1g_2)=s(g_0,g_2g_1).$$

From (i), $\|g_1g_2\|_d$ and $\|g_2g_1\|_d$ are both in $\|g_2\|_d + I(9\gamma)$. On the other hand, as $s(g_0, g_1) = 0$, we have $s(g_0, g_1)\|g_1\|_d + s(g_0, g_2)\|g_2\|_d = s(g_0, g_2)\|g_2\|_d$. The desired conclusion follows. $\square$

The next corollary will be important in the subsequent development.

Corollary 6.47. *Suppose d is γ-linear and γ-monotone, g_0 and g_0' are elements in $N(\rho/4 - \gamma) \setminus N(4\gamma)$, and $(g_1, \dots, g_n)$ is a sequence with $g_i \in N(\rho/4-\gamma) \setminus N(4\gamma)$ for $i \in \{1, \dots, n\}$. Then*

$$\left|\sum_{i=1}^{n} s(g_0, g_i)\|g_i\|_d\right| = \left|\sum_{i=1}^{n} s(g_0', g_i)\|g_i\|_d\right|.$$

Proof. As $s(g_0, g_i) = s(g_0', g_i) = 0$ whenever $\|g_i\|_d < 4\gamma$, we can reduce to the case where $\min_{1\le i\le n} \|g_i\|_d \ge 4\gamma$. Using Proposition 6.45(iii) to replace g_0 with g_0^{-1} if necessary, we can assume that $s(g_0, g_1) = s(g_0', g_1)$. Then by Proposition 6.45(iii), $s(g_0, g_i) = s(g_0', g_i)$ for all $i \in \{1, \dots, n\}$. This gives us the desired conclusion. $\square$

The following auxiliary lemma allows us to choose g_0 as in Corollary 6.47.

Lemma 6.48. *The set $N(\rho/4 - \gamma) \setminus N(4\gamma)$ is not empty.*

Proof. It suffices to show that $\mu_G(N(4\gamma)) < \mu_G(N(\rho/4-\gamma)$. Since id_G is in $N(4\gamma)$, $N(4\gamma)$ is a nonempty open set and has $\mu_G(N(4\gamma)) > 0$. Therefore, $N^2(4\gamma)$ and $N^4(4\gamma)$ are also open. By γ-linearity, we have

$$N^2(4\gamma) \subseteq N_{9\gamma} \quad \text{and} \quad N^4(4\gamma) \subseteq N_{19\gamma}.$$

As $19\gamma < \rho$, we have $N^4(4\gamma) \neq G$. Using Proposition 6.17, we get

$$\mu_G(N^2(4\gamma)) \le 2/3 \quad \text{and} \quad \mu_G(N(4\gamma)) < 1/3.$$

Hence, by Kemperman's inequality $\mu_G(N(4\gamma)) < \mu_G(N^2(4\gamma)) \le \mu_G(N(\rho/4-\gamma))$, which is the desired conclusion. $\square$

Suppose $(g_1, \ldots, g_n)$ is a sequence of elements in $N(\rho/4 - \gamma) \setminus N(4\gamma)$. We set

$$t(g_1, \ldots, g_n) = \left| \sum_{i=1}^{n} s(g_0, g_i) \|g_i\|_d \right|$$

with g_0 is an arbitrary element in $N(\rho/4 - \gamma) \setminus N(4\gamma)$, and call this the **total weight** associated to $(g_1, \ldots, g_n)$. This is well defined by Corollary 6.47 and Lemma 6.48.

6.7.3 Almost linear pseudometrics: group homomorphisms into tori

In this section, we will use the relative sign function and the total weight function defined in Section 7.2 to define a universally measurable multivalued group homomorphism into $\mathbb{T}$. We will then use a number or results in descriptive set theory and geometry to refine this into a continuous group homomorphism.

We keep the setting of Section 7.2, and assume further that G is compact. Let s and t be the relative sign function and the total weight function defined earlier. Set $\lambda = \rho/36$, and $N[\lambda] = \{g \in G : \|g\|_d \leq \lambda\}$. The set $N[\lambda]$ is compact, and hence measurable. Moreover, Lemma 6.46 is applicable when g_0 is an arbitrary element in $N(\rho/4 - \gamma) \setminus N(4\gamma)$, and g_1 are g_2 are in $N[\lambda]$.

A sequence $(g_1, \ldots, g_n)$ of elements in G is a **λ-sequence** if g_i is in $N[\lambda]$ for all $i \in \{1, \ldots, n\}$. We are interested in expressing an arbitrary g of G as a product of a λ-sequence where all components are "in the same direction". The following notion captures that idea. A λ-sequence $(g_1, \ldots, g_n)$ is **irreducible** if for all $2 \leq j \leq 4$, we have

$$g_{i+1} \cdots g_{i+j} \notin N(\lambda).$$

A **concatenation** of a λ-sequence $(g_1, \ldots, g_n)$ is a λ-sequence $(h_1, \ldots, h_m)$ such that there are $0 = k_0 < k_1 < \cdots < k_m = n$ with

$$h_i = g_{k_{i-1}+1} \cdots g_{k_i} \text{ for } i \in \{1, \ldots, m\}.$$

The next lemma allows us to reduce an arbitrary sequence to irreducible λ-sequences via concatenation.

Lemma 6.49. *Suppose d is γ-linear and γ-monotone, and $(g_1, \ldots, g_n)$ is a λ-sequence. Then $(g_1, \ldots, g_n)$ has an irreducible concatenation $(g'_1, \ldots, g'_m)$ with*

$$t(g'_1, \ldots, g'_m) \in t(g_1, \ldots, g_n) + I(25(n-m)\gamma).$$

Proof. The statement is immediate when $n = 1$. Using induction, suppose we have proven the statement for all smaller values of n. If $(g_1, \ldots, g_n)$ is irreducible, we are done. Consider the case where $g_{i+1}g_{i+2}$ is in $N(\lambda)$ for some $0 \leq i \leq n-2$. Fix g_0 in $N(\lambda/4 - \gamma) \setminus N(4\gamma)$. Using Lemma 6.46(ii)

$$s(g_0, g_{i+1}g_{i+2})\|g_{i+1}g_{i+2}\|_d \in s(g_0, g_{i+1})\|g_{i+1}\|_d + s(g_0, g_{i+2})\|g_{i+2}\|_d + I(25\gamma).$$

From here, we get the desired conclusion. The cases where either $g_{i+1}g_{i+2}g_{i+3}$ for some $0 \leq i \leq n-3$ or $g_{i+1}g_{i+2}g_{i+3}g_{i+4}$ is in $N(\lambda)$ for some $0 \leq i \leq n-4$ can be dealt with similarly. $\square$

The following lemma makes the earlier intuition of "in the same direction" precise:

Lemma 6.50. *Suppose d is γ-linear and γ-monotone, g_0 is in $N(\rho/4 - \gamma) \setminus N(4\gamma)$, and $(g_1, \ldots, g_n)$ is an irreducible λ-sequence. Then for all i, i', j, and j' such that $2 \leq j, j' \leq 4$, $0 \leq i \leq n-j$, and $0 \leq i' \leq n-j'$, we have*

$$s(g_0, g_{i+1} \cdots g_{i+j}) = s(g_0, g_{i'+1} \cdots g_{i'+j'}).$$

Proof. It suffices to show for fixed i, j with $0 \leq i \leq n-j-1$ and $2 \leq j \leq 3$ that

$$s(g_0, g_{i+1} \cdots g_{i+j}) = s(g_0, g_{i+1} \cdots g_{i+j+1}).$$

Note that both $g_{i+1} \cdots g_{i+j}$ and $g_{i+1} \cdots g_{i+j+1}$ are in $N(\rho/4 - \gamma) \setminus N(4\gamma)$. Hence, applying Proposition 6.45(iv), we reduce the problem to showing

$$s(g_{i+j}^{-1} \cdots g_{i+1}^{-1}, g_{i+1} \cdots g_{i+j+1)}) = -1.$$

This is the case because otherwise, $\|g_{i+j+1}\|_d \geq 2\lambda - \gamma > \lambda$, a contradiction. $\square$

We now get a lower bound for the total distance of an irreducible λ-sequence:

Corollary 6.51. *Suppose d is γ-linear and γ-monotone, and $(g_1, \ldots, g_n)$ is an irreducible λ-sequence. Then*

$$t(g_1, \ldots, g_n) > n\lambda/4.$$

Proof. If $n = 2k$, let $h_i = g_{2i-1}g_{2i}$ for $i \in \{1, \ldots, k\}$. If $n = 2k+1$, let $h_i = g_{2i-1}g_{2i}$ for $i \in \{1, \ldots, k-1\}$, and $h_k = g_{2n-1}g_{2n}g_{2n+1}$. From Lemma 6.46, we have

$$t(h_1, \ldots, h_k) \in t(g_1, \ldots, g_n) + I(25(n-k)\gamma). \tag{6.13}$$

As $(g_1, \ldots, g_n)$ is irreducible, h_i is in $N(3\lambda) \setminus N(\lambda)$ for $i \in \{1, \ldots, k\}$. By Lemma 6.50, we get $s(g_0, h_i) = s(g_0, h_j)$ for all i and j in $i \in \{1, \ldots, k\}$. Thus by the definition of the total weight again, $t(h_1, \ldots, h_k) > n\lambda/3$. Combining with the assumption on λ and (6.13), we get $t(g_1, \ldots, g_n) > n\lambda/3 - 11n\gamma > n\lambda/4$. $\square$

When $(g_1 \ldots, g_n)$ is an irreducible λ-sequence, $g_1 \cdots g_m$ is intuitively closer to g_0 than $g_1 \cdots g_{m+k}$ for some positive k. However, as G is compact, the sequence may "return back" to id_G when n is large. The next proposition provides a lower bound estimate on such n.

Lemma 6.52 (Monitor lemma)**.** *Suppose d is γ-linear and γ-monotone, and $(g_1, \dots, g_n)$ is an irreducible λ-sequence with $g_1 \cdots g_n = \mathrm{id}_G$. Then $n \geq 1/\mu_G(N(4\lambda))$.*

Proof. Let $m > 0$. For convenience, when $m > n$ we write g_m to denote the element g_i with $i \leq n$ and $i \equiv m \pmod{n}$. Define

$$N^{(m)}(4\lambda) = \{g \in G \mid d(g, g_1 \cdots g_m) < 4\lambda\}.$$

Note that we have $N^{(m)}(4\lambda) = N^{(m')}(4\lambda)$ when $m \equiv m' \pmod{n}$. By invariance of d and μ_G, clearly $\mu_G(N^{(m)}(4\lambda)) = \mu_G(N(4\lambda))$ for all m. We also write $N^{(0)}(4\lambda) = N(4\lambda)$. We will show that

$$G = \bigcup_{m \in \mathbb{Z}} N^{(m)}(4\lambda) = \bigcup_{m=0}^{n-1} N^{(m)}(4\lambda),$$

which yields the desired conclusion.

As $g_1 \cdots g_n = \mathrm{id}_G$, we have id_G is in $N^{(0)}(2\lambda)$, and hence in $\bigcup_{m \in \mathbb{Z}} N^{(m)}(4\lambda)$. As every elements in G can be written as a product of finitely many elements in $N(\lambda)$, it suffices to show for every $g \in \bigcup_{m \in \mathbb{Z}} N^{(m)}(4\lambda)$ and $g' = gh$ with $h \in N(\lambda)$ that g' is in $\bigcup_{m \in \mathbb{Z}} N^{(m)}(4\lambda)$. The desired conclusion then follows from the induction on the number of translations in $N(\lambda)$.

Fix m which minimizes $d(g, g_1 \dots g_m)$. We claim that $d(g, g_1 \dots g_m) < 2\lambda + \gamma$. This claim gives us the desired conclusion because we then have $d(g', g_1 \dots g_m) < 3\lambda + 2\gamma < 4\lambda$ by the γ-linearity of d.

We now prove the claim that $d(g, g_1 \dots g_m) < 2\lambda + \gamma$. Suppose to the contrary that $d(g, g_1 \dots g_m) \geq 2\lambda + \gamma$. Let $u = (g_1 \cdots g_m)^{-1} g$. Now by Lemma 6.50 we have either $s(u, g_{m+1} g_{m+2}) = 1$, or $s(u, g_m^{-1} g_{m-1}^{-1}) = 1$. Suppose it is the former, since the latter case can be proved similarly. Then $s(u, g_{m+2}^{-1} g_{m+1}^{-1}) = -1$. Note that $g = g_1 \cdots g_m u = (g_1 \cdots g_{m+2}) g_{m+2}^{-1} g_{m+1}^{-1} u$. By the definition of u, and the linearity of d, we have $\|u\|_d \geq$

$2\lambda+\gamma > \|g_{m+1}g_{m+2}\|_d$, therefore by the irreducibility we have

$$\begin{aligned} d(g, g_1\cdots g_{m+2}) &= \|g_{m+2}^{-1}g_{m+1}^{-1}u\|_d \\ &< \|u\|_d - \|g_{m+2}^{-1}g_{m+1}^{-1}\|_d + \gamma < \|u\|_d - \lambda + \gamma < \|u\|_d. \end{aligned}$$

This contradicts our choice of m having $d(g, g_1, \dots, g_m)$ minimized. $\square$

In the later proofs of this section, we will fix an irreducible λ sequence $g_1\cdots g_n = \mathrm{id}_G$ to serve as "monitors". As each element of G will be captured by one of the monitors, this will help us to bound the error terms in the final almost homomorphism we obtained from the pseudometric.

Suppose d is γ-linear and γ-monotone, and $N(\rho/4-\gamma)\setminus N(4\gamma) \neq \varnothing$. Define the **returning weight** of d to be

$$\omega = \inf\{t(g_1, \dots, g_n) : (g_1, \dots, g_n) \text{ is an irreducible } \lambda\text{-sequence with } g_1\cdots g_n = \mathrm{id}_G\}.$$

The following corollary translate Lemma 6.52 to a bound on such ω:

Corollary 6.53. *Suppose d is γ-linear and γ-monotone, and ω is the returning weight of d. Then we have the following:*

(i) $\lambda/4\mu_G(N(4\lambda)) \leq \omega \leq 4\lambda/\mu_G(N(\lambda))$.

(ii) *There is an irreducible λ-sequence $(g_1, \dots, g_n)$ such that $\omega = t(g_1, \dots, g_n)$ and*

$$1/\mu_G(N(4\lambda)) \leq n \leq 4/\mu_G(N(\lambda)).$$

Proof. Note that each irreducible λ-sequence $(g_1, \dots, g_n)$ has $n \geq 1/\mu_G(N(4\lambda))$ by using Lemma 6.52. Hence, by Corollary 6.51, we get $\omega \geq \lambda/4\mu_G(N(4\lambda))$. On the other hand, by Proposition 6.17, $G = (N(\lambda))^k$ for all $k > 1/\mu_G(N(\lambda))$. Hence, with Lemma 6.49, there is an irreducible λ-sequence $(g_1, \dots, g_n)$ with $g_1 \cdots g_n = \mathrm{id}_G$ and $n \leq 4/\mu_G(N(\lambda))$. From the definition of t, we get $\omega \leq 4\lambda/\mu_G(N(\lambda))$.

Now if an irreducible λ-sequence $(g_1, \dots, g_n)$ has $n > 4/\mu_G(N(\lambda))$, then by (i) and Corollary 6.51,

$$t(g_1, \dots, g_n) > \frac{4\lambda}{\mu_G(N(\lambda))} \geq \omega,$$

a contradiction. Therefore, we have

$$\begin{aligned} \omega = \inf\{t(g_1, \dots, g_n) : (g_1, \dots g_n) &\text{ is an irreducible } \lambda\text{-sequence with} \\ & n \leq 4/\mu_G(N(\lambda)) \text{ and } g_1 \cdots g_n = \mathrm{id}_G\}. \end{aligned}$$

For fixed n the set of irreducible λ-sequence of length n is closed under taking limit. Hence, we obtain desired $(g_1, \dots, g_n)$ using the Bozalno–Wierstrass Theorem. ☐

The next lemma allows us to convert between $\mu_G(N(\lambda))$ and $\mu_G(N(4\lambda))$:

Lemma 6.54. *Suppose d is γ-linear and γ-monotone. Then*

$$\mu_G(N(4\lambda)) \leq 16\mu_G(N(\lambda)).$$

Proof. Fix $h \in N(\lambda) \setminus N(\lambda/2 - \gamma)$. Such h exists since by γ-monotonicity we have $N^2(\lambda/2 - \gamma) \subseteq N(\lambda)$, and by Kemperman's inequality, $\mu_G(N(\lambda) > 2\mu_G(N(\lambda/2 - \gamma))$. Let g be an arbitrary element in $N(4\lambda)$, and assume first $s(g, h) = 1$. Let $k \geq 0$ be an integer, and define $g_k = g(h^{-1})^k$. Then by Lemma 6.45 and Lemma 6.46,

$$\|g_k\|_d \in \|g\|_d - k\|h\|_d + I(5k\gamma) \text{ for } k < \|g\|_d/\|h\|_d.$$

Hence, there is $k < 8$ such that $g_k \in N(\lambda)$. When $s(g, h) = -1$, one can similarly construct

g'_k as gh^k, and find $k < 8$ such that $g'_k \in N(\lambda)$. Therefore

$$N(4\lambda) \subseteq \Big(\bigcup_{i=0}^{7} N(\lambda)h^i\Big) \cup \Big(\bigcup_{j=0}^{7} N(\lambda)h^{-j}\Big).$$

Thus, $\mu_G(N(4\lambda)) \leq 16\mu_G(N(\lambda))$. $\square$

The following proposition implicitly establish that t defines an approximate multivalue group homomorphism from G to $\mathbb{R}/\omega\mathbb{Z}$.

Proposition 6.55. *Suppose d is γ-linear and γ-monotone, ω is the returning weight of d, and $(g_1, \ldots, g_n)$ is a λ-sequence with $g_1 \ldots g_n = \mathrm{id}_G$ and $n \leq 4/\mu_G(N(\lambda))$. Then*

$$t(g_1, \ldots, g_n) \in \omega\mathbb{Z} + I(\omega/400).$$

Proof. Let g_0 be in $N(\rho/4 - \gamma) \setminus N(4\gamma)$. Using Proposition 6.45(iii) to replace g_0 with g_0^{-1} if necessary, we can assume that

$$t(g_1, \ldots, g_n) = \sum_{i=1}^{n} s(g_0, g_i)\|g_i\|_d.$$

As $n \leq 4/\mu_G(N(\lambda))$, we have $t(g_1, \ldots, g_n) \leq 4\lambda/\mu_G(N(\lambda))$. From Corollary 6.53(i), we have $\lambda \leq 4\omega\mu_G(N(4\lambda))$. Hence,

$$t(g_1, \ldots, g_n) < \frac{16\omega\mu_G(N(4\lambda))}{\mu_G(N(\lambda))}. \tag{6.14}$$

Using Corollary 6.53 again, we obtain an irreducible λ-sequence $(h_1, \ldots, h_m)$ such that $t(h_1, \ldots, h_m) = \omega$ and $1/\mu_G(N(4\lambda)) \leq m \leq 4/\mu_G(N(\lambda))$. Using Proposition 6.45(iii) to replace $(h_1, \ldots, h_m)$ with $(h_m^{-1}, \ldots, h_1^{-1})$ if necessary, we can assume that

$$t(h_1, \ldots, h_m) = -\sum_{i=1}^{n} s(g_0, h_i)\|h_i\|_d.$$

We now define a sequence $(g'_1, \dots, g'_{n'})$ such that

1. $n' = n + km$ for some integer $k \geq 0$.

2. $g'_i = g_i$ for $1 \leq i \leq n$.

3. For $i \geq n+1$, $g'_i = h_j$ with $j \equiv i - n \pmod{m}$.

From the definition of the total weight, for $k < t(g_1, \dots, g_n)/\omega$, we have

$$t(g'_1, \dots, g'_{n'}) = t(g_1, \dots, g_n) - k\omega.$$

We choose integer $k < t(g_1, \dots, g_n)/\omega + 1$ such that $|t(g'_1, \dots, g'_{n'})| \leq \omega/2$. Then by (6.14), and the trivial bound $\mu_G(N(\lambda)) < \mu_G(N(4\lambda))$, we have

$$n' < n + km < \frac{4}{\mu_G(N(\lambda))} + \left(\frac{16\mu_G(N(4\lambda))}{\mu_G(N(\lambda))} + 1\right)\frac{4}{\mu_G(N(\lambda))} \leq \frac{72\mu_G(N(4\lambda))}{\mu_G^2(N(\lambda))}.$$

Note that $(g'_1, \dots, g'_{n'})$ is a λ-sequence with $g'_1 \dots g'_{n'} = \mathrm{id}_G$. We assume further that $0 \leq t(g'_1, \dots, g'_{n'}) < \omega/2$ as the other case can be dealt with similarly. Obtain an irreducible concatenation $(h'_1, \dots, h'_{m'})$ of $(g'_1, \dots, g'_{n'})$. From Lemma 6.49, we get

$$t(h'_1, \dots, h'_{m'}) < t(g'_1, \dots, g'_{n'}) + 25(n' - m')\gamma \leq \frac{\omega}{2} + \frac{1800\mu_G(N(4\lambda))\gamma}{\mu_G^2(N(\lambda))}.$$

Using Corollary 6.53(i) and Lemma 6.54, we have

$$\frac{1800\mu_G(N(4\lambda))\gamma}{\mu_G^2(N(\lambda))} \leq \frac{1800\mu_G(N(4\lambda))\gamma}{\mu_G^2(N(4\lambda))/16^2} < \frac{5 \cdot 10^5\gamma}{N(4\lambda)} \leq 5 \cdot 10^5\gamma\frac{4\omega}{\lambda}.$$

As $\gamma < 10^{-8}\rho$, and $\lambda = \rho/16 - \gamma$, one can check that the lass expression is at most $\omega/400$. Hence, $t(h'_1, \dots, h'_{m'}) < \omega$. From the definition of ω, we must have $t(h'_1, \dots, h'_{m'}) = 0$. Thus by Lemma 6.49 again,

$$t(g'_1, \dots, g'_n) \in I(25n'\gamma) \subseteq I(\omega/400),$$

which completes the proof. $\square$

Recall that a **Polish space** is a topological space which is separable and completely metrizable. In particular, the underlying topological space of any connected compact Lie group is a Polish space. Let X be a Polish space. A subset B of X is **Borel** if B can be formed from open subsets of X (equivalently, closed subsets of X) through taking countable unions, taking countable intersections, and taking complement. A function $f : X \to Y$ between Polish space is Borel, if the inverse image of any Borel subset of Y is Borel. A subset A of X is **analytic** if it is the continuous image of another Polish space Y. Below are some standard facts about these notions; see [110] for details.

Fact 6.56. *Suppose X, Y are Polish spaces, and $f : X \to Y$ is continuous. We have the following:*

(i) *Every Borel subset of X is analytic.*

(ii) *Equipping $X \times Y$ with the product topology, the graph of a Borel function from X to Y is analytic.*

(iii) *The collection of analytic subsets of X is closed under taking countable unions, taking intersections and cartersian products.*

(iv) *Images of analytic subsets in X under f is analytic.*

Given $x \in \mathbb{R}$, let $\|x\|_{\mathbb{T}}$ be the distance of x to the nearest element in $\mathbb{Z}$. We now obtain a consequence of Lemma 6.55.

Corollary 6.57 (Analytic multivalued almost homomorphism)**.** *There is an analytic subset Γ of $G \times \mathbb{T}$ satisfying the following properties:*

(i) *The projection of Γ on G is surjective.*

(ii) $(\mathrm{id}_G, \mathrm{id}_{\mathbb{R}/\omega\mathbb{Z}})$ *is in Γ.*

(iii) *If* $g_1, g_2 \in G$ *and* $t_1, t_2, t_3 \in \mathbb{R}$ *are such that* $(g_1, t_1/\omega + \mathbb{Z})$, $(g_2, t_2/\omega + \mathbb{Z})$, *and* $(g_1g_2, t_3/\omega + \mathbb{Z})$ *in* Γ, *then*

$$\|(t_1 + t_2 - t_3)/\omega\|_{\mathbb{T}} < 1/400.$$

(iv) *There are* $g_1, g_2 \in G$ *and* $t_1, t_2 \in \mathbb{R}$ *with that* $(g_1, t_1/\omega + \mathbb{Z}), (g_2, t_2/\omega + \mathbb{Z}) \in \Gamma$ *and* $\|(t_1 - t_2)/\omega\|_{\mathbb{T}} > 1/3$.

Proof. Let Γ consist of $(g, t/\omega + \mathbb{Z}) \in G \times \mathbb{T}$ with $g \in G$ and $t \in \mathbb{R}$ such that there is $n \leq 1/\mu_G(N(\lambda)) + 1$ and an irreducible λ-sequence $(g_1, \ldots, g_n)$ satisfying

$$g = g_1 \cdots g_n \quad \text{and} \quad t = t(g_1, \ldots, g_n).$$

Note that the relative sign function $s : G \times G \to \mathbb{R}$ is Borel, the set $N[\gamma]$ is compact, and the function $x \to \|x\|_d$ is continuous. Hence, by Fact 6.56(i,ii), the function $(g_1, \ldots, g_n) \mapsto t(g_1, \ldots, g_n)$ is Borel, and its graph is analytic. For each n, by Fact 6.56(iii)

$$\begin{aligned}\widetilde{\Gamma}_n := \{&(g, t, g_1, \ldots, g_n) \in G \times \mathbb{R} \times G^n : \\ &\|g_i\|_d < \lambda \text{ for } 1 \leq i \leq n, g = g_1 \cdots g_n, t = t(g_1, \ldots, g_n)\}\end{aligned}$$

is analytic. Let Γ_n be the image of $\widetilde{\Gamma}_n$ under the continuous map

$$(g, t, g_1, \ldots, g_n) \mapsto (g, t/\omega + \mathbb{Z}).$$

Then by Fact 6.56(iv), Γ_n is analytic. Finally, $\Gamma = \bigcup_{n<1/\mu_G(N(\lambda))+1} \Gamma_n$ is analytic by Fact 6.56(iii).

We now verify that Γ satisfies the desired properties. It is easy to see that (i) and (ii) are immediately from the construction, and (iii) is a consequence of Lemma 6.55. We now prove (iv). Using Corollary 6.53, we obtain an irreducible λ-sequence $(g_1, \ldots, g_n)$ with

$t(g_1, \ldots, g_n) = \omega$ and $n < 4/\mu_G(N(\lambda))$. Note that

$$|t(g_1, \ldots, g_{k+1}) - t(g_1, \ldots, g_k)| \le \lambda.$$

Hence, there must be $k \in \{1, \ldots n\}$ such that $\omega/3 < t(g_1, \ldots, g_k) < 2\omega/3$. Set $t_1 = 0$ and $t_2 = t(g_1, \ldots, g_k)$ for such k. It is then easy to see that $\|(t_1 - t_2)/\omega\|_{\mathbb{T}} > 1/3$. $\square$

To construct a group homomorphism from G to $\mathbb{T}$, we will need three more facts. Recall the following measurable selection theorem from descriptive set theory; see [23, Theorem 6.9.3].

Fact 6.58 (Kuratowski and Ryll–Nardzewski measurable selection theorem)**.** *Let $(X, \mathcal{A})$ be a measurable space, Y a complete separable metric space equipped with the usual Borel σ-algebra, and F a function on X with values in the set of nonempty closed subsets of Y. Suppose that for every open $U \subseteq Y$, we have*

$$\{a \in X : F(a) \cap U \neq \varnothing\} \in \mathcal{A}.$$

Then F has a selection $f : X \to Y$ which is measurable with respect to $\mathcal{A}$.

A **Polish group** is a topological group whose underlying space is a Polish space. In particular, Lie groups are Polish groups. A subset A of a Polish space X is **universally measurable** if A is measurable with respect to every complete probability measure on X for which every Borel set is measurable. In particular, every analytic set is universally measurable; see [147] for details. A map $f : X \to Y$ between Polish spaces is **universally measurable** if inverse images of open sets are universally measurable. We have the following recent result from descriptive set theory by [147]; in fact, we will only apply it to Lie groups so a special case which follows from an earlier result by Weil [168, page 50] suffices.

Fact 6.59 (Rosendal)**.** *Suppose G and H are Polish groups, $f : G \to H$ is a universally measurable group homomorphism. Then f is continuous.*

Finally, we need the following theorem from geometry by Grove, Karcher, and Ruh [85] and independently by Kazhdan [109], that in a compact Lie groups an almost homomorphism is always close to a homomorphism uniformly. We remark that the result is not true for general compact topological groups, as a counterexample is given in [167].

Fact 6.60 (Grove–Karcher–Ruh; Kazhdan). *Let G, H be compact Lie groups. There is a constant c only depending on H, such that for every real number q in $[0, c]$, if $\pi : G \to H$ is a q-almost homomorphism, then there is a homomorphism $\chi : G \to H$ which is $1.36q$-close to π. Moreover, if π is universally measurable, then χ is universally measurable. When $H = \mathbb{T}$, we can take $c = \pi/6$.*

The next theorem is the main result in this subsection. It tells us from an almost linear pseudometric, one can construct a group homomorphism to $\mathbb{T}$ or to $\mathbb{R}$; this can also be seen as a stability theorem of Proposition 6.42.

Theorem 6.61. *Suppose d is γ-linear and γ-monotone. Then there is a continuous surjective group homomorphism $\chi : G \to \mathbb{T}$ such that for all $g \in \ker(\chi) \cap N(\lambda)$, we have $\|g\|_d \in N(\lambda/2)$.*

Proof. Let ω be the returning weight of d, and let Γ be as in the proof of Corollary 6.57. Equip G with the σ-algebra $\mathcal{A}$ of universally measurable sets. Then $\mathcal{A}$ in particular consists of analytic subsets of G. Define F to be the function from G to the set of closed subsets of $\mathbb{T}$ given by

$$F(g) = \overline{\{t/\omega + \mathbb{Z} : t \in \mathbb{R}, (g, t/\omega + \mathbb{Z}) \in \Gamma\}}.$$

If U is an open subset of $\mathbb{T}$, then $\{g \in G : F(g) \cap U \neq \varnothing\}$ is in $\mathcal{A}$ being the projection on G of the analytic set $\{(g, t/\omega + \mathbb{Z}) \in G \times \mathbb{T} : (g, t/\omega + \mathbb{Z}) \in \Gamma \text{ and } t \in U\}$. Applying Fact 6.58, we get a universally measurable $1/400$-almost homomorphism $\pi : G \to \mathbb{T}$. Using Fact 6.60, we get a universal measurable group homomorphism $\chi : G \to \mathbb{T}$ satisfying

$$\|\chi(g) - \pi(g)\|_{\mathbb{T}} < 1.36/400 = 0.0068.$$

The group homomorphism χ is automatically continuous by Fact 6.59. Combining with Corollary 6.57(iv), we see that χ cannot be the trivial group homomorphism, so χ is surjective.

Finally, for g be in $\ker(\chi) \cap (N(\lambda))$, we need to verify that g is in $N(\lambda/2)$. Suppose to the contrary that $g \notin N(\lambda/2)$. Choose $n = \lfloor 1/\mu_G(N(4\lambda)) \rfloor$, and $(g_1, \dots, g_n)$ the λ-sequence such that $g_i = g$ for $i \in \{1, \dots, n\}$. By Proposition 6.45, $(g_1, \dots, g_n)$ is irreducible. Hence, by Lemma 6.52, $t(g_1, \dots, g_n) < \omega$. As $n \leq 1/\mu_G(N(\lambda) + 1)$, by construction and Corollary 6.57(iii), we have

$$\pi(g^n) \in t(g_1, \dots, g_n)/\omega + I(1/400) + \mathbb{Z} = n\|g\|_d/\omega + I(1/400) + \mathbb{Z}.$$

Since $g^n \in \ker\chi$, we have $\|\pi(g^n)\|_{\mathbb{T}} < 0.0068$, so $n\|g\|_d/\omega < (0.0068 + 1/400)$. By Corollary 6.53(i) and Lemma 6.54, this implies

$$\|g\|_d \leq \frac{(0.0068 + 1/400) \cdot 4\lambda}{\mu_G(N(\lambda))} \mu_G(N(4\lambda)) < \frac{\lambda}{2},$$

which is a contradiction. This completes our proof. $\square$

6.8 Geometry of minimal and nearly minimal expansion pairs II

In this section, we study the shape of a nearly minimally expanding pair relative to a connected closed proper subgroup of the ambient Lie group such that the cosets of the subgroup intersect the nearly minimal expanding pair "transversally in measure". Section 8.1 shows that in a compact connected Lie group, such a subgroup exists and can in fact be chosen to be a one-dimensional torus. In Section 8.2 and 8.3, we obtain shape description for the minimally expanding pair and nearly minimally expanding pair respectively. Using that we will construct linear and almost linear pseudometric as described in Section 7.

Throughout this section, G is a *connected unimodular Lie* group, and H is a *connected unimodular* closed subgroup of G. In particular, the left Haar measures μ_G and μ_H are Haar measures. We let A, and B be σ-compact subsets of G.

We set $\mu_{G/H}$ and $\mu_{H\backslash G}$ to be the Radon measures on G/H and $H\backslash G$ such that we have the quotient integral formulas (Fact 6.8 and Lemma 6.9(vi)). We also remind the reader that we normalize the measure whenever a group under consideration is compact. Hence, if H is compact, then we have

$$\mu_G(AH) = \mu_{G/H}(\pi A) \text{ and } \mu_{H\backslash B}(\widetilde{\pi}B) = \mu_{H\backslash G}(\widetilde{\pi}B),$$

and if $\chi : H \to \mathbb{R}$ is a continuous and surjective group homomorphism with compact kernel, then the pushforward of μ_H is the Lesbegue measure $\mu_{\mathbb{R}}$.

6.8.1 Toric transversal intersection in measure

In this section, we assume that G is *compact*. We will consider a more general situation than what we need assuming

$$\mu_G(A) = \mu_G(B) = \kappa \text{ and } \mu_G(AB) < M\kappa.$$

where M is a constant. We will prove that if κ is sufficiently small measure, then there is a torus $H \subseteq G$ such that we are in the short fiber scenario (i.e., for every $x, y \in G$,

$$\min\{\mu_H(A \cap xH), \mu_H(Hy \cap B)\} < \lambda \tag{6.15}$$

for some given constant λ). Assume (6.15) fails for every maximal tori H, which means we have a long fiber "in every direction". Then both A and B can be seen as a variant of *Kakeya sets* in Lie groups; see [140] for some properties of Kakeya sets in this setting. In general, it is well-known that Kakeya sets can have arbitrarily small measure; but when A, B has nearly minimally expansion, we will show in this section that both A and B must not be too small. Our result in particular applies to approximate groups, as described in Section 6.2.

We use the following lemma, which can be seen as a corollary of the quotient integral formula (Fact 6.8).

Lemma 6.62. *For every $b \in G$, the following identity holds*

$$\mu_G\big(A(B \cap Hb)\big) = \int_G \mu_H\big((A \cap aH)(B \cap Hb)\big)\, \mathrm{d}\mu_G(a).$$

Proof. Let $C = A(B \cap Hb)$. From Fact 6.8, one has

$$\mu_G(C) = \int_G \mu_H(C \cap ab^{-1}Hb)\, \mathrm{d}\mu_G(a) = \int_G \mu_H(C \cap aHb)\, \mathrm{d}\mu_G(a).$$

Hence, it suffices to check that

$$C \cap aHb = (A \cap aH)(B \cap Hb) \quad \text{ for all } a, b \in G.$$

The backward inclusion is clear. Note that $aHHb = aHb = ab(b^{-1}Hb)$ for all a and b in G. For all a, a', and b in G, we have we have $a'Hb = aHb$ when $aH = a'H$ and $aHb \cap a'Hb = \varnothing$ otherwise. An arbitrary element $c \in C$ is in $(A \cap a'H)(B \cap Hb)$ for some $a' \in A$. Hence if c is also in aHb, we must have $a'H = aH$. So we also get the forward inclusion. $\square$

Suppose r and s are in $\mathbb{R}$, the sets $A_{(r,s]}$ and $\pi A_{(r,s]}$ are given by

$$A_{(r,s]} := \{a \in A : \mu_H(A \cap aH) \in (r, s]\}$$

and

$$\pi A_{(r,s]} := \{aH \in G/H : \mu_H(A \cap aH) \in (r, s]\}.$$

In particular, $\pi A_{(r,s]}$ is the image of $A_{(r,s]}$ under the map π. By Lemma 6.9, $\pi A_{(r,s]}$ is $\mu_{G/H}$-measurable, and $A_{(r,s]}H = \pi^{-1}(\pi A_{(r,s]})$ and $A_{(r,s]}$ are μ_G-measurable.

Lemma 6.63. *Suppose $\lambda < 1$ is a constant, and either there is $a \in A$ such that $\mu_H(A \cap aH) > \lambda$, or there is $b \in B$ such that $\mu_H(B \cap Hb) > \lambda$. Then,*

$$\min\left\{\frac{\mu_G(A)}{\mu_{G/H}(\pi A)}, \frac{\mu_G(B)}{\mu_{H\backslash G}(\widetilde{\pi} B)}\right\} \geq \frac{\lambda}{(M+2)^2}.$$

Proof. Without loss of generality, suppose $\mu_H(B \cap Hb) > \lambda$ for a fixed $b \in B$. By the quotient integral formula, $\kappa = \mu_G(A)$ is at least

$$\int_{\pi A_{(1/2,1]}} \mu_H(A \cap xH) \,\mathrm{d}\mu_{G/H} xH > \frac{1}{2}\mu_{G/H}(\pi A_{(1/2,1]}) = \frac{1}{2}\mu_G(A_{(1/2,1]}H),$$

Hence, $\mu_G(A_{(1/2,1]}H) < 2\kappa$. We now prove that $\mu_G(A_{(0,1/2]}H) < M\kappa/\lambda$. Suppose that it is not the case. By Lemma 6.62 we get

$$\mu_G(A(B \cap Hb)) \geq \int_{A_{(0,1/2]}H} \mu_H((A \cap aH)(B \cap Hb)) \,\mathrm{d}\mu_G(a)$$

Observe that $\mu_H((A \cap aH)(B \cap Hb)) > \lambda$ since $\mu_H(B \cap Hb) > \lambda$, we have

$$\mu_G(AB) \geq \mu_G(A(B \cap Hb)) \geq \lambda \mu_G(A_{(0,1/2]}H) > M\kappa,$$

contradicting the assumption that $\mathfrak{d}_G(AB) < \kappa$. Hence $\mu_G(AH) < (M+2)\kappa/\lambda$. This implies that there is $a \in A$ such that $\mu_H(A \cap aH) > \lambda/(M+2)$. Now we apply the same argument switching the role of A and B, we get $\mu_{H\backslash G}(\widetilde{\pi} B) < (M+2)^2 \mu_G(B)/\lambda$ which completes the proof. $\square$

Lemma 6.63 leads us to consider the problem of obtaining lower bound for the measure of toric nonexpanders, which is of independent interest.

Definition 6.64. We say A is called a **toric K-expander**, if there is a one-dimensional torus subgroup H of G such that $\mu_G(AH) \geq K\mu_G(A)$.

Lemma 6.65. *Suppose A is not a toric K-expander, $g_1, \ldots, g_n$ are in G, and $A' = \bigcup_{j=1}^d Ag_j$. Then A' is not a toric (nK)-expander.*

Proof. We need to verify for each T that $\mu_G(A'T) < nK\mu_G(A')$. Note that

$$A'T = \Big(\bigcup_{i=1}^n Ag_i\Big)T = \bigcup_{i=1}^n A(g_iT) = \bigcup_{i=1}^n A(g_iTg_i^{-1})g_i = \bigcup_{i=1}^n AT_ig_i,$$

where T_i is the torus subgroup $g_iTg_i^{-1}$ of G. Hence,

$$\mu_G(A'T) < nK\mu_G(A) \leq nK\mu_G(A')$$

as desired. $\square$

Fact 6.66 (Bhatia–Davis inequality)**.** *Suppose $(X, \mathcal{A}, \mu)$ is a measure space, α and β are constants, and $f : X \to \mathbb{R}^{>0}$ is a measurable function with*

$$\alpha \leq \inf_{x\in X} f(x) < \sup_{x\in X} f(x) \leq \beta$$

Then $(\mathbb{E}_x f^2(x)) - (\mathbb{E}_x f(x))^2 \leq (\beta - \mathbb{E}_x f(x))(\mathbb{E}_x f(x) - \alpha)$.

The following lemma will help us to translate the set along some direction.

Lemma 6.67. *Suppose K, α, and β are constant with $K > 1$, $0 < \alpha < \beta < 1$, $\mu_G(AH) = K\mu_G(A)$, and*

$$\alpha \leq \inf_{g\in A} \mu_H(A \cap gH) < \sup_{g\in A} \mu_H(A \cap gH) \leq \beta.$$

Then for every number $\gamma \geq (\alpha + \beta - K\alpha\beta)\mu_G(A)$, there is $h \in H$ with $\mu_G(A \cap Ah) = \gamma$.

Proof. Let μ_G, μ_H be normalized Haar measures of G and H. Choose h from H uniformly at random. Note that

$$\begin{aligned}\mathbb{E}_{h\in H}\mu_G(A\cap Ah) &= \int_H \mu_G(A\cap Ah)\,\mathrm{d}\mu_H(h)\\ &= \int_H\int_G \mathbb{1}_A(g)\mathbb{1}_A(gh)\,\mathrm{d}\mu_G(g)\,\mathrm{d}\mu_H(h).\end{aligned}$$

Using the quotient integral formula (Fact 6.8), the above equality is

$$\begin{aligned}\int_G \mathbb{1}_A(g)\mu_H(A\cap gH)\,\mathrm{d}\mu_G(g) &= \int_{G/H} \mu_H^2(A\cap gH)\,\mathrm{d}\mu_{G/H}(gH)\\ &= \mathbb{E}_{gH\in G/H}\big(\mu_H^2(A\cap gH)\big)\\ &= \mu_{G/H}(\pi A)\mathbb{E}_{gH\in\pi A}\big(\mu_H^2(A\cap gH)\big)\end{aligned}$$

Note that $\mu_{G/H}(\pi A) = \mu_G(AH) = K\mu_G(A)$, and

$$\mathbb{E}_{gH\in\pi A}\big(\mu_H(A\cap gH)\big) = 1/K.$$

Hence, applying the Bhatia–Davis inequality (Fact 6.66), we get

$$\mathbb{E}_{h\in H}\mu_G(A\cap Ah) \leq (\alpha+\beta-K\alpha\beta)\mu_G(A).$$

The desired conclusion follows from the continuity of $H\to\mathbb{R}, h\mapsto \mu_G(A\cap Ah)$. $\square$

The following lemma says that for a toric nonexpander A and a torus subgroup H of G, one can slightly modify A to get A' such that most of $A'H$ can be covered by finitely many right translations of A'.

Lemma 6.68. *Suppose $K>1$ is a constant, A is not a toric K-expander, and H is a one-dimensional torus subgroup of G. Then for every $0<\varepsilon<1/2K$, there is a σ-compact $A'\subseteq A$, integer $m=m(K,\varepsilon)>0$, and $h_1,\dots,h_m\in H$, satisfying*

(i) *A' is not a toric $2K$-expander*

(ii) $\mu_G(A') > (1-\varepsilon K)\mu_G(A)$

(iii) $\mu_G\left(A'H \setminus \bigcup_{i=1}^m A'h_i\right) < \varepsilon\mu_G(AH)$.

Proof. Let μ_G, μ_H be normalized Haar measures on G and H, and let $\mu_{G/H}$ be the invariant Radon measure induced by μ_G and μ_H on the homogeneous space G/H. Let

$$\pi A_{(\varepsilon,1]} = \{g \in G/H \mid \varepsilon < \mu_H(A \cap gH) \leq 1\},$$
$$\pi A_{(0,1]} = \{g \in G/H \mid A \cap gH \neq \varnothing\}.$$

Let $x = \mu_{G/H}(\pi A_{(0,\varepsilon)})/\mu_{G/H}(\pi A_{(0,1]})$, then $\mu_{G/H}(\pi A_{(\varepsilon,1]}) = (1-x)(\mu_{G/H}(\pi A_{(0,1]})$. One has

$$\frac{\mu_{G/H}(\pi A_{(0,1]})}{K} < \mu_G(A) \leq (\varepsilon x + 1 - x)\mu_{G/H}(\pi A_{(0,1]})).$$

It follows that

$$x = \frac{\mu_{G/H}(\pi A_{(0,\varepsilon]})}{\mu_{G/H}(\pi A_{(0,1]})} < \frac{K-1}{K(1-\varepsilon)}.$$

Choose σ-compact $A' \subseteq A_{(\varepsilon,1]}$ such that $\mu_G(A_{(\varepsilon,1]} \setminus A') = 0$. One has

$$\begin{aligned} \mu_G(A') &\geq \mu_G(A) - \varepsilon\frac{K-1}{K(1-\varepsilon)}\mu_{G/H}(\pi A_{(0,1]}) \\ &\geq \left(1 - \varepsilon\frac{K-1}{1-\varepsilon}\right)\mu_G(A) = \frac{1-\varepsilon K}{1-\varepsilon}\mu_G(A) \geq (1-\varepsilon K)\mu_G(A). \end{aligned}$$

Hence we have

$$\mu_G(A') \geq (1-\varepsilon K)\mu_G(A) \geq \frac{1-\varepsilon K}{K}\mu_G(AT) > \frac{1}{2K}\mu_G(A'T),$$

for every torus T when $\varepsilon < 1/2K$.

It remains to obtain $m = m(\varepsilon, K)$ and $h_1, \ldots, h_m$ such that (iii) is satisfied. Construct a sequence (A'_n) of σ-compact subset of G with $A'_n H = A'H$ as follows. Let $A'_0 = A'$. Suppose

A'_n has been constructed. Set

$$\varepsilon_n = \inf_{gH \in \pi_A} \mu_H(A'_n \cap gH) \text{ and } K_n = \mu_G(A'_n H)/\mu_G(A'_n).$$

Using Lemma 6.67, obtain $h'_n \in H$ such that

$$\mu_G(A'h'_n \setminus A') = \varepsilon_n(K'_n - 1)\mu_G(A'_n).$$

Finally, let $A'_{n+1} = A'_n \cup A'_n h'_n$. Then, $A_{n+1}H = A_n H = A'H$. It is also easy to see that $\varepsilon_n \geq \varepsilon$ for all n. Now, if $\mu_G(A_n) < (1-\varepsilon)\mu_G(A'H)$ for some n, then

$$K_n = \frac{\mu_G(A'H)}{\mu_G(A_n)} \geq \frac{1}{1-\varepsilon},$$

and hence,

$$\mu_G(A'_{n+1}) \geq \frac{1-\varepsilon+\varepsilon^2}{1-\varepsilon}\mu_G(A'_n).$$

As $(1-\varepsilon+\varepsilon^2)/(1-\varepsilon) > 1$, this cannot be the case for all n. Let N be the first n such that $\mu_G(A'_N) > (1-\varepsilon)\mu_G(A'H)$. Then

$$2K\mu_G(A') > \mu_G(A'H) \geq \mu_G(A'_N) \geq \left(\frac{1-\varepsilon+\varepsilon^2}{1-\varepsilon}\right)^N \mu_G(A').$$

This implies that

$$N \leq \frac{\log 2K}{\log(1-\varepsilon+\varepsilon^2) - \log(1-\varepsilon)}.$$

Finally set $m = 2^N$, and choose $h_1, \ldots, h_m$ such that $A'_N = \bigcup_{i=1}^m A'h_i$, we get the desired conclusion. $\square$

The next simple lemma shows that we can find finitely many torus such that the product of them is G.

Fact 6.69. *Let G is compact. Then there is a constant n depending only on the dimension* *of G such that there are n tori $H_1, \ldots, H_n$ in G with $H_1 \cdots H_n = G$.*

Let A be a toric nonexpander. The next important "cage" lemma provides an inductive construction to construct a set C from A, such that the size of C is bounded from above, and any right translations of A cannot "escape" C.

Lemma 6.70 (Cage lemma)**.** *Suppose $A \subseteq G$ is not a toric K-expander, then there is a σ-compact $C \subseteq A$ such that $\mu_G(C) = O_K(1)\mu_G(A)$ and for all $g \in G$*

$$\frac{\mu_G(C \cap Ag)}{\mu_G(A)} > \frac{1}{2}.$$

Proof. Using Fact 6.69, we obtain one-dimensional torus subgroups $H_1, \ldots, H_n$ of G such that $G = H_1 \cdots H_n$. For every constant $\varepsilon_0, \ldots, \varepsilon_{n-1}$, we construct a sequence $(A_i)_{i=0}^n$ of σ-compact subsets of G and a sequence $(K_i)_{i=0}^n$ of constants satisfying the following conditions

1. $A_0 = A$ and $K_0 = K$.
2. A_i is not a toric K_i-expander for $0 \leq i \leq n$.
3. $A_i \subseteq A_{i+1}$ with $\mu_G(A_{i+1}) \leq K_i \mu_G(A_i)$ for $0 \leq i \leq n-1$.
4. $\mu_G(A_i \setminus A_{i+1}h_{i+1}) \leq \varepsilon_i \mu_G(A_i)$ for any $h_{i+1} \in H_{i+1}$ and $0 \leq i \leq n-1$.
5. $K_{i+1} = K_i(\varepsilon_0, \ldots, \varepsilon_i)$ for $0 \leq i \leq n-1$.

Suppose we have $A_0, \ldots, A_i$ and $K_0, \ldots, K_i$ satisfying all the conditions restricted down to i. We are going to construct A_{i+1}. By (2), A_i is a toric K_i-nonexpander. Let $\delta > 0$ be a parameter which we will determine later. Using Lemma 6.68, we obtain a σ-compact $A_i' \subseteq A_i$, $m = m(K_i, \delta)$, and $h_1', \ldots, h_m' \in H_{i+1}$ such that

$$\mu_G(A_i') > (1 - \delta K_i)\mu_G(A_i),$$

A'_i is a toric $2K_i$-nonexpander, and

$$\mu_G\left(A'_iH_{i+1} \setminus \bigcup_{j=1}^{m} A'_ih'_j\right) < \delta\mu_G(A'_iH_{i+1}).$$

By adding one element of H_{i+1} if necessary, we can arrange that e_G is in $\{1, \ldots, m\}$. Set $A_{i+1} = \bigcup_{j=1}^{m} A'_ih'_j$ and set $K_{i+1} = mK_i$. By Lemma 6.65, A_{i+1} is not a toric K_{i+1}-expander, so (2) is satisfied. By construction $A_i \subseteq A_{i+1}$, and

$$\mu_G(A_{i+1}) \leq \mu_G(A'_iH_{i+1}) \leq \mu_G(A_iH_{i+1}) < K_i\mu_G(A_i),$$

so we have (3). For every $h' \in H_{i+1}$, since $A_{i+1}h' \subseteq A'_iH_{i+1}$, we have

$$\mu_G(A'_i \setminus A_{i+1}h') < \delta\mu_G(A'_iH_{i+1}) \leq \delta\mu_G(A_iH_{i+1}) < \delta K_i\mu_G(A_i).$$

Therefore,

$$\mu_G(A_i \setminus A_{i+1}h') < 2\delta K_i\mu_G(A_i).$$

Note that the construction so far depends on δ. Now, by choosing $\delta = \delta(K_i, \varepsilon_i)$ sufficiently small, we can make

$$\mu_G(A_i \setminus A_{i+1}h') < \varepsilon_i\mu_G(A_i),$$

so we get (4). Finally, note that $K_{i+1} = mK_i$, $m = m(K_i, \delta)$, $\delta = \delta(K_i, \varepsilon_i)$, and $K_i = K_i(\varepsilon_0, \ldots, \varepsilon_{i-1})$, so

$$K_{i+1} = K_{i+1}(\varepsilon_0, \ldots, \varepsilon_i),$$

which gives us (5).

We now proceed with the proof of the lemma. Let $\varepsilon_0, \ldots, \varepsilon_{n-1}$ be parameters which we will determine later, and obtain $(A_i)_{i=0}^{n}$ and $(K_i)_{i=0}^{n}$ as in the earlier step. Set $C = A_n$. Note that

$$G = G^{-1} = (H_1 \dots H_n)^{-1} = H_n \cdots H_1.$$

Hence, an arbitrary $g \in G$ can be written as a product $h_n \cdots h_1$ with $h_i \in H_i$ for $i \in \{1, \dots, n\}$. Now consider $Cg = A_n h_n \cdots h_1$. By (4), $\mu_G(A_{n-1} \setminus A_n h_n) < \varepsilon_{n-1}\mu_G(A_{n-1})$. Next, for $A_n h_n h_{n-1}$, again by (4),

$$\begin{aligned}\mu_G(A_{n-2} \setminus A_n h_n h_{n-1}) &\leq \mu_G(A_{n-2} \setminus A_{n-1}h_{n-1}) + \mu_G(A_{n-1}h_{n-1} \setminus A_n h_n h_{n-1}) \\ &< \varepsilon_{n-2}\mu_G(A_{n-2}) + \varepsilon_{n-1}\mu_G(A_{n-1}).\end{aligned}$$

Hence by induction and (3), we conclude that

$$\mu_G(A \setminus Cg) < \sum_{i=0}^{n-1} \varepsilon_i \mu_G(A_i) \leq \left(\sum_{i=0}^{n-1} \varepsilon_i \prod_{j=0}^{i-1} K_j \right) \mu_G(A).$$

Using (5), we can choose ε_i sufficiently small such that $\left(\sum_{i=0}^{n-1} \varepsilon_i \prod_{j=0}^{i-1} K_j \right) < 1/2$. Then, for all $g \in G$.

$$\frac{\mu_G(C \cap Ag)}{\mu_G(A)} = \frac{\mu_G(Cg^{-1} \cap A)}{\mu_G(A)} > \frac{1}{2}.$$

Finally, note that we can choose $\varepsilon_0, \dots, \varepsilon_{n-1}$ depending on K. Hence,

$$\mu_G(C) = O_K(1)\mu_G(A),$$

which completes the proof. □

Suppose μ and ν are measures on G. Their **convolution** $\mu * \nu$ is the unique measure satisfying the property

$$\int_G f(x) \, \mathrm{d}\mu * \nu(x) = \int_{G \times G} f(xy) \, \mathrm{d}\mu(x) \, \mathrm{d}\nu(y).$$

The convolution exists for all the case we care about. If $\mu(A) > 0$, the **uniform measure** on A is defined by

$$\mu_A(X) := \frac{\mu_G(A \cap X)}{\mu_G(A)} \quad \text{for measurable } X \subseteq G.$$

The following lemma is an immediate consequence of the definition and Fubini's theorem.

Lemma 6.71. *Let μ_A be the uniform measure on A. Then $\mu_A * \mu_G = \mu_G$.*

Proof. Let X be a measurable set in G, then

$$\mu_A * \mu_G(X) = \int_G \left(\int_G \mathbb{1}_X(xy) \, \mathrm{d}\mu_G(y) \right) \mathrm{d}\mu_A(x) = \mu_G(X) \int_G \mathrm{d}\mu_A(x) = \mu_G(X)$$

as desired. □

The next theorem is the main result of this subsection. It gives us a lower bound control on the toric nonexpanders. Together with Lemma 6.63, this quantitative result shows that a sufficiently small set which inside some nearly minimal expansion pair cannot contain a long fiber in every direction.

Proposition 6.72 (Bounding size of toric nonexpanders)**.** *For each K, there is $S = O_K(1)$ such that if A is not a toric K-expander, then $\mu_G(A) > S$.*

Proof. Let C be as in Lemma 6.70. Recall μ_A is the uniform measure on A, and $\mu_A * \mu_G$ is the convolution measure. Then

$$\begin{aligned} \int_G \mathbb{1}_C(x) \, \mathrm{d}\mu_A * \mu_G(x) &= \int_G \int_G \mathbb{1}_C(xy) \, \mathrm{d}\mu_A(x) \, \mathrm{d}\mu_G(y) \\ &= \int_G \frac{\mu_G(Cy^{-1} \cap A)}{\mu_G(A)} \, \mathrm{d}\mu_G(y) \geq \frac{1}{2}. \end{aligned}$$

This means $\mu_A * \mu_G(C) \geq 1/2$. By Lemma 6.71, this implies $\mu_G(C) \geq 1/2$. Since $\mu_G(C) = O_K(1)\mu_G(A)$ we get the desired conclusion. □

We now deduce the main theorem of this section.

Theorem 6.73. *There is $S = O_{M,\lambda}(1)$ such that if $\mu_G(A) = \mu_G(B) = \kappa < S$ and $\mu_G(AB) < M\kappa$, then there is a one-dimensional torus subgroup H of G such that for all $x, y \in G$*

$$\mu_H(A \cap xH) < \lambda \text{ and } \mu_H(B \cap Hy) < \lambda.$$

Proof. Suppose for every one-dimensional torus subgroup H of G, either $\mu_H(A \cap xH) > \lambda$ some $x \in G$ or $\mu_H(B \cap By) > \lambda$ for some $y \in G$. Then by Lemma 6.63, A is not a toric K-expander with $K = (M+2)^2/\lambda$. Hence, by Proposition 6.72, we have $\mu_G(A) > S$ with $S = O_{M,\lambda}(1)$. Thus, we get the desired conclusion. □

We get the following immediate corollary for approximate groups. Since our proof is quantitative, we can make the constant below quantitative if we wish.

Corollary 6.74. *There is $S = O_K(1)$ such that if A is a K-approximate group with $\mu_G(A) < S$, then there is a one-dimensional torus subgroup H of G such that for all $x, y \in G$*

$$\mu_H(A \cap xH) < \lambda \text{ and } \mu_H(B \cap Hy) < \lambda.$$

6.8.2 Linear pseudometric from minimal expansions

Throughout the subsection G is a connected *noncompact* unimodular group, and H is a closed subgroup of G which is either isomorphic to $\mathbb{R}$, or some smaller dimension connected unimodular group, so that by induction on dimension we may assume Theorem 6.1 holds on H. Suppose (A, B) is minimally expanding, that is,

$$\mu_G(AB) = \mu_G(A) + \mu_G(B),$$

and both A, B have positive measure.

This section can also be seen as a preview of Section 8.3. The strategy of this section also works for compact G replacing H with result of Section 8.1. For convenience of notation, we will treat the compact case in Section 8.3 together with the situation where (A, B) is nearly minimally expanding.

Lemma 6.75. *For all $a \in A$ and $b \in B$, we have*

1. $\mu_H\big((A \cap aH)(B \cap Hb)\big) \geq \mu_H(A \cap aH) + \mu_H(B \cap Hb)$.

2. $\mu_G\big(A(B \cap Hb)\big) \geq \mu_G(A) + \mu_{G/H}(\pi A)\mu_H(B \cap Hb)$.

3. $\mu_G\big((A \cap aH)B\big) \geq \mu_H(A \cap aH)\mu_{H\backslash G}(\widetilde{\pi} B) + \mu_G(B)$.

The equality in (2) holds if and only if the equality in (1) holds for almost all $a \in AH$. A similar conclusion holds for (3).

Proof. The first inequality comes from a direct application of Kemperman inequality. For the second inequality, by right translating B and using the unimodularity of G, we can arrange that $Hb = H$. The desired conclusion follows from applying (1) and Lemma 6.62. □

The next lemma gives us the important geometric properties of A and B.

Theorem 6.76 (Rigidity fiberwise)**.** *There is a continuous surjective group homomorphism $\chi : H \to \mathbb{R}$, two compact intervals $I, J \subseteq \mathbb{R}$ with*

$$\mu_{\mathbb{R}}(I) = \frac{\mu_G(A)}{\mu_{G/H}(\pi A)} \quad \text{and} \quad \mu_{\mathbb{R}}(J) = \frac{\mu_G(B)}{\mu_{G\backslash H}(\widetilde{\pi} B)}$$

σ-compact $A' \subseteq A$ and $B' \subseteq B$ with

$$\mu_{G/H}(\pi A') = \mu_{G/H}(\pi A) \quad \text{and} \quad \mu_{H\backslash G}(\widetilde{\pi} B') = \mu_{H\backslash G}(\widetilde{\pi} B)$$

such that the following hold:

(i) $\mu_{G/H}(\pi A) = \mu_{H\backslash G}(\widetilde{\pi} B)$;

(ii) *for each* $a \in A'H$, *we have*

$$\mu_H(A \cap aH) = \frac{\mu_G(A)}{\mu_{G/H}(\pi A)},$$

and there is $\zeta_a \in \mathbb{R}$ *such that*

$$\mu_H\big((A \cap aH) \triangle a\chi^{-1}(\zeta_a + I)\big) = 0.$$

(iii) *for each* $b \in HB'$, *we have*

$$\mu_H(B \cap Hb) = \frac{\mu_G(B)}{\mu_{H\backslash G}(\widetilde{\pi} B)},$$

and there is $\widetilde{\zeta}_b \in \mathbb{R}$ *such that*

$$\mu_H\big((B \cap Hb) \triangle \chi^{-1}(\widetilde{\zeta}_b + J)b\big) = 0.$$

Proof. Without loss of generality we assume $\mu_{G/H}(\pi A) \geq \mu_{H\backslash G}(\widetilde{\pi} B)$. Below, we let Hb range over $\widetilde{\pi} B$, and choose Hb uniformly at random in the expectation. By Lemma 6.62 and the quotient integration formula, we have

$$\begin{aligned}\sup_{Hb} \mu_G(A(B \cap Hb)) &\geq \mu_G(A) + \mu_{G/H}(\pi A) \sup_{Hb} \mu_H(B \cap Hb) \\ &\geq \mu_G(A) + \mu_{G/H}(\pi A)\mathbb{E}_{Hb}\mu_H(B \cap Hb) \\ &= \mu_G(A) + \frac{\mu_{G/H}(\pi A)}{\mu_{H\backslash G}(\widetilde{\pi} B)}\mu_G(B) \geq \mu_G(A) + \mu_G(B).\end{aligned} \tag{6.16}$$

Note that $\mu_G(AB) \geq \sup_{Hb} \mu_G(A(B \cap Hb))$. Since $\mu_G(AB) = \mu_G(A) + \mu_G(B)$, the equality must hold at each step. In particular, we have

$$\mu_{G/H}(\pi A) = \mu_{H\backslash G}(\widetilde{\pi} B), \tag{6.17}$$

and for $\mu_{H\backslash B}$-almost all $Hb \in \widetilde{\pi}(B)$, we have

$$\mu_H(B \cap Hb) = \mathbb{E}_{Hb'}(B \cap Hb') = \frac{\mu_G(B)}{\mu_{H\backslash G}(\widetilde{\pi} B)}.$$

Now, using (6.17) and applying the same argument again switching the role of A and B, we conclude that for $\mu_{G/H}$-almost all aH in πA, we have

$$\mu_H(A \cap aH) = \mathbb{E}_{a'H}\mu_H(A \cap a'H) = \frac{\mu_G(A)}{\mu_{G/H}(\pi A)}. \tag{6.18}$$

Moreover, the fact that equality holds in (6.16) shows that for $\mu_{G/H}$-almost all $aH \in \pi A$ and $\mu_{H\backslash G}$-almost all $Hb \in \widetilde{\pi} B$ we have

$$\mu_H((A \cap aH)(B \cap Hb)) = \mu_H(A \cap aH) + \mu_H(B \cap Hb).$$

By the relationship between μ_G and $\mu_{G/H}$, in the preceding statement, we can replace $\mu_{G/H}$-almost all $aH \in \pi A$ with μ_G-almost all $a \in AH$. We can do the same for μ_G and $\mu_{H\backslash G}$.

Now, as H satisfies Theorem 6.1, for a an b such that (6.18) holds, we can choose a continuous surjective group homomorphism $\chi_{a,b} : H \to \mathbb{R}$ and compact intervals $I_{a,b}$, and $J_{a,b}$ in $\mathbb{R}$ with

$$\mu_{\mathbb{R}}(I_{a,b}) = \mu_H(A \cap aH), \quad \mu_{\mathbb{R}}(J_{a,b}) = \mu_H(B \cap Hb),$$

and

$$\mu_H\big((A \cap aH) \triangle \chi_{a,b}^{-1}(I_{a,b})\big) = 0, \quad \mu_H\big((B \cap Hb) \triangle \chi_{a,b}^{-1}(J_{a,b})\big) = 0.$$

Applying Lemma 6.33, we deduce that $\chi_{a,b}$ is the same, and $I_{a,b}$ and $J_{a,b}$ have constant lengths for μ_G-almost all $a \in AH$ and μ_G-almost all $b \in HB$. It follows that we can choose χ, I, J, A', and B' as described in the statement of the Theorem. $\square$

Corollary 6.77 (Global structure of AH)**.** *For all $g \in G$, we have*

$$\mu_{G/H}(\pi A \triangle \pi(gA)) = 0.$$

Proof. Set $\rho = \mu_G(A)$. Recall that $\mathrm{Stab}_G^{<2\rho}(A)$ is open in G, and every $g \in G$ can be expressed as a finite products of elements in $\mathrm{Stab}_G^{<2\rho}(A)$ since G is connected. Therefore, it suffcies to consider the case where g is in $\mathrm{Stab}_G^{<2\rho}(A)$. Clearly, (gA, B) is a minimally expanding pair. In the current case, we also have $d_A(\mathrm{id}_G, g) < \rho$ and $\mu_G(A \cap gA) > 0$. By Lemma 6.21, $(A \cup gA, B)$ is also a minimally expanding pair. Theorem 6.76 (i) then gives us

$$\mu_{G/H}(\pi A) = \mu_{G/H}(\pi(gA)) = \mu_{G/H}(\pi A \cup \pi(gA)) = \mu_{H\backslash G}(\pi B).$$

This gives us $\mu_{G/H}(\pi A \triangle \pi(gA)) = 0$ as desired. $\square$

Theorem 6.76 and Corollary 6.77 essentially allows us to define a "directed linear pseudometric" on G by "looking at the generic fiber" as discussed in the following remark:

Remark 6.78. Fix $a \in AH$ and let the notation be as in Lemma 6.76. For g_1, g_2 in G such that $g_1^{-1}a, g_2^{-1}a \in A'H$, set

$$\delta_{a,A}(g_1, g_2) = \zeta_{g_1^{-1}a} - \zeta_{g_2^{-1}a}.$$

We have the following linearity property of $\delta_{a,A}$ when the relevant terms are defined, which is essentially the linearity property of the metric from $\mathbb{R}$.

1. $\delta_{a,A}(g_1, g_1)$=0.

2. $\delta_{a,A}(g_1, g_2) = \delta_{a,A}(g_2, g_1)$.

3. $\delta_{a,A}(g_1, g_3) = \delta_{a,A}(g_1, g_2) + \delta_{a,A}(g_2, g_3)$.

Properties (1) and (2) are immediate, and property (3) follows from the easy calculation below:

$$
\begin{aligned}
\delta_{a,A}(g_1, g_2) &= \zeta_{g_1^{-1}a} - \zeta_{g_2^{-1}a} \\
&= \zeta_{g_1^{-1}a} - \zeta_{g_3^{-1}a} + \zeta_{g_3^{-1}a} - \zeta_{g_2^{-1}a} \\
&= \delta_{a,A}(g_1, g_3) \pm \delta_{a,A}(g_3, g_2).
\end{aligned}
$$

Properties (3) also implies that

$$|\delta_{a,A}(g_1, g_3)| = \Big| \pm |\delta_{a,A}(g_1, g_2)| \pm |\delta_{a,A}(g_2, g_3)| \Big|.$$

which tells us that $|\delta_{a,A}|$ is a linear-pseudometric. The problem with the above definitions is that they are not defined everywhere.

There are two ways to overcome this difficulty. The new approach, using difference in measure, will be presented later on. The old approach, present in an earlier version of this paper, is to define a pseudometric on G directly by setting

$$d(g_1, g_2) = \xi \text{ if for } \mu_G\text{-almost all } a \in AH,\ |\delta_{a,A}(g_1, g_2)| = \xi.$$

This is indeed possible. In fact, one can bypass the pseudometric machinery altogether and define the group homomorphism $\chi : G \to \mathbb{T}$ directly by setting

$$\chi(g) = \zeta \text{ if for } \mu_G\text{-almost all } a \in AH,\ |\delta_{a,A}(\mathrm{id}_G, g)| = \zeta.$$

However, this does not come for free, and one need to work equally hard to verify that χ is indeed a group homomorphism.

The old approach can moreover be extended to the case of nearly minimal expansion. However, we can only handle a quadratic error with this old approach because we only have

Corollary 6.82, which lacks the global property of Corollary 6.77. The real problem solved by introducing the pseudometric is to get the linear error bound for the nearly minimal expansion problem. ⋈

Lemma 6.79. *Let $\chi : H \to \mathbb{R}$ be as in Theorem 6.81. For all $g_1, g_2 \in G$ with $\mu_G(g_1A \cap g_2A) > 0$, there is a σ-compact $A'' \subseteq A$ with*

$$\mu_{G/H}(\pi A'') = \mu_{G/H}(\pi A)$$

such that for all $a \in A''H$, the following holds

(i) *there are $\zeta_{g_1^{-1}a}, \zeta_{g_2^{-1}a} \in \mathbb{T}$ such that for $i \in \{1, 2\}$;*

$$\mu_H\big((A \cap aH) \triangle a\chi^{-1}(\zeta_{g_i^{-1}a} + I)\big) = 0.$$

(ii) *with any $\zeta_{g_1^{-1}a}, \zeta_{g_2^{-1}a}$ satisfying (i) and $\delta_{a,A}(g_1, g_2) = \zeta_{g_2^{-1}a} - \zeta_{g_1^{-1}a}$, if we have $\mu_G(g_1A \cap g_2A) > 0$, then*

$$d_A(g_1, g_2) = \mu_{G/H}(\pi A)|\delta_{a,A}(g_1, g_2)|.$$

Proof. Obtain $, \chi, A', I, J$ as in Theorem 6.81. Let $A'' \subseteq G$ be the σ-compact set

$$\{a \in A : g_1^{-1}a, g_2^{-1}a \in A'H\}.$$

By the preceding lemma, $\mu_{G/H}(\pi A'') = \mu_{G/H}(\pi A)$. Fix $a \in A''$. We then have

$$A \cap g_1^{-1}aH = g_1^{-1}a\chi^{-1}(\zeta_{g_1^{-1}a} + I) \text{ and } A \cap g_2^{-1}aH = g_2^{-1}a\chi^{-1}(\zeta_{g_2^{-1}a} + I).$$

Multiplying by g_1 and g_2 respectively, we get (i).

As $\mu_G(g_1A \cap g_2A) > 0$, by Lemma 6.21, $(g_1A \cap g_2A, B)$ is minimally expanding. From

Theorem 6.76(ii), for $\mu_{G/H}$-almost all $aH \in \pi(g_1A \cap g_2A)$, we have

$$\mu_H(g_1A \cap aH) = \frac{\mu_G(g_1A)}{\mu_{G/H}(\pi(g_1A))} \text{ and } \mu_H(g_1A \cap g_2A \cap aH) = \frac{\mu_G(g_1A \cap g_2A)}{\mu_{G/H}(\pi(g_1A \cap g_2A))}.$$

Note that $\pi(g_1A \cap g_2A) \subseteq \pi(g_1A) \cap \pi(g_2A)$. However, by Lemma 6.76,

$$\mu_{G/H}(\pi(g_1A \cap g_2A)) = \mu_{H\backslash G}(\widetilde{\pi}B) = \mu_{G/H}(\pi(g_1A)) = \mu_{G/H}(\pi(g_2A)).$$

Combining with Lemma 6.77, we get

$$\mu_{G/H}(\pi(g_1A)\triangle\pi A) = 0 \text{ and } \mu_{G/H}(\pi(g_1A \cap g_2A)\triangle\pi A) = 0.$$

Hence, shrinking the above A'' if necessary, we can arrange that for all $a \in A''H$,

$$\mu_H(g_1A \cap aH) = \frac{\mu_G(g_1A)}{\mu_{G/H}(\pi A)} \text{ and } \mu_H(g_1A \cap g_2A \cap aH) = \frac{\mu_G(g_1A \cap g_2A)}{\mu_{G/H}(\pi(A))}.$$

Finally, from (i), for all $a \in A''H$, we have

$$\mu_H(g_1A) - \mu_H(g_1A \cap g_2A \cap aH) = |\zeta_{g_1^{-1}a} - \zeta_{g_1^{-1}a}|.$$

Recalling that $d_A(g_1, g_2) = \mu_G(g_1A) - \mu_G(g_1A \cap g_2A)$, we learn that (ii) is satisfied. $\square$

We now construct the pseudometric as promised.

Proposition 6.80 (Linearity of the pseudometric)**.** *For all, g_1, g_2, g_3 in $\mathrm{Stab}_G^{<\rho}(A)$, we have*

$$d_A(g_1, g_2) \in \{d_A(g_1, g_3) + d_A(g_3, g_2), |d_A(g_1, g_3) - d_A(g_3, g_2)|\}.$$

Proof. For $i \in \{1, \ldots, 3\}$, g_i is in $\mathrm{Stab}_G^{<\rho}(A)$ by assumption, so $d_A(\mathrm{id}_G, g_i) < \rho/2$. Hence, for

$i, j \in \{1, \dots, 3$ we have

$$d_A(g_i, g_j) < \rho \text{ and } \mu_G(g_i A \cap g_j A) > 0.$$

Applying the preceding Corollary(ii), we get σ-compact $A''' \subseteq A$ with $\mu_G(A'''H) \neq \varnothing$ such that for each $a \in A'''H$, we have

$$\mu_G(g_i A \cap g_j A) = \mu_{G/H}(\pi A)|\zeta_{g_i^{-1}a} - \zeta_{g_j^{-1}a}| \text{ for } i, j \in \{1, 2, 3\}.$$

The desired conclusion immediately follows. □

6.8.3 Almost linear pseudometric from near minimal expansions

In this subsection, G is always a *connected compact* Lie group. Let H be a closed subgroup of G, and H is isomorphic to the one dimension torus $\mathbb{T}$. Throughout the subsection, $A, B \subseteq G$ are σ-compact subsets such that

$$\kappa/2 < \mu_G(A) < 2\kappa \text{ and } \mu_G(B) = \kappa,$$

and (A, B) is $\eta\kappa$-nearly minimally expanding for some sufficiently small constant $\eta > 0$, that is

$$\mu_G(AB) \leq \mu_G(A) + \mu_G(B) + \eta\kappa.$$

In this section, we assume $\eta < 10^{-100}$. We did not try to optimise η, so it is very likely that by a more careful computation, one can make η much larger. But we believe this method does not allow η to be very close to 1.

By Fact 6.14, let $\tau = 2$, and $c = c(\tau)$ be the constant obtained from the theorem. In this subsection, we consider the case when

$$\max\{\mu_H(A \cap aH), \mu_H(B \cap Hb)\} < c$$

for all $a, b \in G$. The proofs in this section is more involved compared to the equality case, and the main difficulty is to control the error term coming from the near minimally expansion pair. For the readers who do not care the exact quantitative bound on the error terms, one can always view η as an infinitesimal element, then one can use equalities to replace all the inequalities in the proofs by pretending to take the standard part, and apply the methods given in the previous section.

Towards showing that sets A and B behave rigidly, our next theorem shows that most of the non-empty fibers in A and B have the similar lengths, and the majority of them behaves rigidly fiberwise.

Theorem 6.81 (Near rigidity fiberwise)**.** *There is a continuous surjective group homomorphism $\chi : H \to \mathbb{T}$, two compact intervals $I, J \subseteq \mathbb{T}$ with*

$$\mu_{\mathbb{T}}(I) = \frac{\mu_G(A)}{\mu_{G/H}(\pi A)} \quad \textit{and} \quad \mu_{\mathbb{T}}(J) = \frac{\mu_H(B)}{\mu_{H\backslash G}(\widetilde{\pi} B)}.$$

σ-compact $A' \subseteq A$ and $B' \subseteq B$ with

$$\mu_{G/H}(\pi A') > 99\mu_{G/H}(\pi A)/100 \quad \textit{and} \quad \mu_{H\backslash G}(\widetilde{\pi} B') > 99\mu_{H\backslash G}(\widetilde{\pi} B)/100,$$

and a constant $\nu < 10^{-6}$ such that the following statements hold:

(i) $(1-\eta)\mu_{H\backslash G}(\widetilde{\pi} B) \le \mu_{G/H}(\pi A) \le (1+\eta)\mu_{H\backslash G}(\widetilde{\pi} B)$.

(ii) *For every a in $A'H$,*

$$(1-\nu)\frac{\mu_G(A)}{\mu_{G/H}(\pi A)} \le \mu_H(A \cap aH) \le (1+\nu)\frac{\mu_G(A)}{\mu_{G/H}(\pi A)},$$

and there is $\zeta_a \in \mathbb{T}$ with

$$\mu_H\big((A \cap aH) \triangle a\chi^{-1}(\zeta_a + I)\big) < \nu \min\Big\{\frac{\mu_G(A)}{\mu_{G/H}(\pi A)}, \frac{\mu_G(B)}{\mu_{H\backslash G}(\widetilde{\pi} B)}\Big\}.$$

(iii) *For every b in HB',*

$$(1-\nu)\frac{\mu_G(B)}{\mu_{H\backslash G}(\widetilde{\pi} B)} \leq \mu_H(B \cap Hb) \leq (1+\nu)\frac{\mu_G(B)}{\mu_{H\backslash G}(\widetilde{\pi} B)},$$

and there is $\widetilde{\zeta}_b \in \mathbb{T}$ with

$$\mu_H\big((B \cap Hb) \triangle \chi^{-1}(\widetilde{\zeta}_b(B) + J)b\big) < \nu \min\Big\{\frac{\mu_G(A)}{\mu_{G/H}(\pi A)}, \frac{\mu_G(B)}{\mu_{H\backslash G}(\widetilde{\pi})}\Big\}.$$

Proof. Without loss of generality, we assume that $\mu_{G/H}(\pi A) \geq \mu_{H\backslash G}(\widetilde{\pi} B)$. Let β be a constant such that $\beta < \kappa/800\mu_{H\backslash G}(\widetilde{\pi} B)$. Obtain $b^* \in G$ such that

$$\mu_H(B \cap Hb^*) \geq \mu_H(B \cap Hb) - \beta \text{ for all } b \in G,$$

and the fiber $B \cap Hb^*$ has at least the average length, that is

$$\mu_H(B \cap Hb^*) \geq \mathbb{E}_{b \in BH}\mu_H(B \cap Hb) = \frac{\mu_G(B)}{\mu_{H\backslash G}(\widetilde{\pi} B)}. \tag{6.19}$$

Set $\delta = 400\eta\kappa/\mu_{G/H}(\pi A)$. As $\eta < 10^{-100}$, we get $\nu < 10^{-6}$ such that

$$\delta < \nu\mu_G(A)/\mu_{G/H}(\pi A).$$

Set

$$N = \{a \in AH : \mathfrak{d}_H(A \cap aH, B \cap Hb^*) > \beta\}.$$

Note that N is measurable by Lemma 6.9. By Lemma 6.62 we have

$$\begin{aligned}&\mu_G\big(A(B\cap Hb^*)\big)\\ &= \int_N \mu_H\big((A\cap aH)(B\cap Hb^*)\big)\,\mathrm{d}\mu_G(a) + \int_{G\setminus N} \mu_H\big((A\cap aH)(B\cap Hb^*)\big)\,\mathrm{d}\mu_G(a).\end{aligned}$$

Since $A\cap aH$ is nonempty for every $a\in AH$, using Kemperman's inequality on H we have that $\mu_G(A(B\cap Hb^*))$ is at least

$$\int_N \big(\mu_H(A\cap aH)+\mu_H(B\cap Hb^*)+\delta\big)\,\mathrm{d}\mu_G(a) + \int_{G\setminus N} \big(\mu_H(A\cap aH)+\mu_H(B\cap Hb^*)\big)\,\mathrm{d}\mu_G(a).$$

Suppose we have $\mu_G(N) > \mu_{G/H}(\pi A)/400$. Therefore, by the choice of b^* we get

$$\begin{aligned}\mu_G\big(A(B\cap Hb^*)\big) &> \mu_G(A) + \frac{\delta\mu_{G/H}(\pi A)}{100} + \mu_H(B\cap Hb^*)\mu_{G/H}(\pi A)\\ &\geq \mu_G(A)+\mu_G(B)\frac{\mu_{G/H}(\pi A)}{\mu_{H\setminus G}(\widetilde{\pi}B)} + \eta\kappa.\end{aligned} \tag{6.20}$$

Since $\mu_{G/H}(\pi A) > \mu_{H\setminus G}(\widetilde{\pi}B)$, and $A(B\cap Hb^*)\subseteq AB$, we have

$$\mu_G(AB) > \mu_G(A)+\mu_G(B)+\eta\kappa.$$

This contradicts the assumption that (A,B) is $\eta\kappa$-nearly minimally expanding. From equation (6.20) we also get

$$\mu_{H\setminus G}(\widetilde{\pi}B) \leq \mu_{G/H}(\pi A) \leq \big(1+\eta\big)\mu_{H\setminus G}(\widetilde{\pi}B), \tag{6.21}$$

which proves (i).

From now on, we assume that $\mu_G(N)\leq \mu_{G/H}(\pi A)/400$. Since (A,B) is $\eta\kappa$-minimally expanding, by (6.20), we have

$$\mu_H(B\cap Hb^*)\mu_{H\backslash G}(\widetilde{\pi}B) \leq \mu_G(B)+\eta\kappa,$$

and this in particular implies that for every $b\in G$, we have

$$\mu_H(B\cap Hb) \leq \frac{\mu_G(B)}{\mu_{H\backslash G}(\widetilde{\pi}B)} + \frac{\eta\mu_G(B)}{\mu_{H\backslash G}(\widetilde{\pi}B)} + \eta < \big(1+\nu\big)\frac{\mu_G(B)}{\mu_{H\backslash G}(\widetilde{\pi}B)}.$$

Thus there is $Y\subseteq B$ with $\mu_G(HY)<\mu_{H\backslash G}(\widetilde{\pi}B)/400$ such that for every $b\in HY$,

$$\mu_H(B\cap Hb) \geq \frac{\mu_G(B)}{\mu_{H\backslash G}(\widetilde{\pi}B)} - 400\frac{\eta\mu_G(B)}{\mu_{H\backslash G}(\widetilde{\pi}B)} - 400\eta > \big(1-\nu\big)\frac{\mu_G(B)}{\mu_{H\backslash G}(\widetilde{\pi}B)}.$$

Next, we apply the similar argument to A. Let $\alpha<(\mu_G(A)-2\eta\kappa)/200\mu_{G/H}(\pi A)$, and choose a^* such that $\mu_H(A\cap a^*H)>\mu_H(A\cap aH)-\alpha$ for all $a\in AH$, and

$$\mu_H(A\cap a^*H)\geq \mathbb{E}_{a\in AH}\mu_H(A\cap aH)=\frac{\mu_G(A)}{\mu_{G/H}(\pi A)}.$$

Let $N'\subseteq HB$ such that for every b in N', $\mathfrak{d}_H(A\cap a^*H, B\cap Hb)<\delta$ minimally expanding. Hence we have

$$\begin{aligned}
&\mu_G\big((A\cap a^*H)B\big)\\
&=\int_{N'}\mu_H\big((A\cap a^*H)(B\cap Hb)\big)\,\mathrm{d}\mu_G(b)+\int_{G\backslash N'}\mu_H\big((A\cap a^*H)(B\cap Hb)\big)\,\mathrm{d}\mu_G(b)\\
&\geq \mu_G(B)+\mu_H(A\cap a^*H)\mu_{H\backslash G}(\widetilde{\pi}B)+\delta\mu_G(N') \qquad (6.22)\\
&\geq \mu_G(A)+\mu_G(B)-\frac{\mu_G(A)\eta\kappa}{\mu_G(A)+\eta\kappa}+\delta\mu_G(N').
\end{aligned}$$

By the fact that $\mu_G(AB)\geq\mu_G((A\cap a^*H)B)$ and (A,B) is $\eta\kappa$-nearly minimally expanding, we have that

$$\mu_G(N')\leq\frac{1}{200}\mu_{G/H}(\pi A)\leq\frac{1}{150}\mu_{H\backslash G}(\widetilde{\pi}B).$$

Now, by equation (6.22), and the choice of a^*, we have that for all $a \in AH$,

$$\mu_H(A \cap aH) \leq \frac{\mu_G(A)}{\mu_{G/H}(\pi A)} + \frac{\eta\mu_G(A)}{\mu_{H\backslash G}(\widetilde{\pi}B)} + \alpha < \big(1+\nu\big)\frac{\mu_G(A)}{\mu_{G/H}(\pi A)}.$$

Again by equation (6.22), there is $X \subseteq A$ with $\mu_G(XH) \leq \mu_{G/H}(\pi A)/200$, such that for every $a \in X$,

$$\mu_H(A \cap aH) \geq \frac{\mu_G(A)}{\mu_{G/H}(\pi A)} - 200\frac{\eta\mu_G(A)}{\mu_{H\backslash G}(\widetilde{\pi}B)} - 200\alpha \geq \big(1-\nu\big)\frac{\mu_G(A)}{\mu_{G/H}(\pi A)}.$$

Let $A' = A \cap (AH \setminus (XH \cup N'))$, and let $B' = B \cap (HB \setminus (HY \cup N))$. Then

$$\mu_G(A') \geq \frac{99}{100}\mu_{G/H}(\pi A), \quad \mu_G(B') \geq \frac{99}{100}\mu_{H\backslash G}(\widetilde{\pi}B),$$

Let a be in $A'H$ and b be in $B'H$. By our construction, the first parts of (ii) and (iii) are satisfied. Moreover, $(A \cap aH, B \cap Hb^*)$ and $(A \cap a^*H, B \cap Hb)$ are δ-minimally expanding pairs. By the way we construct A' and B', we have that $a^* \in A'$ and $b^* \in B'$. Recall that $\mu_H(A \cap aH), \mu_H(B \cap Hb) < \lambda$ for every $a, b \in G$. Therefore, by the inverse theorem on $\mathbb{T}$ (Fact 6.14), and Lemma 6.33, there is a group homomorphism $\chi : H \to \mathbb{T}$, and two compact intervals I_A, I_B in $\mathbb{T}$, with

$$\mu_{\mathbb{T}}(I_A) = \frac{\mu_G(A)}{\mu_{G/H}(\pi A)}, \quad \mu_{\mathbb{T}}(I_B) = \frac{\mu_G(B)}{\mu_{H\backslash G}(\widetilde{\pi}B)},$$

such that for every $a \in A'$ and $b \in B'$, there are elements $\zeta_a, \widetilde{\zeta}_b(B)$ in $\mathbb{T}$, and

$$\mu_H(A \cap aH \triangle a\chi^{-1}(\zeta_a + I_A)) < \delta, \quad \mu_H(B \cap Hb \triangle \chi^{-1}(\widetilde{\zeta}_b + I_B)b) < \delta,$$

as desired. □

The next corollary gives us an important fact of the structure of the projection of A on

G/H.

Corollary 6.82 (Global structure of AH). *Suppose $\mu_G(A) = \kappa$. Then for all $g \in \mathrm{Stab}_G^{\kappa/2}(A)$, we have*

$$\mu_G(AH \triangle gAH) \leq \mu_{G/H}(\pi A)\frac{1}{100}.$$

Proof. By Lemma 6.21, $(A \cup gA, B)$ is a $2\eta\kappa$-minimal expansion pair, and by Theorem 6.81, we have

$$\frac{1}{1+2\eta} < \frac{\mu_G(AH \cup gAH)}{\mu_{H\backslash G}(\widetilde{\pi}B)} < 1 + 2\eta.$$

On the other hand, since (A, B) and (gA, B) are $\eta\kappa$-nearly minimal expanding, we have

$$1/(1+\eta) < \frac{\mu_{G/H}(\pi A)}{\mu_{H\backslash G}(\widetilde{\pi}B)}, \frac{\mu_G(gAH)}{\mu_{H\backslash G}(\widetilde{\pi}B)} < 1 + \eta$$

Since $\eta < 10^{-100}$, we get the desised conclusion. □

Lemma 6.83. *Suppose $\mu_G(A) = \kappa$, $g_1, g_2 \in \mathrm{Stab}_G^{\kappa/4}(A)$, and $\chi : H \to \mathbb{T}$ and $I \subseteq \mathbb{T}$ are as in Theorem 6.81. Then there is a σ-compact $A'' \subseteq A$ with*

$$\mu_{G/H}(\pi A'') = 96\mu_{G/H}(\pi A)/100$$

such that for all $a \in A''H$, the following holds

(i) *there are $\zeta_{g_1^{-1}a}, \zeta_{g_2^{-1}a} \in \mathbb{T}$ such that for $i \in \{1, 2\}$;*

$$\mu_H\big((A \cap aH) \triangle a\chi^{-1}(\zeta_{g_i^{-1}a} + I)\big) < \frac{\nu\kappa}{\mu_{G/H}(\pi A)}.$$

(ii) *with any $\zeta_{g_1^{-1}a}, \zeta_{g_2^{-1}a}$ satisfying (i) and $\delta_{a,A}(g_1, g_2) = \zeta_{g_2^{-1}a} - \zeta_{g_1^{-1}a}$, we have*

$$d_A(g_1, g_2) \in \mu_{G/H}(\pi A)|\delta_{a,A}(g_1, g_2)| + I\big(2\nu\kappa\big).$$

Proof. Obtain $,\chi, A', I, J$ as in Theorem 6.81. Let $A'' \subseteq G$ be the σ-compact set

$$\{a \in A : g_1^{-1}a, g_2^{-1}a \in A'H\}.$$

It is easy to see that $\mu_{G/H}(\pi A'') > 98/100\mu_{G/H}(\pi A)$. Fix $a \in A''$. We then have

$$A \cap g_1^{-1}aH = g_1^{-1}a\chi^{-1}(\zeta_{g_1^{-1}a} + I) \text{ and } A \cap g_2^{-1}aH = g_2^{-1}a\chi^{-1}(\zeta_{g_2^{-1}a} + I).$$

Multiplying by g_1 and g_2 respectively, we get (i).

Now suppose further that $\mu_G(g_1A \cap g_2A) > 0$. Recall that $(g_1A \cap g_2A, B)$ is then a minimally expanding pair by Lemma 6.21. By Theorem 6.76(ii), for $\mu_{G/H}$-almost all $aH \in \pi(g_1A \cap g_2A)$, we have

$$(1-\nu)\frac{\mu_G(g_1A)}{\mu_{G/H}(\pi(g_1A))} \leq \mu_H(g_1A \cap aH) \leq (1+\nu)\frac{\mu_G(g_1A)}{\mu_{G/H}(\pi(g_1A))}$$

and

$$(1-2\nu)\frac{\mu_G(g_1A \cap g_2A)}{\mu_{G/H}(\pi(g_1A \cap g_2A))} \leq \mu_H(g_1A \cap g_2A \cap aH) = (1+2\nu)\frac{\mu_G(g_1A \cap g_2A)}{\mu_{G/H}(\pi(g_1A \cap g_2A))}.$$

Note that $\pi(g_1A \cap g_2A) \subseteq \pi(g_1A) \cap \pi(g_2A)$. However, by Lemma 6.76,

$$\mu_{G/H}(\pi(g_1A \cap g_2A)) = \mu_{H\backslash G}(\widetilde{\pi}B) = \mu_{G/H}(\pi(g_1A)) = \mu_{G/H}(\pi(g_2A)).$$

Combining with Lemma 6.77, we get

$$\mu_{G/H}(\pi(g_1A)\triangle\pi A) = 0 \quad\text{and}\quad \mu_{G/H}(\pi(g_1A \cap g_2A)\triangle\pi A) = 0.$$

Hence, shrinking the above A'' if necessary, we can arrange that for all $a \in A''H$,

$$(1-5\nu)\frac{\mu_G(g_1A)}{\mu_{G/H}(\pi A)} \leq \frac{\mu_G(g_1A)}{\mu_{G/H}(\pi A)}\mu_H(g_1A \cap aH) \leq (1+5\nu)\frac{\mu_G(g_1A)}{\mu_{G/H}(\pi A)}$$

and

$$(1-10\nu)\frac{\mu_G(g_1A \cap g_2A)}{\mu_{G/H}(\pi(A))} \leq \mu_H(g_1A \cap g_2A \cap aH) \leq (1+10\nu)\frac{\mu_G(g_1A \cap g_2A)}{\mu_{G/H}(\pi(A))}.$$

Finally, from (i) for all $a \in A''H$, we have

$$\mu_H(g_1A) - \mu_H(g_1A \cap g_2A \cap aH) \in \mu_{G/H}(\pi A)|\zeta_{g_1^{-1}a} - \zeta_{g_2^{-1}a}| + I\big(2\nu\mu_G(A)\big).$$

Recall that $d_A(g_1, g_2) = \mu_H(g_1A) - \mu_H(g_1A \cap g_2A \cap aH)$. Hence, (ii) is satisfied. □

We now show deduce property of the pseudometric d_A.

Proposition 6.84 (Almost linearity and path monotonicity of the pseudometric)**.** *Assume that $\mu_G(A) = \kappa$. Then we have the following:*

(i) *For all g_1, g_2, g_3 in $\mathrm{Stab}_G^{\kappa/2}(A)$, we have*

$$d_A(g_1, g_2) \in |\pm d_A(g_1, g_3) \pm d_A(g_2, g_3)| + I\big(6\alpha\mu_G(A)\big),$$

(ii) *Let $\mathfrak{g}$ be the Lie algebra of G, and let $\exp : \mathfrak{g} \to G$ be the exponential map. For every $X \in \mathfrak{g}$, either $d_A(\exp(Xt), \mathrm{id}_G) < \kappa/4$ for every t, or there is $t_0 > 0$ with $d_A(\exp(Xt_0), \mathrm{id}_G) \geq \kappa/4$ such that for every $t \in [0, t_0]$,*

$$\begin{aligned} &d_A(\exp(X(t+t_0)), \mathrm{id}_G) \\ &\in d_A(\exp(X(t+t_0)), \exp(Xt_0)) + d_A(\exp(Xt_0), \mathrm{id}_G) + I\big(12\nu\mu_G(A)\big). \end{aligned}$$

Proof. We first prove (i). Let χ and I be as in Theorem 6.81. Applying Lemma 6.83, we get $a \in AH$ and $\zeta_{g_1^{-1}a}, \zeta_{g_2^{-1}a}, \zeta_{g_1^{-1}a} \in \mathbb{T}$ such that for $i \in \{1, \dots, 3\}$, we have

$$\mu_H\big((A \cap aH) \triangle a\chi^{-1}(\zeta_{g_i^{-1}a} + I)\big) < \frac{\nu\kappa}{\mu_{G/H}(\pi A)},$$

and for $i, j \in \{1, \dots, 3\}$, we have

$$d_A(g_i, g_j) \in \mu_{G/H}(\pi A)|\delta_{a,A}(g_i, g_j)| + I\big(2\nu\kappa\big).$$

with $\delta_{a,A}(g_i, g_j) = \zeta_{g_j^{-1}a} - \zeta_{g_i^{-1}a}$. As $\delta_{a,A}(g_1, g_2) = \delta_{a,A}(g_1, g_3) + \delta_{a,A}(g_3, g_2)$, we get the desired conclusion.

Next, we prove (ii). Let $X \in \mathfrak{g}$, and suppose there is $t > 0$ such that

$$d_A(\exp(Xt), \mathrm{id}_G) \geq \kappa.$$

Using the continuity of $g \mapsto \mu_G(A \setminus gA)$ (Fact 6.4(vii)), we obtain $t_0 > 0$ such that t_0 the smallest positive real number with $d_A(\mathrm{id}_G, \exp(Xt_0)) \geq \kappa/10$. Fix $t \in [0, t_0]$, and set

$$g_0 = \exp(Xt_0) \text{ and } g = \exp(Xt).$$

Note that $gg_0 = g_0 g$ as g_0 and g are on the same one parameter subgroup of G. One can easily check that g_0, g, $g_0 g$ are in $\mathrm{Stab}_G^{\kappa/2}(A)$. Again, let χ and I be as in Theorem 6.81 and apply Lemma 6.83 to get $a \in AH$ and $\zeta_{g_i^{-1}a} \in \mathbb{T}$ for $g_i \in \{\mathrm{id}_G, g, g_0, gg_0\}$ such that

$$\mu_H\big((A \cap aH) \triangle (a\chi^{-1}(\zeta_{g_i^{-1}a} + I)\big) < \frac{\nu\kappa}{\mu_{G/H}(\pi A)}, \tag{6.23}$$

and for $g_i, g_j \in \{\mathrm{id}_G, g, g_0, gg_0\}$, we have

$$d_A(g_i, g_j) \in \mu_{G/H}(\pi A)|\delta_{a,A}(g_i, g_j)| + I\big(2\nu\kappa\big) \tag{6.24}$$

with $\delta_{a,A}(g_i, g_j) = \zeta_{g_j^{-1}a} - \zeta_{g_i^{-1}a}$. As $gg_0 = g_0 g$, we have

$$\delta_{a,A}(\mathrm{id}_G, g) + \delta_{a,A}(g, gg_0) = \delta_{a,A}(\mathrm{id}_G, g_0 g) = \delta_{a,A}(\mathrm{id}_G, g_0) + \delta_{a,A}(g_0, g_0 g) \tag{6.25}$$

Using (6.23), (6.24), and the fact that $d_A(\mathrm{id}_G, g_0) = d_A(g, gg_0)$, we get

$$\delta_{a,A}(g, gg_0) \in \pm\delta_{a,A}(\mathrm{id}_G, g_0) + I\big(6\alpha\mu_G(A)\big).$$

By a similar argument,$\delta_{a,A}(g_0, g_0 g) \in \pm\delta_{a,A}(\mathrm{id}_G, g) + I\big(6\alpha\mu_G(A)\big)$. Combining with (6.25), we get that $\delta_{a,A}(\mathrm{id}_G, gg_i)$ is in both

$$\delta_{a,A}(\mathrm{id}_G, g_0) \pm \delta_{a,A}(\mathrm{id}_G, g) + I\big(6\alpha\mu_G(A)\big)$$

and

$$\delta_{a,A}(\mathrm{id}_G, g) \pm \delta_{a,A}(\mathrm{id}_G, g_0) + I\big(6\alpha\mu_G(A)\big).$$

Using the fact that ν is very small, and considering all the four possibilites, we deduce

$$\delta_{a,A}(\mathrm{id}_G, gg_0) = \delta_{a,A}(\mathrm{id}_G, g_0) + \delta_{a,A}(\mathrm{id}_G, g) + I\big(12\alpha\mu_G(A)\big).$$

Applying (6.23) and (6.24) again, we get the desired conclusion. □

6.9 Proof of the main theorems

6.9.1 Minimal expansion pairs in noncompact groups

In this subsection, we prove Theorems 6.1 when G is noncompact. The next theorem is a restatement of Theorem 6.1(vi), which is the main result in this subsection.

Theorem 6.85 (Main theorem for noncompact groups)**.** *Suppose G is a connected unimod-*

ular noncompact group, μ_G is a Haar measure on G, and $A, B \subseteq G$ are σ-compact subsets of G with positive finite measures, such that

$$\mu_G(AB) = \mu_G(A) + \mu_G(B)$$

Then there is a continuous surjective group homomorphism $\chi : G \to \mathbb{R}$ with compact kernel, and compact intervals $I, J \subseteq \mathbb{R}$ with $\mu_G(A) = \mu_{\mathbb{R}}(I)$ and $\mu_G(B) = \mu_{\mathbb{R}}(J)$, such that

$$A \subseteq \chi^{-1}(I), \quad \textit{and} \quad B \subseteq \chi^{-1}(J).$$

Moreover, if A and B are compact, then $A = \chi^{-1}(I)$ and $B = \chi^{-1}(J)$.

Proof. By the Gleason–Yamabe Theorem (Fact 6.28), there is a connected compact normal subgroup H of G such that $L = G/H$ is a connected Lie group. By Fact 6.10, L is unimodular. Let $\pi : G \to L$ be the quotient map. Using Corollary 6.27, there are σ-compact subsets A', B' of L such that

$$\mu_G(A \triangle \pi^{-1}(A')) = 0 \quad \text{and} \quad \mu_G(B \triangle \pi^{-1}(B')) = 0,$$

and we still have $\mu_L(A'B') = \mu_L(A') + \mu_L(B')$.

When L is a simple Lie group, by the Iwasawa decomposition, $L = KAN$ where AN is a simply connected closed nilpotent group. Thus AN contains $\mathbb{R}$ as a closed subgroup, and so does L. When L is not simple, then L contains a connected closed normal subgroup H, and by Fact 6.10, H is unimodular, and of smaller dimension. Applying induction on dimension, we may assume H satisfies the statement of the theorem. For $g_1, g_2 \in L$, let

$$d_A(g_1, g_2) = \mu_L(A) - \mu_L(g_1 A \cap g_2 A).$$

Then by Proposition 6.37, d_A is a pseudometric on L, with radius $\mu_G(A)$. By Proposition 6.80, d_A is locally linear. Using Proposition 6.42, we have $\ker d_A$ is a compact normal subgroup of L, and $L/\ker d_A$ is isomorphic to $\mathbb{R}$ as topological groups. By the third isomorphism theorem (Fact 6.7), $\mathbb{R}$ is a quotient group of G, and the corresponding quotient map χ has a compact kernel. Applying Corollary 6.27 again, as well as the inverse theorem on $\mathbb{R}$ (Fact 6.15), there are I, J compact intervals of $\mathbb{R}$ such that

$$\mu_G(A \triangle \chi^{-1}(I)) = 0 \quad \text{and} \quad \mu_G(B \triangle \chi^{-1}(J)) = 0.$$

This also implies that $\mu_G(A) = \mu_{\mathbb{R}}(I)$ and $\mu_G(B) = \mu_{\mathbb{R}}(J)$.

Suppose $g \in A$ and $g \notin \chi^{-1}(I)$. Since I is compact, $\chi(g) \notin I$, and there is $\alpha > 0$ such that the distance between $\chi(g)$ and the nearest element in I is at least α. Thus

$$\mu_{\mathbb{R}}(\chi(g)\chi(B) \setminus \chi(A)\chi(B)) \geq \alpha.$$

and this implies that $\mu_G(gB \setminus \chi^{-1}(I)\chi^{-1}(J)) \geq \alpha$. Therefore,

$$\mu_G(AB) \geq \mu_G(\chi^{-1}(I)\chi^{-1}(J)) + \mu_G(gB \setminus \chi^{-1}(I)\chi^{-1}(J)) \geq \mu_G(A) + \mu_G(B) + \alpha,$$

and this contradicts the fact that (A, B) is minimally expanding. Hence we have $A \subseteq \chi^{-1}(I)$ and $B \subseteq \chi^{-1}(J)$ as desired. When both A and B are compact, we have $A = \chi^{-1}(I)$ and $B = \chi^{-1}(J)$ by the compactness. $\square$

We remark that the same argument almost works for compact groups when (A, B) is a minimal expansion pair, except that when we choose the closed subgroup H, we need to choose one such that we are in the toric transversal scenario. This can be done by using Theorem 6.73 (See Section 6.9.2).

6.9.2 Nearly minimal expansion pairs in compact groups

In this subsection, we prove the main theorems for compact groups. We first prove Theorem 6.3.

Proof of Theorem 6.3. Let $d > 0$ be an integer, let c be the real number fixed at the beginning of Section 8.3, let α be in Proposition 6.84. Let G be a connected compact simple Lie group of dimension at most d, and A is a compact subset of G of measure at most C, and $\mu_G(A^2) < (2+\eta)\mu_G(A)$, where $C = C(d)$ is the constant in Theorem 6.73, and η is the constant fixed in Section 8.3.

By Theorem 6.73, when A satisfies $\mu_G(A) < C$, there is a closed subgroup H which is isomorphic to $\mathbb{T}$, such that for all $g \in G$, $\mu_H(gA \cap H) < c$. We fix such a closed subgroup H of G. Let

$$d_A(g_1, g_2) = \mu_G(A) - \mu_G(g_1 A \cap g_2 A).$$

By Proposition 6.37, d_A is a pseudometric. Since $\mu_G(A^2) < (2+\eta)\mu_G(A)$, Proposition 6.84 shows that d_A is a $6\alpha\mu_G(A)$-linear pseudometric, and it is $12\alpha\mu_G(A)$-monotone along each one parameter subgroup. By Proposition 6.44, d_A is globally $96\alpha\mu_G(A)$-monotone. Let $\gamma = 96\alpha\mu_G(A)$. Then d_A is γ-monotone γ-linear, and of radius $\rho = \mu_G(A)$. Theorem 6.61 thus implies that there is a continuous surjective group homomorphism mapping G to $\mathbb{T}$, and this contradicts the fact that G is simple. $\square$

Now we are going to prove the inverse theorem. In Proposition 6.36, given a continuous surjective group homomorphism from G to $\mathbb{T}$, we obtain a structural characterization with further assumption that the images of both A and B are small. The next lemma says that, with the homomorphism obtained from almost linear pseudometric in Sections 7 and 8, both A and B should have small image.

Lemma 6.86. *Let G be a connected compact groups, $\pi : G \to \mathbb{T}$ is a surjective continuous group homomorphism, and A, B are σ-compact subgroups of G with $\mathfrak{d}_G(A, B) = \delta$ for some*

$\delta < \min\{\mu_G(A), \mu_G(B)\}$, and $\max\{\mu_G(A), \mu_G(B)\} < 1/100$. Suppose for every $g \in \ker \pi$ with $\mu_G(A \backslash gA) < \mu_G(A)/16$, we have $\mu_G(A \backslash gA) < \mu_G(A)/32$. Then $\mu_{\mathbb{T}}(\pi(A)) + \mu_{\mathbb{T}}(\pi(B)) < 1/5$.

Proof. Let $H = (\ker \pi)_0$. As G is compact, we will have $G/H \cong \mathbb{T}$ as topological groups. For every set X in G, and interval $I \subseteq [0, 1]$, we define

$$\pi X_I = \{g \in G : \mu_H(g^{-1}X \cap H) \in I\}.$$

It is clear that $\pi X_I \subseteq XH$. Since G is compact and H is normal, we also have $XH = HX$ and $\mu_H(X \cap gH) = \mu_H(X \cap Hg)$.

Assume first that we have $\sup_g \mu_H(B \cap Hg) > 1/2$. Choose b_0 such that $\mu_H(B \cap Hb_0) > 1/2$. We claim that $\mu_G(\pi A_{[0,1/2)}) \leq 2\mu_G(B) + 2\delta$. Otherwise, applying Kemperman's inequality on H we have

$$\mu_G(AB) \geq \mu_G(A(B \cap Hb_0)) > \mu_G(A) + \mu_G(B) + \delta,$$

this contradicts that $\mathfrak{d}_G(A, B) \leq \delta$. Similarly we also have $\mu_G(\pi A_{[1/2,1]}) \leq 2\mu_G(A) + 2\delta$, as otherwise

$$\mu_G(AB) \geq \frac{1}{2}\mu_G(\pi A_{[1/2,1]}) + \mu_G(B) > \mu_G(A) + \mu_G(B) + \delta,$$

contradiction. Assume we also have $\sup_g \mu_H(A \cap gH) > 1/2$. Then using the same argument, we conclude that $\mu_G(\pi B_{[0,1/2)}) < 2\mu_G(A) + 2\delta$, and $\mu_G(\pi B_{[1/2,1]}) \leq 2\mu_G(B) + 2\delta$. Thus we have

$$\mu_{\mathbb{T}}(\pi(A)) + \mu_{\mathbb{T}}(\pi(B)) < 4\mu_G(A) + 4\mu_G(B) + 8\delta < 1/5.$$

Next, we may assume that $\sup_g \mu_H(A \cap gH) \leq 1/2$. Note that

$$\frac{\mu_G(AH)}{\mu_G(A)} > \frac{1}{\sup_g \mu_H(A \cap gH)}.$$

Hence by Lemma 6.67, for every $\ell > 1/2$, there is $h \in H$ such that $\mu_G(A \cap hA) = \ell \mu_G(A)$, and in particular, there is $h \in H$ with

$$\frac{15\mu_G(A)}{16} < \mu_G(A \cap hA) < \frac{31\mu_G(A)}{32},$$

which contradicts the assumption.

Finally, let us consider the case when $\sup_g \mu_H(B \cap Hg) < 1/2$. From the above argument, we may also assume that $\sup_g \mu_H(A \cap gH) \geq 1/2$. This implies that $\mu_G(\pi B_{[0,1/2)}) < 2\mu_G(A) + 2\delta$, and in particular

$$\mathbb{E}_{g \in HB} \mu_H(B \cap Hg) > \frac{\mu_G(B)}{2\mu_G(A) + 2\delta}.$$

Hence, we have

$$\mu_G(\pi A_{[0,1/2)}) \leq \frac{(\mu_G(B) + \delta)(2\mu_G(A) + 2\delta)}{\mu_G(B)} < 4\mu_G(A) + 4\delta.$$

Therefore,

$$\mu_{\mathbb{T}}(\pi(A)) + \mu_{\mathbb{T}}(\pi(B)) < 8\mu_G(A) + 6\delta < 1/5,$$

as desired. $\square$

With all tools in hand, we are going to prove the following theorem, which is a restatement of Theorem 6.2 and Theorem 6.1(v) for compact groups.

Theorem 6.87 (Main theorem for compact groups)**.** *Let G be a connected compact group, and A, B be compact subsets of G with positive measure. Set*

$$s = \min\{\mu_G(A), \mu_G(B), 1 - \mu_G(A) - \mu_G(B)\}.$$

Given $0 < \varepsilon < 1$, there is a constant $K = K(s)$ does not depends on G, such that if $\delta < K\varepsilon$ and

$$\mu_G(AB) < \mu_G(A) + \mu_G(B) + \delta \min\{\mu_G(A), \mu_G(B)\}.$$

Then there is a surjective continuous group homomorphism $\chi : G \to \mathbb{T}$ together with two compact intervals $I, J \in \mathbb{T}$ with $\mu_{\mathbb{T}}(I) = \mu_G(A)$, $\mu_{\mathbb{T}}(J) = \mu_G(B)$, and

$$\mu_G(A \triangle \chi^{-1}(I)) < \varepsilon \mu_G(A), \quad \mu_G(B \triangle \chi^{-1}(J)) < \varepsilon \mu_G(B).$$

Moreover, when $\mu_G(AB) < 1$ and

$$\mu_G(AB) = \mu_G(A) + \mu_G(B),$$

we have $A \subseteq \chi^{-1}(I)$ and $B \subseteq \chi^{-1}(J)$.

Proof. Given $\varepsilon > 0$ and s defined in the statement of the theorem. Let $\delta > 0$ to be determined later. Suppose $\mu_G(A) < \mu_G(B)$, and $\mathfrak{d}_G(A, B) \leq \delta \mu_G(A)$. By Lemma 6.21, there is a constant K_1 only depending on s such that there are small σ-compact sets A_1, B_1 in G, such that $\mu_G(A_1) = \mu_G(B_1) < \tau$,

$$\mu_G(A_1 B_1) < \mu_G(A_1) + \mu_G(B_1) + K_1 \delta \min \mu_G(A_1),$$

and both $\mathfrak{d}_G(A_1, B)$ and $\mathfrak{d}_G(A, B_1)$ are at most $K_1 \delta \mu_G(A_1)$, where τ is the constant from Proposition 6.32. Again using Proposition 6.32, there is a constant $d > 0$, and a connected compact subgroup H of G, such that $L = G/H$ is a Lie group of dimension at most d, and we can find σ-compact sets A_2, B_2 in L, such that $\mu_L(A_2) \leq \mu_L(B_2)$,

$$\max\{\mu_G(A_1 \triangle \pi^{-1}(A_2)), \mu_G(B_1 \triangle \pi^{-1}(B_2))\} < 3K_1 \delta \mu_G(A_1), \tag{6.26}$$

and

$$\mu_L(A_2B_2) < \mu_L(A_2) + \mu_L(B_2) + 7K_1\delta\mu_G(A_1) < \mu_L(A_2) + \mu_L(B_2) + 7K_1\delta\mu_L(A_2).$$

Now we apply Lemma 6.21 again. Then there is a constant K_2 only depending on s, and two small σ-compact sets A_3, B_3 in L, such that $\mu_L(A_3) = \mu_L(B_3) < S$, and all of $\mathfrak{d}_L(A_3, B_3)$, $\mathfrak{d}_L(A_3, B_2)$, and $\mathfrak{d}_L(A_2, B_3)$ are at most $K_1K_2\delta\mu_L(A_3)$, where S is the constant from Proposition 6.72.

Let c be the constant fixed in the beginning of Section 8.3. By Theorem 6.73, there is a closed subgroup T of L, such that T is isomorphic to $\mathbb{T}$, and for every $g \in L$, we have

$$\max\{\mu_T(A \cap gT), \mu_T(B \cap Tg)\} < c.$$

Now we are in the toric transversal scenario. Similarly as what we did in the proof of Theorem 6.3, we define the pseudometric d_A, and from Proposition 6.84 and Proposition 6.44, when $K_1K_2\delta < \eta$ where η is from Proposition 6.84, we obtain a γ-linear γ-monotone pseudometric, where $\gamma = 96\alpha$, and α is also from Proposition 6.84. Therefore, Theorem 6.61 gives us a surjective continuous group homomorphism $\phi : L \to \mathbb{T}$, such that for every $g \in \ker\phi$, we have $\mu_L(A_3 \setminus gA_3) < \mu_L(A_3)/10$.

By the third isomorphism theorem (Fact 6.7), $\mathbb{T}$ is also a quotient group of G. It remains to determine the structure of A and B. By Lemma 6.86, we have $\mu_{\mathbb{T}}(\phi(A_3)) + \mu_{\mathbb{T}}(\phi(B_3)) < 1/5$. Then by Proposition 6.36, there are compact intervals I_3, J_3 in $\mathbb{T}$, such that

$$\mu_{\mathbb{T}}(I_3) = \mu_L(A_3) \quad \text{and} \quad \mu_{\mathbb{T}}(J_3) = \mu_L(B_3),$$

and $\mu_L(A_3 \triangle \phi^{-1}(I_3)) < K_0K_1K_2\delta\mu_L(A_3)$, $\mu_L(B_3 \triangle \phi^{-1}(J_3)) < K_0K_1K_2\delta\mu_L(A_3)$. Thus by

Lemma 6.33, there are compact intervals I_2, J_2 in $\mathbb{T}$ with

$$\mu_{\mathbb{T}}(I_2) = \mu_L(A_2) \quad \text{and} \quad \mu_{\mathbb{T}}(J_2) = \mu_L(B_2),$$

and

$$\mu_L(A_2 \triangle \phi^{-1}(I_2)) < (2K_0K_1K_2 + 45)\delta\mu_L(A_2),$$
$$\mu_L(B_2 \triangle \phi^{-1}(J_2)) < (2K_0K_1K_2 + 45)\delta\mu_L(A_2).$$

Let $\chi = \pi \circ \phi$. By (6.26) and Lemma 6.33 again, there are intervals I' and J' in $\mathbb{T}$, such that

$$\mu_{\mathbb{T}}(I') = \mu_G(A) \quad \text{and} \quad \mu_{\mathbb{T}}(J') = \mu_G(B),$$

and

$$\mu_G(A \triangle \chi^{-1}(I')) < (4K_0K_1K_2 + 90)\delta\mu_G(A),$$
$$\mu_G(B \triangle \chi^{-1}(J')) < (4K_0K_1K_2 + 90)\delta\mu_G(A) \leq (4K_0K_1K_2 + 90)\delta\mu_G(B).$$

Note that all K_0, K_1, and K_2 only depend on s, then one can take

$$\delta = \min\left\{\frac{\varepsilon}{4K_0K_1K_2 + 90}, \frac{\eta}{2K_0K_1K_2}\right\}.$$

Finally, we consider the case when $\delta = 0$, that is, $\mu_G(AB) = \mu_G(A) + \mu_G(B)$. The proof follows the same argument, by replacing Proposition 6.84 with Proposition 6.80 to construct the locally linear pseudometric, and by replacing Proposition 6.84 and Theorem 6.61 by Proposition 6.42. Proposition 6.44 is not needed in this case, as in this case, in order to construct a homomorphism to $\mathbb{T}$, the monotonicity of the pseudometric is not needed. $\square$

Chapter 7

A nonabelian Brunn–Minkowski inequality

Henstock and Macbeath [90] asked in 1953 whether the Brunn–Minkowski inequality can be generalized to nonabelian locally compact groups; questions in the same line were also asked by Hrushovski [101], McCrudden [132], and Tao [160]. We obtain here such an inequality and prove that it is sharp for helix-free locally compact groups, which includes real linear algebraic groups, Nash groups, semisimple Lie groups with finite center, solvable Lie groups, etc. The proof follows an induction on dimension strategy; new ingredients include an understanding of the role played by maximal compact subgroups of Lie groups, a necessary modified form of the inequality which is also applicable to nonunimodular locally compact groups, and a proportionated averaging trick. This chapter is based on joint work with Tran and Zhang [106].

7.1 Introduction

7.1.1 Background

Let μ be the usual Lebesgue measure on $\mathbb{R}^d$, let X and Y be nonempty and compact subsets of $\mathbb{R}^d$, and set $X+Y := \{x+y : x \in X, y \in Y\}$. The Brunn–Minkowski inequality says that

$$\mu(X+Y)^{1/d} \geq \mu(X)^{1/d} + \mu(Y)^{1/d}. \tag{7.1}$$

For fixed $\mu(X)$ and $\mu(Y)$, the inequality provides us with the minimum value of $\mu(X+Y)$ which is obtained, for example, when X, Y, and $X+Y$ are d-dimensional hypercubes with side length $\mu(X)^{1/d}$, $\mu(Y)^{1/d}$, and $\mu(X)^{1/d} + \mu(Y)^{1/d}$, respectively.

Under the further assumption that X and Y are convex, the inequality in an equivalent form was proven by Brunn [38] in 1887. In the celebrated Geometrie der Zahlen (Geometry of Numbers) [136] published in 1896, Minkowski introduced the current form of the inequality and established that the equality happens if and only if X and Y are homothetic convex sets. Removing the convexity assumption was done by Lyusternik [128] in 1935. However, his proof that the same condition for equality still holds was seen to contain some flaws, a situation eventually corrected by Henstock and Macbeath [90] in 1953. The Brunn–Minkowski inequality is widely considered a cornerstone of convex geometry. See [72] for an excellent survey on its numerous generalizations and applications.

In this chapter, we consider the problem of generalizing the Brunn–Minkowski inequality to a locally compact group G. Here, up to a multiplication by positive constants, we have a unique left Haar measure μ generalizing the Lebesgue measure in $\mathbb{R}^d$; see Section 5.2 for the precise definitions.

We temporarily further assume that μ is also invariant under right translations. Such G is called *unimodular*. This assumption holds when $G = \mathbb{R}^d$ and in many other situations (e.g, when G is compact, discrete, a nilpotent Lie group, a semisimple Lie group, etc). Set $XY = \{xy : x \in X, y \in Y\}$ for nonempty compact $X, Y \subseteq G$. The translation invariance property of μ implies that

$$\mu(XY) \geq \max\{\mu(X), \mu(Y)\}$$

and should intuitively be even larger, hinting at a meaningful generalization of the Brunn–Minkowski inequality to this setting. This will be shown to be the case.

For an arbitrary locally compact group G, μ might no longer be right invariant. Hence, we still have $\mu(XY) \geq \mu(Y)$, but we might have $\mu(XY) < \mu(X)$. By a result by Macbeath [130] in 1960, the trivial inequality $\mu(XY) \geq \mu(Y)$ for nonunimodular G is already sharp in the sense that for any $\alpha, \beta, \varepsilon > 0$, there are nonempty compact $X, Y \subseteq G$ with

$$\mu(X) = \alpha, \mu(Y) = \beta, \text{ and } \mu(XY) < \mu(Y) + \varepsilon.$$

We will later see in this chapter that there is still a meaningful generalization of the Brunn–Minkowski inequality involving both μ and a right Haar measure ν. Surprisingly, it turns out that if one only cares about unimodular cases, the nonunimodular cases are still needed for our proof. We will keep the settings and notations of this paragraph throughout the rest of the chapter.

The problem of generalizing the Brunn–Minkowski inequality was proposed in 1953 by Henstock and Macbeath [90]; different variations of this problem were also later suggested by Hrushovski [101], by McCrudden [132], and by Tao [160]. In the direction of the intuition described earlier, Kemperman [112] showed in 1964 that $\mu(XY) \geq \mu(X) + \mu(Y)$ when G is connected, unimodular and noncompact. Even more important for us is the followin generalization to all connected noncompact locally compact groups, which reads

$$\frac{\nu(X)}{\nu(XY)} + \frac{\mu(Y)}{\mu(XY)} \leq 1.$$

While applicable to all locally compact groups, Kemperman's inequalities are not sharp even for $\mathbb{R}^2$ giving a weaker conclusion than the Brunn–Minkowski inequality. The most definite result toward the correct lower bound was obtained by McCrudden [132] in 1969. In effect, he showed that when G is a unimodular solvable Lie group of dimension d, and m is the dimension of the maximal compact subgroup, we have

$$\mu(XY)^{1/(d-m)} \geq \mu(X)^{1/(d-m)} + \mu(Y)^{1/(d-m)}.$$

The above differs from McCrudden's original statement in that m was defined using an inductive idea in [132]; the current form is more suitable to get the later generalization and to show that it is indeed sharp. A number of special cases of this result were rediscovered by Gromov [84], by Hrushovski [102], by Leonardi and Mansou [122], and by Tao [160]. Sharpness for nilpotent groups was essentially proven by Monti [138]; see also Tao [160].

7.1.2 Statement of main results

Suppose G is Lie group with connected component G_0. Following Levi decomposition (Fact 5.22), we have an exact sequence of Lie groups

$$1 \to Q \to G_0 \to S \to 1$$

where Q is solvable and S is semisimple. It is known that the center $Z(S)$ is a discrete abelian group of finite rank h; see Facts 5.26 and 5.27. We call h the **helix dimension** of G. As an example, $\mathrm{SL}_2(\mathbb{R})$ has helix dimension 0 while its universal cover has helix dimension 1. If $h = 0$, equivalently S has finite center, we say that G is **helix-free**. Real linear algebraic groups and more generally, Nash groups (equivalently, semialgebraic Lie groups or groups definable in the field of real numbers) are helix free; see [10, Lemma 4.5] and the subsequent discussion in the same paper. Our first main results is a generalization of Brunn–Minkowski inequality to Lie groups whose exponent will be seen to be sharp for helix-free Lie groups:

Theorem 7.1. *Suppose G is a Lie group, μ is a left Haar measure, ν is a right Haar measure, the dimension of G is d, the maximal dimension of a compact subgroup of G is m, the helix dimension of G is h, and X, Y are compact subsets of G with positive measure. Then*

$$\frac{\nu(X)^{1/(d-m-h)}}{\nu(XY)^{1/(d-m-h)}} + \frac{\mu(Y)^{1/(d-m-h)}}{\mu(XY)^{1/(d-m-h)}} \leq 1; \tag{7.2}$$

the left-hand-side is interpreted as $\max\{\nu(X)/\nu(XY), \mu(Y)/\mu(XY)\}$ *if* $d-m-h=0$. *In particular, if G is unimodular, then* $\mu(XY)^{\frac{1}{d-m-h}} \geq \mu(X)^{\frac{1}{d-m-h}} + \mu(Y)^{\frac{1}{d-m-h}}$.

Now consider an arbitrary locally compact group G. Using the Gleason–Yamabe Theorem

(Fact 5.10), one can choose an open subgroup G' of G and a normal compact subgroup H of G' such that G'/H is a Lie group. It is shown in Proposition 7.13 that

$$n = \dim(G'/H) - \max\{\dim(K) : K \text{ is a compact subgroup of } G'/H\}$$

is independent of the choice of G' and H satisfying the above properties. We call n the **noncompact Lie dimension** of G. Let Q be the radical (i.e, the maximal connected closed solvable normal subgroup, see Fact 5.21) of G'/H. Note that $(G'/H)_0/Q)$ has discrete center $Z((G'/H)_0/Q$ by Facts 5.26 and 5.27. We call

$$h = \mathrm{rank}(Z((G'/H)_0/Q))$$

the **helix dimension** of G. We will also show that the helix dimension h of G'/H is independent of the choice of G' and H in Proposition 7.13. Our second main result reads:

Theorem 7.2. *Suppose G is a locally compact group with noncompact Lie dimension n and helix dimension h, μ is a left Haar measure, ν is a right Haar measure, and X, Y are compact subsets of G with positive measure. Then*

$$\frac{\nu(X)^{1/(n-h)}}{\nu(XY)^{1/(n-h)}} + \frac{\mu(Y)^{1/(n-h)}}{\mu(XY)^{1/(n-h)}} \leq 1;$$

the left-hand-side is interpreted as $\max\{\nu(X)/\nu(XY), \mu(Y)/\mu(XY)\}$ *when* $n - h = 0$*. In particular, if G is unimodular, then* $\mu(XY)^{\frac{1}{n-h}} \geq \mu(X)^{\frac{1}{n-h}} + \mu(Y)^{\frac{1}{n-h}}$.

When G is as in Theorem 7.1, the noncompact Lie dimension n is simply $d - m$, so Theorem 7.2 is a generalization of Theorem 7.1. On the other hand, Theorem 7.2 is equally applicable to totally disconnected locally compact groups, which are the polar opposite of Lie groups.

Our last main result tells us that when G is helix-free, the exponent $1/(n - h) = 1/n$ in Theorem 7.1 and Theorem 7.2 are sharp even when we assume further that $X = Y$. As usual in the current setting, we write X^k for the k-fold product of X.

Theorem 7.3. *Suppose G is a locally compact group with noncompact Lie dimension n, μ is a left Haar measure, and ν is a right Haar measure. Then*

1. *When $n = 0$, there is a compact set X with positive left and right measure in G such that $\mu(X^2) = \mu(X)$ and $\nu(X^2) = \nu(X)$.*

2. *When $n > 0$, for every $\varepsilon > 0$, there is a compact set X with positive left and right measure in G such that*
$$\frac{\nu(X)^{\frac{1}{n}-\varepsilon}}{\nu(X^2)^{\frac{1}{n}-\varepsilon}} + \frac{\mu(X)^{\frac{1}{n}-\varepsilon}}{\mu(X^2)^{\frac{1}{n}-\varepsilon}} > 1.$$

As a corollary, if G is unimodular with $n > 0$, for every $\varepsilon' > 0$, there is a compact set X in G such that $\mu(X^2) < (2^n + \varepsilon')\mu(X)$.

The upper bound given in Theorem 7.3 matches the lower bound given in Theorem 7.2 when the group is helix-free, that is a group has helix dimension 0, which essentially means the semisimple part of the group has finite center. Hence, for these groups, our theorems resolve the problem of generalizing the Brunn–Minkowski inequality, which was suggested by Henstock and Macbeath [90], by Hrushovski [101], by McCrudden [130], and by Tao [160].

We believe that the exponent in Theorem 7.3 should be correct for all locally compact groups, which is made precise by the following conjecture:

Conjecture 7.4 (Nonabelian Brunn–Minkowski Conjecture)**.** *Suppose G is a locally compact group with noncompact Lie dimension n, μ is a left Haar measure, ν is a right Haar measure, and X, Y are compact subsets of G with positive measure. Then*

$$\frac{\nu(X)^{1/n}}{\nu(XY)^{1/n}} + \frac{\mu(Y)^{1/n}}{\mu(XY)^{1/n}} \leq 1;$$

the left-hand-side is interpreted as $\max\{\nu(X)/\nu(XY), \mu(Y)/\mu(XY)\}$ *when $n = 0$.*

We remark that, the exponent in the inequality obtained in Theorem 7.2, if is not sharp, still has the correct order of magnitude, as the helix dimension h of G is always at most $n/3$, where n is the noncompact Lie dimension of G; see Corollary 7.20.

The next result shows that one can reduce Conjecture 7.4 to all simply connected simple Lie groups. Unexpectedly, the hardest remaining cases are what one might initially regard to be the simplest cases.

Theorem 7.5. *Suppose the nonabelian Brunn–Minkowski conjecture holds for all simply connected simple Lie groups, then it holds for all locally compact groups.*

In the statements of our main results, we require the sets X and Y to be compact. The reason is that, when X and Y are just measurable, the set XY may not be measurable. We remark that by using the regularity property of Haar measure, the conclusions in our main theorems still hold for measurable X and Y if we replace $\mu(XY)$ and $\nu(XY)$ by inner Haar measures.

The results of this chapter continue a line of work by the first two authors [105] on small measure expansions in locally compact groups. Through classifying groups G and compact subsets X and Y of G with nearly minimal expansion, it is shown there that when G is a simple compact Lie group and $\mu(X)$ sufficiently small,

$$\mu(X^2) > (2+c)\mu(X)$$

for a positive constant c. This can be seen as a continuous analog of the expansion gap results. For noncompact simple Lie groups, Theorem 7.1 provides a significant strengthening counterpart where we have $\mu(X^2) \geq 4\mu(X)$. As we will see later, some of the techniques used in this chapter are further developments from techniques used in [105].

The equality for Theorems 7.1 and 7.2 can happen for $\mathbb{R}^d$, but might be impossible for general G. In fact, from McCrudden's result [134], the equality cannot happen even when G is the Heisenberg group. It would also be interesting to understand when equality nearly happens and develop a theory similar to that of Christ, Figalli, and Jerison [42, 65, 66] for $\mathbb{R}^d$.

Like the Brunn–Minkowski inequality for $\mathbb{R}^d$, our results do not rely on the normalization of Haar measures. However, by fixing a Haar measure μ on a unimodular group G, it would be interesting to determine the value of

$$\min\{\mu(XY) : X, Y \subseteq G \text{ are compact}, \mu(X) = \alpha, \mu(Y) = \beta\},$$

for given $\alpha, \beta \in \mathbb{R}^{>0}$, and to classify the situations where the equality happens. We do not pursue this question here.

7.1.3 Overview of the proof

In this subsection, we discuss the idea of the proof of the main results and the organization of the chapter. For expository purpose, we restrict our attention to helix-free locally compact groups, where we can fully prove Conjecture 7.4. The proof of the full versions of Theorems 7.1 and 7.2 requires a more involved discussion on the helix dimension, which is developed in Section 7.2.

In the current situation, for all our three theorems, the exponent of the inequalities are controlled by n of G instead of just its topological dimension d as in the simpler versions for $\mathbb{R}^d$. Recall that, for a Lie group G, $n = d - m$ where m the maximum dimension of a compact subgroup of G. The proof of Theorem 7.3 explains the critical role of m: Our construction is essentially a small neighborhood of a compact subgroup of G having maximal dimension, see Figure 7.1. One may then naturally conjecture that the above is the best we can do. Theorems 7.1 and 7.2 confirm this intuition for helix-free groups.

To motivate our proofs of Theorems 7.1 and 7.2, we first recall some proofs of the known cases of the Brunn–Minkowski inequality. Over $\mathbb{R}^d$, the usual strategy is to induct on dimensions. This is generalized by McCrudden to obtain the following "unimodular exponent

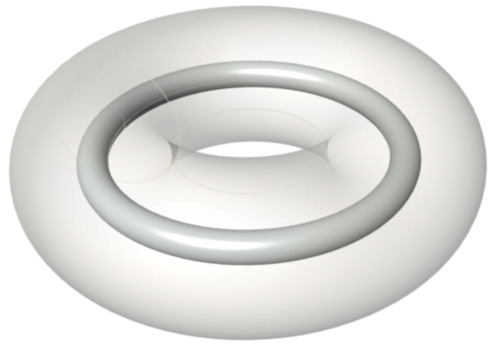


Figure 7.1: Let $G = \mathrm{SL}(2,\mathbb{R})$ (the open region bounded by the outer torus), and let $K = \mathrm{SO}(2,\mathbb{R})$ be the maximal compact subgroup of G. If we take X to be a small closed neighborhood of K (closed region bounded by the inner torus), Theorem 7.3 says when X is sufficiently small, $\mu_G(X^2)$ will be very close to $4\mu_G(X)$ instead of $8\mu_G(X)$, although G has topological dimension 3.

splitting" result: Given an exact sequence of *unimodular* locally compact groups

$$1 \to H \to G \to G/H \to 1,$$

if H and G/H satisfy Brunn–Minkowski inequalities with exponents $1/n_1$ and $1/n_2$, respectively, then the group G satisfies a Brunn–Minkowski inequality with exponent $1/(n_1+n_2)$.

McCrudden's proof of the above result can be seen as the following "spillover" argument: For each g in G, we call $X \cap gH$ a *fiber* of X, and refer to the size of $g^{-1}X \cap H$ in H as its *length*. Let $\pi : G \to G/H$ be the quotient map. We now partition X and Y each into N parts. Suppose $X = \bigcup_{i=1}^{N} X_i$ and $Y = \bigcup_{i=1}^{N} Y_i$, we require that the images under π of the X_i's are pairwise disjoint, the shortest fiber-length in each X_i is at least the longest fiber-length in X_{i-1}, and likewise for the Y_i's.

The induction hypotheses, i.e., the Brunn–Minkowski inequalities, in H and G/H give us a lower bound l_N on fiber-lengths in X_NY_N and a lower bound w_N on the size of $\pi(X_NY_N)$ in G/H. Their product l_Nw_N is a lower bound for $\mu(X_NY_N)$. Next we consider $(X_{N-1} \cup X_N)(Y_{N-1} \cup Y_N)$. Again a lower bound l_{N-1} on fiber-lengths in this set and a lower bound w_{N-1} on the size of its image under π can be obtained from the induction hypotheses on

H and G/H. From our method, we have $l_{N-1} \le l_N$ and $w_{N-1} \ge w_N$. The $l_{N-1}w_{N-1}$ will be a weak lower bound for $\mu((X_{N-1} \cup X_N)(Y_{N-1} \cup Y_N))$ since the fibers in $X_N Y_N$ are "exceptionally long". Taking all of these into account, a stronger lower bound is

$$l_N w_N + l_{N-1}(w_{N-1} - w_N).$$

Repeating the above process and taking the limit $N \to \infty$ we have the "spillover" argument which enables McCrudden to obtain his result.

McCrudden applied this result to obtain the Brunn–Minkowski inequality for unimodular solvable groups with sharp exponents. A simpler proof of his result is given in Section 7.4 for completeness. In the proof of our main theorems, one important ingredient will be an exponent splitting result (that is a generalization of his).

McCrudden's method completely stops working when one is looking to prove Brunn–Minkowski for simple groups since there is no nontrivial closed normal subgroup to induct from. Next we explain how we overcome this main difficulty. Our method turns out to work also for semisimple groups in the same way and we will explain it in this more general setting.

Let us assume G is a connected semisimple Lie group with finite center (hence helix-free and automatically unimodular) and think about how we can prove the Brunn–Minkowski for it. One can consider the Iwasawa decomposition $G = KAN$ where K has a compact Lie algebra and $Q = AN$ is solvable and try to connect the Brunn–Minkowski of S to a similar property of Q. However, Q may not be unimodular in general. Let Δ_Q be the modular function on G. One can choose to compromise by choosing $Q' = \ker(\Delta_Q)$ that is unimodular and try to use the Brunn–Minkowski for Q' to prove the Brunn–Minkowski on G. This is indeed a good direction to go but along this direction one inevitably gives up on the sharp exponent $1/n$ and can at best prove a weaker inequality with the worse exponent $1/(n-1)$.

Because of this, it is necessary to formulate an inequality for nonunimodular groups that is a good analogue of (7.1). We propose the inequality (7.2), which seems to be new in the literature. To prove (7.2) for AN, we need a nonunimodular exponent splitting result for the exact sequence coming out from the modular function. It turns out that the spillover method can also be used to reduce the problem to the case where the modular function is almost constant on X and Y. We work this out in Section 7.5. In the next more involved step in the same section, we obtain an approximate version of McCrudden's result, which involves another use of the spillover method, to finish off the proof.

In the next crucial step, we prove that the Brunn–Minkowski for a semisimple G follows from (7.2) for the solvable AN. Our method was motivated by a recent paper [105] by the first two authors, which characterizes nearly minimal expansion sets. Over there, a key idea is to choose a fiber f uniformly at random in Y and uses Xf to estimate XY. For our current proof, we also choose two fibers f_X and f_Y randomly from X and Y, but with respect to two carefully chosen probability measures p_X and p_Y that are in general nonuniform. We show that by constructing p_X and p_Y based on the structural information of X and Y, $\mu(XY)$ can be estimated by the expected size of $f_X f_Y$ in AN, and the latter is well controlled by the Brunn–Minkowski inequality (7.2) for AN. This part is done in Section 7.6. It worth noting that in this case our inequality matches the upper bound construction when the semisimple group has a finite center.

With the above preparation, we can explain how we prove Brunn–Minkowski for a general helix-free Lie group G. Using reductions proved in Sections 7.4, 7.5, and 7.6, we can reduce the problem to the case where G is unimodular and connected. Such G can be decomposed into a semi-direct product of a unimodular solvable group Q and a semisimple group S via the Levi decomposition. We already know how to handle S from the discussion in Section 7.6. McCrudden's result can then be used to deal with Q and to deduce the desired inequality for G.

In many of our reductions, we have an exact sequence of groups $1 \to H \to G \to G/H \to 1$ and want to deduce the Brunn–Minkowski for G from the Brunn–Minkowski for H and G/H.

One tricky issue is that this inductive method only gives sharp results if the sum of the noncompact Lie dimensions and helix dimensions of H and G/H is equal to the noncompact Lie dimension of G. Unfortunately this is not always true (see examples in page 15). With this warning in mind, we must ensure the above property is always satisfied in the whole reduction. Our discussion in Section 7.2 guarantees this.

In the remaining part, we discuss some new challenges in the proof of Theorem 7.2 for a helix-free locally compact group G. The Gleason–Yamabe Theorem tells us that G contains an open subgroup G' that has a Lie quotient G'/H with H compact. For the start, we need to handle the nonuniqueness in the choice of G' and H and make sure that every choice gives the same desired result. This requires some nontrivial effort and makes heavy use of the Gleason–Yamabe Theorem, and we prove it in Section 7.2.

The rest of the proof of Theorem 7.2 has two steps. In the first step, we reduce the problem to unimodular groups. This is done with a similar strategy as used in the proof of the Lie group case with the additional help of a dichotomy result proved in Section 7.7. To motivate the second step, recall that in the Lie group case we first reduce the problem to connected groups. In our second step, unlike in the Lie group case, the identity component of our group here may not be open. Hence the correct analogue is to reduce the situation to open subgroups with a Lie quotient, which requires some additional results in Section 7.4. The desired result then follows from the Lie group case.

7.2 Noncompact Lie dimension and helix dimension

In this section, we show that noncompact Lie dimensions and helix dimensions are well defined in locally compact groups and that they behave well in many exact sequences. The latter is the nontrivial underlying reason that the lower bound in Theorem 7.1 and Theorem 7.2 matches the upper bound in Theorem 7.3 for helix-free locally compact groups.

Throughout the section, all groups are locally compact, we will use various definitions and facts from Sections 5.3, 5.4, and 5.5. The following lemma discusses the behavior of Iwasawa decomposition under taking quotient by a compact normal subgroup.

Lemma 7.6. *Suppose G is a connected semisimple Lie group, H is a (not necessarily connected) compact subgroup of G. Then we have the following.*

1. *There is an Iwasawa decomposition $G = KAN$ such that $H \leq K$.*

2. *Assume further that H is a normal subgroup of G, $G = KAN$ is an Iwasawa decomposition such that $H \leq K$, $G' = G/H$, and $\pi : G \to G'$ is the quotient map. Then there is an Iwasawa decomposition $G' = K'A'N'$ such that $\pi(K) = K'$.*

Proof. We first prove (1). Let $Z(G)$ be the center of G, $G' = G/Z(G)$ and $\rho : G \to G'$ be the quotient map, and $H' = \rho(H)$. By Facts 5.26 and 5.27, ρ is a covering map and G' is centerless. Let $\mathfrak{g}$ be the common Lie algebra of G and G', and $\exp : \mathfrak{g} \to G$ and $\exp' : \mathfrak{g} \to G'$ be the exponential maps. Using Fact 5.31.2 about Iwasawa decomposition, it suffices to construct a Cartan involution τ of $\mathfrak{g}$ such that if $\mathfrak{k}$ is the subalgebra of $\mathfrak{g}$ fixed by τ and $\exp(\mathfrak{k}) = K$, then $H \leq K$. Take a maximal compact subgroup K' of G' that contains H'. Let τ_0 be an arbitrary Cartan involution of G (this exists because of Fact 5.30). Let $\mathfrak{k}_0$ be the the subalgebra of $\mathfrak{g}$ fixed by τ_0, and $K'_0 = \exp(\mathfrak{k}_0)$ in G'. Then by Fact 5.31.2 about Iwasawa decomposition and the earlier observation that G' is centerless, K'_0 is a maximal compact subgroup of G'. By Fact 5.14.1 and the assumption that G is connected, there is an automorphism σ' of G' such that $\sigma'(K'_0) = K'$. Let α be the automorphism of $\mathfrak{g}$ obtain by taking the tangent map of σ', and let

$$\tau = \alpha\tau_0\alpha^{-1} \text{ and } \mathfrak{k} = \alpha(\mathfrak{k}_0)$$

As every Cartan–Killing form is invariant under automorphisms of $\mathfrak{g}$, we get that τ is a Cartan involution. It is also easy to check that $\mathfrak{k}$ is the subalgebra of $\mathfrak{g}$ fixed by τ. Using the functoriality of the exponential function (Fact 5.19), we get $K' = \exp(\mathfrak{k})$. Now set

$K = \exp(\mathfrak{k})$. By Fact 5.31.2, we get an Iwasawa decomposition $G = KAN$. Therefore, by the functoriality of the exponential function (Fact 5.19), $K' = \rho(K)$. Now as $H' \leq K'$, every element of H is in $Z(G)K$. By Fact 5.31.2 about Iwasawa decomposition, we have $Z(G) \subseteq K$, so $H \leq K$ as desired.

We now prove (2). Set $K' = \pi(K)$. Let $\mathfrak{g}$, $\mathfrak{h}$, and $\mathfrak{k}$ be the Lie algebras of G, H, and K, and let $\kappa_{\mathfrak{g}}$, $\kappa_{\mathfrak{h}}$, $\kappa_{\mathfrak{k}}$ be the Cartan–Killing form of $\mathfrak{g}$, $\mathfrak{h}$, and $\mathfrak{k}$. Then, $\mathfrak{g}' = \mathfrak{g}/\mathfrak{h}$ is the Lie algebra of G', and $\mathfrak{k}' = \mathfrak{k}/\mathfrak{h}$ is the Lie algebra of K' by Fact 5.18. Let τ be a Cartan involution of $\mathfrak{g}$ that fixes $\mathfrak{k}$. We will construct from this a Cartan involution τ' of $\mathfrak{g}'$ which fixes $\mathfrak{k}'$. If we have done so, then using Fact 5.31.2, we obtain A' and N' such that $G' = K'A'N'$ is an Iwasawa decomposition, which completes the proof.

Now we construct τ' as described earlier. As $\mathfrak{g}$ is semisimple, the Lie algebras $\mathfrak{h}$ and $\mathfrak{k}$ are also semisimple. With $\mathfrak{q}$ the orthogonal complement of $\mathfrak{k}$ in $\mathfrak{g}$ with respect to $\kappa_{\mathfrak{g}}$ and $\mathfrak{c}$ the orthogonal complement of $\mathfrak{h}$ in $\mathfrak{k}$ with respect to $\kappa_{\mathfrak{k}}$, we have $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ and $\mathfrak{k} = \mathfrak{h} \oplus \mathfrak{c}$ by Fact 5.24. By the same fact, with $\kappa_{\mathfrak{p}}$ and $\kappa_{\mathfrak{c}}$ the Cartan–Killing forms of $\mathfrak{p}$ and $\mathfrak{c}$, we have $\kappa_{\mathfrak{g}} = \kappa_{\mathfrak{k}} \oplus \kappa_{\mathfrak{p}}$ and $\kappa_{\mathfrak{k}} = \kappa_{\mathfrak{h}} \oplus \kappa_{\mathfrak{c}}$. It is then easy to see that every elements of $\mathfrak{c} \oplus \mathfrak{p}$ is orthogonal to $\mathfrak{h}$ with respect to $\kappa_{\mathfrak{g}}$. A dimension comparison gives us $\mathfrak{c} \oplus \mathfrak{p} = \mathfrak{d}$ with $\mathfrak{d}$ the orthogonal complement of $\mathfrak{h}$ in $\mathfrak{g}$. In summary, we have

$$\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p} = \mathfrak{h} \oplus \mathfrak{c} \oplus \mathfrak{p} = \mathfrak{h} \oplus \mathfrak{d} \quad \text{and} \quad \kappa_{\mathfrak{g}} = \kappa_{\mathfrak{k}} \oplus \kappa_{\mathfrak{p}} = \kappa_{\mathfrak{h}} \oplus \kappa_{\mathfrak{c}} \oplus \kappa_{\mathfrak{p}} = \kappa_{\mathfrak{h}} \oplus \kappa_{\mathfrak{d}}.$$

As a particular consequence, the quotient map from $\mathfrak{g}$ to $\mathfrak{g}'$ restricts to isomorphisms of Lie algebras from $\mathfrak{d}$ to $\mathfrak{g}' = \mathfrak{g}/\mathfrak{h}$ and from $\mathfrak{c}$ to $\mathfrak{k}' = \mathfrak{k}/\mathfrak{h}$. Since $\mathfrak{h}$ is a subalgebra of $\mathfrak{k}$, τ fixes $\mathfrak{h}$. As Cartan–Killing forms are invariant under automorphisms, τ restricts to an endomorphism of $\mathfrak{d}$, which the the orthogonal complement of $\mathfrak{h}$ in $\mathfrak{g}$ under $\kappa_{\mathfrak{g}}$. Therefore, $\tau|_{\mathfrak{d}}$ is an involution of $\mathfrak{d}$. The bilinear form

$$\mathfrak{d} \times \mathfrak{d} : (x, y) \mapsto -\kappa_{\mathfrak{d}}(x, \tau|_{\mathfrak{d}}(y))$$

is positive definite as it is simply the restriction to $\mathfrak{d}$ of the positive definite bilinear form

$\mathfrak{g} \times \mathfrak{g} : (x, y) \mapsto -\kappa_{\mathfrak{d}}(x, \tau(y))$. Hence, $\tau|_{\mathfrak{d}}$ is a Cartan involution of $\mathfrak{d}$. It is clear that the subalgebra of $\mathfrak{d}$ fixed by $\tau|_{\mathfrak{d}}$ is $\mathfrak{c}$. Finally, let τ' be the pushforward of $\tau|_{\mathfrak{d}}$ under the quotient map from $\mathfrak{g}$ to $\mathfrak{g}'$. It is easy to see that τ' satisfies the desired requirement. $\square$

The following lemma allows us to compute noncompact Lie dimensions for the universal cover of a compact Lie group.

Lemma 7.7. *Suppose that K is a covering group of a compact Lie group K' with the covering map $\rho : K \to K'$, and that K and K' are connected. If $\ker(\rho)$ is a discrete group of rank h, and m is the maximum dimension of a compact subgroup of K. Then $h = \dim(K) - m$.*

Proof. We first consider the case when K is a solvable group. Then $K' \cong \mathbb{T}^k$ where k is the dimension of K by Fact 5.15.2. Recall that K is a quotient of the universal cover of K', which is $\mathbb{R}^k$. Hence, $K \cong \mathbb{R}^h \times \mathbb{T}^{k-h}$. It is easy to see that the maximum dimension of a compact subgroup of K is $k - h$, which gives us the desired conclusion in this case.

We now prove the statement of the Lemma. Let Q_K be the radical of K, $Q_{K'}$ the radical of K', $S_K = K/Q_K$, and $S_{K'} = K'/Q_{K'}$. Note that K and K' have the same Lie algebra $\mathfrak{k}$. By Fact 5.21, Q_K and $Q_{K'}$ have the same Lie algebra $\mathfrak{q}$, which is the radical of $\mathfrak{k}$. Moreover, by the functoriality of the exponential function (Fact 5.19), ρ restrict to a covering map from Q_K to $Q_{K'}$ with kernel $\ker \rho \cap Q_K$. By Fact 5.18, the Lie algebras of S_K and $S_{K'}$ are both isomorphic to $\mathfrak{k}/\mathfrak{q}$. Hence, S_K is a connected semisimple Lie group with compact Lie algebra. Using Fact 5.28, we get S_K is compact with finite center $Z(S_K)$. Let $\pi : K \to S_K$ be the quotient map. Note that $\ker \rho$ is a subgroup of the center of K by Fact 5.16. Hence, the image of $\pi|_{\ker \rho}$ is a subset of $Z(S_K)$, which is finite. As a consequence, $\ker \rho \cap Q_K$, which is the kernel of $\pi|_{\ker \rho}$, has the same rank h as $\ker \rho$. Let m_1 and m_2 be the maximum dimensions of a compact subgroup of Q_K and of S_K respectively. Then $m = m_1 + m_2$ by Fact 5.14.2. By the special case for the solvable group K proven earlier, $h + m_1 = \dim Q_K$. As S_K is compact, $m_2 = \dim S_K$. Thus, $h + m = h + m_1 + m_2 = \dim(Q_K) + \dim(S_K)$ as desired. $\square$

The following proposition links the noncompact Lie dimension and the helix dimension.

Proposition 7.8. *Suppose G is a connected semisimple Lie group of dimension d, m is the maximal dimension of a compact subgroup of G, h is the helix dimension of G, and $G = KAN$ is an Iwasawa decomposition of G. Then $h = \dim K - m$, or equivalently, $d - m - h = \dim(AN)$.*

Proof. Let $Z(G)$ be the center of G. Then $Z(G)$ has rank h by the definition. By Fact 5.31.2, we have $Z(G)$ is a subset of K. Let $G' = G/Z(G)$, and $K' = K/Z(G)$. Using Lemma 7.6.2, we obtain A' and N' such that $G' = K'A'N'$ is an Iwasawa decomposition. Let $\rho : G \to G'$ be the quotient map. The group $Z(G)$ is discrete by Fact 5.26, so ρ and $\rho|_K$ are covering maps.

Now, the maximum dimension of a compact subgroup of G is the same as that of K by Lemma 7.6.1. Applying Lemma 7.7 to K, we have that $h = \dim K - m$. Note that $d = \dim(K) + \dim(AN)$ by Fact 5.31, so we also get $d - m - h = \dim(AN)$. □

The next lemma discusses the noncompact Lie dimensions and the helix dimensions of a Lie group and its open subgroups.

Lemma 7.9. *Suppose G is a Lie group, and G' is an open subgroup of G. Then G and G' have the same dimension, the same maximum dimension of a compact subgroup, and the same helix dimension.*

Proof. It is clear that G and G' have the same dimension. Any compact subgroup of G' is a compact subgroup of G. If K is a compact subgroup of G, then $K \cap G'$ is an open subgroup of K, hence $K \cap G'$ has the same dimension as K. Therefore the maximum dimension of a compact subgroup of G is the same as that of G'. Finally, note that G and G' have the same identity component G_0, and the helix dimension is defined using G_0. Thus, G and G' have the same helix dimension. □

The following Lemma tells us the behavior of radical under quotient by a compact normal subgroup.

Lemma 7.10. *Suppose G is a Lie group, H is a compact normal subgroup of G, $G' = G/H$, $\pi : G \to G'$ is the quotient map. Let Q be the radical of G, $S = G/Q$. Then we have the following:*

1. *with $Q' = \pi(Q)$, and $S' = G'/Q'$, we have HQ is closed in G, $Q' = HQ/H$, and $S' = G'/(HQ/H) = (G/H)/(HQ/H))$ is canonically isomorphic as a topological group to both G/HQ and $(G/Q)/(HQ/Q) = S/(HQ/Q)$;*

2. *Q' is the radical of G';*

Proof. We prove (1). As H is compact, we get HQ is closed in G by Lemma 5.3. Then $Q' = HQ/H$, and $S' = G'/(HQ/H) = (G/H)/(HQ/H))$. The remaining part of (1) is a consequence of the third isomorphism theorem (Fact 5.1.3).

We next prove (2). As Q' is a quotient of the solvable group Q, it is solvable. Moreover, Q' is a connected closed normal subsgroup of G' as Q is a connected closed normal subgroup of G. By (1), G'/Q' is a quotient of the semisimple group S. Hence, G'/Q' is semisimple. Therefore, Q' is the maximal connected solvable closed normal subgroup of G'. In other words, Q' is the radical of G'. $\square$

The next lemma says in a Lie group, taking quotient by a normal compact group does not change the helix dimension. Doing so also does not change the difference between the dimension and the dimension of a maximum compact subgroup.

Lemma 7.11. *Suppose G is a Lie group, H is a compact normal subgroup of G, and $G' = G/H$. Let d, m, and h be the dimension, the maximal dimension of a compact subgroup, and the helix dimension of G, respectively. Define d', m', and h' likewise for G'. Then:*

1. *$d = d' + \dim(H)$ and $m = m' + \dim(H)$;*

2. *$h = h'$.*

Proof. We prove (1). Clearly, $d = d' + \dim(H)$. If K is a compact subgroup of G and $K' = \pi(K)$, then K' is a compact subgroup of G', then $\dim(K') + \dim(H) = \dim(K)$. Conversely, if K' is a compact subgroup of G', then $K = \pi^{-1}(K')$ is a compact subgroup of G by Lemma 5.4, and Lemma 7.7 that $\dim(K) = \dim(K') + \dim(H)$. Therefore, $m = m' + \dim(H)$.

We now prove (2). First further assume that both G and G' are semisimple. Let $\pi : G \to G'$ be the quotient map. Using Lemma 7.6.1, we obtain an Iwasawa decomposition $G = KAN$ of G such that $H \subseteq K$. By Lemma 7.6.2, we obtain an Iwasawa decomposition $G' = K'A'N'$ with $K' = \pi(K)$, $A' = \pi(A)$, and $N' = \pi_N$. Let m_K be the maximum dimension of a compact subgroup of K, and m'_K be the maximum dimension of a compact subgroup of K'. By Proposition 7.8, $m_K + h = \dim(K)$, and $m_{K'} + h' = \dim(K')$. Now, by (1) applied to K, we have $m_K = m'_K + \dim(H)$. Therefore, we get $h = h'$.

Next, consider the case where G is connected. Let Q be the radical of G, $S = G/Q$, $Q' = \pi(Q)$, and $S' = G'/Q'$. Then by Lemma 7.10.2, Q' is the radical of G'. Hence, it suffices to show that S and S' has the same helix dimension. By Lemma 7.10.1, S' is isomorphic as a topological group to $S/(HQ/Q)$. Note that HQ/Q is isomorphic as a topological group to $H/(H \cap Q)$ by the second isomorphism theorem for Lie groups (Fact 5.13.2). In particular, HQ/Q is compact, and S' is the quotient of S by a compact group. Applying the known case for semisimple and connected groups, we get the desired conclusion.

Finally, we address the general case. Let G_0 be the identity component of G. Then G_0 is open by Fact 5.12, and G_0H/H is an open subgroup of $G' = G/H$. Hence, by Lemma 7.9, G has the same helix dimension as G_0, and G' has the same helix dimension as G_0H/H. By the second isomorphism theorem (Fact 5.1.2), G_0H/H is isomorphic as a topological group to $G_0/(G_0 \cap H)$, which is a quotient of G_0 by a compact subgroup. Thus, we get the desired conclusion for the general case from the known case discussed above for connected groups. ☐

Lemma 7.12. *Suppose G is an almost Lie group, H_1 and H_2 are closed normal subgroup of*

G such that G/H_1 and G/H_2 are Lie groups, and $H = H_1 \cap H_2$. Then G/H is a Lie group.

Proof. By Fact 5.9, G/H is an almost-Lie group. In light of Fact 5.10.2, we want to construct an open neighborhood U of the identity in G/H that contains no nontrivial compact subgroup. Let $\pi : G \to G/H$, $\pi_1 : G \to G/H_1$, and $\pi_2 : G \to G/H_2$ be the quotient maps. Using Fact 5.1.3, we get continuous surjective group homomorphisms $p_1 : G/H \to G/H_1$ and $p_2 : G/H \to G/H_2$ such that

$$\pi_1 = p_1 \circ \pi \quad \text{and} \quad \pi_2 = p_2 \circ \pi.$$

As G/H_1 is a Lie group, we can use Fact 5.10.2 to choose an open neighborhood U_1 of the identity in G/H_1 such that U_1 contains no nontrival compact subgroup of G/H_1. Choose an open neighborhood U_2 of the identity in G/H_2 likewise, and set

$$U = p_1^{-1}(U_1) \cap p_2^{-1}(U_2).$$

If $K \subseteq U$ is a compact subgroup of G/H, then $p_1(K)$ is a compact subgroup of U_1. By our choice of U_1, $p_1(L) = \{\mathrm{id}_{G/H_1}\}$, which implies that $\pi_1^{-1}(p(K)) = \pi^{-1}(K)$ is a subgroup of H_1. A similar argument yields that $\pi_2^{-1}(p_2(K)) = \pi^{-1}(K)$ is a subgroup of H_2. Hence, $\pi^{-1}(K)$ must be a subgroup of $H = H_1 \cap H_2$. It follows that $K = \{\mathrm{id}_{G/H}\}$, which is the desired conclusion. □

Proposition 7.13 below ensures us the notion of noncompact Lie dimension and helix dimension of a locally compact group as described in the introduction are well defined.

Proposition 7.13. *Suppose G' is an open subgroup of G, and $H \triangleleft G'$ is compact such that G'/H is a Lie group with dimension d, with maximum dimension of a compact subgroup m, and helix dimension h. Then $d - m$ and h are independent of the choice of G' and H.*

Proof. We first prove a simpler statement: If G' is an almost Lie subgroup of G, H is a compact subgroup of G', and we define d, m, and h as in the statement of the Proposition, then $d-m$ and h are independent of the choice of H. Let H_1 and H_2 be compact and normal subgroups of G such that both G/H_1 and G/H_2 are Lie groups. Then by Lemma 7.12, $G/(H_1 \cap H_2)$ is also a Lie group. Note that G/H_1 and G/H_2 are quotients of $G/(H_1 \cap H_2)$ by compact subgroups by the third isomorphism theorem (Fact 5.1.3). Hence, it follows from Lemma 7.11 that G/H_1 and $G/(H_1 \cap H_2)$ have the same difference between the dimension and the maximum dimension of a compact subgroup, and the same helix dimension. A similar statement holds for G/H_2 and $G/(H_1 \cap H_2)$. This completes the proof of the simpler statement.

Now we show the statement of the proposition. Let G_1' and G_2' be open subgroups of G, H_1 and H_2 are compact normal subgroup of G_1' and G_2' respectively such that G_1'/H_1 and G_2'/H_2 are Lie groups. Using the Gleason–Yamabe Theorem (Fact 5.10), we get an open subgroup G' of $G_1 \cap G_2$ which is an almost Lie group. Then G' is an open subgroup of G. Note that $G' \cap H_1$ and $G' \cap H_2$ are compact subgroups of G'. Then $G'/(G' \cap H_1)$ is an open subgroup of G_1'/H_1. It follows from Lemma 7.11 that G'/H_1 and $G'/(G' \cap H_1)$ have the same difference between the dimension and the maximum dimension and the same helix dimension. A similar statement hold for G'/H_2 and $G'/(G' \cap H_2)$. Thus, from the simpler statement we proved in the preceding paragraph, G_1'/H_1 and G_2'/H_2 have the same noncompact dimension and and the same helix dimension. □

We have the following two corollaries.

Corollary 7.14. *If H is an open subgroup of G, then H has the same noncompact Lie dimension and helix dimension as G.*

Proof. Proposition 7.13 implies that the noncompact Lie dimension and helix dimension of a locally compact group is the same as its open almost-Lie subgroups, if those exist. Hence, it suffices to show that there is a common almost-Lie open subgroup of G and H. This is an immediate consequence of the Gleason–Yamabe Theorem (Fact 5.10.1). □

Corollary 7.15. *If H is a compact normal subgroup of G, then G/H has the same non-compact Lie dimension and helix dimension as G.*

Proof. Let π be the projection from G to G/H. If G/H is a Lie group, then from the definitions, G has the same noncompact Lie dimension and helix dimension as G/H. Hence, the conclusion holds in this special case.

Suppose there is a compact $K \lhd G/H$ such that $(G/H)/K$ is a Lie group, then $(G/H)/K$ is isomorphic as topological group to $G/\pi^{-1}(K)$ by Fact 5.1.3. By Lemma 5.4, $\pi^{-1}(K)$ is compact. Hence $(G/H)/K$ is a quotient of G by a compact normal subgroup, and we can use the previous case to get the desired conclusion.

Now we treat the general situation. By the Gleason–Yamabe Theorem, we get an almost-Lie open subgroup G' of G. Then $G'H$ is an open subgroup of G and hence has the same noncompact Lie dimension and helix dimension as G by Corollary 7.14. By the second isomorphism theorem (Fact 5.1.2), we get that $G'/(G' \cap H)$ is isomorphic to $G'H/H$ which is an open subgroup of G/H. In particular, $G'/(G' \cap H)$ has the same noncompact Lie dimension and helix dimension as G/H by Corollary 7.14. Note that $G'/(G' \cap H)$ is an almost-Lie group by Fact 5.9. Hence, we can find K such that $(G'/(G' \cap H))/K$ is a Lie group. We are back to the earlier known situation in the second paragraph. $\square$

We have the following lemma about the Iwasawa decompositions.

Lemma 7.16. *Suppose $1 \to H \to G \xrightarrow{\pi} G/H \to 1$ is an exact sequence of connected semisimple Lie groups. Then there are Iwasawa decompositions $G = KAN$, $H = K_1A_1N_1$, and $G/H = K_2A_2N_2$ such that $K_1 = (K \cap H)_0$, and $K_2 = \pi(K)$.*

Proof. Let $\mathfrak{g}$ and $\mathfrak{h}$ be the Lie algebras of G and H, and let $\kappa_\mathfrak{g}$ and $\kappa_\mathfrak{h}$ be the Cartan–Killing form of $\mathfrak{g}$ and $\mathfrak{h}$. Then $\mathfrak{g}/\mathfrak{h}$ is the Lie algebra of G/H, and $\mathfrak{g}$,$\mathfrak{h}$, and $\mathfrak{g}/\mathfrak{h}$ are semisimple. By Fact 5.24,

$$\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{c} \text{ and } \kappa_\mathfrak{g} = \kappa_\mathfrak{h} \oplus \kappa_\mathfrak{c}$$

where $\kappa_{\mathfrak{c}}$ is the orthogonal complement of $\kappa_{\mathfrak{h}}$ with respect to $\kappa_{\mathfrak{g}}$, and $\kappa_{\mathfrak{c}}$ is the Cartan–Killing form of $\kappa_{\mathfrak{c}}$. Therefore, the quotient map from $\mathfrak{g}$ to $\mathfrak{g}/\mathfrak{h}$ induces an isomorphism from $\mathfrak{c}$ to $\mathfrak{g}/\mathfrak{h}$, so we can identify $\mathfrak{g}/\mathfrak{h}$ with $\mathfrak{c}$. Let τ_1 and τ_2 be a Cartan involutions of $\mathfrak{h}$ and $\mathfrak{c}$. Then $\tau = \tau_1 \oplus \tau_2$ is an involution of $\mathfrak{g}$. As τ_1 and τ_2 are Cartan involutions, the bilinear forms $\mathfrak{h} \times \mathfrak{h} : (x_1, y_1) \mapsto -\kappa_{\mathfrak{h}}(x_1, \tau_1(y_1))$ and $\mathfrak{c} \times \mathfrak{c} : (x_2, y_2) \mapsto -\kappa_{\mathfrak{c}}(x_2, \tau_2(y_2))$ are positive definite. Hence, the bilinear from $\mathfrak{g} \times \mathfrak{g} : (x, y) \mapsto -\kappa_{\mathfrak{g}}(x, \tau(y))$ is also positive definite. Therefore, τ is a Cartan involution of $\mathfrak{g}$. Let $\mathfrak{k}$, $\mathfrak{k}_1$, and $\mathfrak{k}_2$ be the Lie subalgebras of $\mathfrak{g}$, $\mathfrak{h}$, and $\mathfrak{c}$ fixed by τ, τ_1, and τ_2 respectively. It is easy to see that $\mathfrak{k} = \mathfrak{k}_1 \oplus \mathfrak{k}_2$. Let $\exp : \mathfrak{g} \to G$, $\exp_1 : \mathfrak{h} \to H$, and $\exp_2 : \mathfrak{c} \to G/H$ be the exponential maps, and set

$$K = \exp(\mathfrak{k}), K_1 = \exp_1(\mathfrak{k}_1) \text{ and } K_2 = \exp(\mathfrak{k}_2).$$

From Fact 5.31, we obtain Iwasawa decompositions $G = KAN$, $H = K_1 A_1 N_1$, and $G/H = K_2 A_2 N_2$. By the functoriality of the exponential function (Fact 5.19), we get $K_1 \leq K \cap H$, and $K_2 = \pi(K)$. Since K_1 is connected, by a dimension calculation we have $K_1 = (K \cap H)_0$. $\square$

In a short exact sequence of locally compact groups, one may hope that the noncompact Lie dimension and the helix dimension of the middle term is the sum of those of the outer terms. This is not true in general. For instance, in the exact sequence

$$1 \to \mathbb{Z} \to \mathbb{R} \to \mathbb{R}/\mathbb{Z} \to 1,$$

the noncompact Lie dimension of $\mathbb{R}$ is 1, while both $\mathbb{Z}$ and $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ has noncompact Lie dimension 0. Another exmaple is the following. Let H be the universal cover of $\mathrm{SL}(2, \mathbb{R})$, and let $G = (H \times \mathbb{R})/\{(n, n) : n \in \mathbb{Z}\}$. Then we have the exact sequence

$$1 \to H \to G \to \mathbb{T} \to 1,$$

the helix dimension of H is 1, but the helix dimensions of G and $\mathbb{T}$ are 0.

Nevertheless, we have the summability of noncompact Lie dimensions and helix dimensions in many short exact sequences of interest:

Proposition 7.17. *Suppose* $1 \to H \to G \xrightarrow{\pi} G/H \to 1$ *is an exact sequence of connected Lie groups. Then we have the following:*

1. *If* n, n_1, *and* n_2 *are the noncompact Lie dimensions of* G, H, *and* G/H *respectively, then* $n = n_1 + n_2$;

2. *If* G *is moreover semisimple, and* h, h_1, *and* h_2 *are the helix dimensions of* G, H, *and* G/H *respectively, then* $h = h_1 + h_2$.

Proof. We first prove (1). Let m be the maximum dimension of a compact subgroup in G. As G is connected, m is also the dimension of an arbitrary maximal compact subgroup of G by Fact 5.14.1. Defining m_1 and m_2 likewise for H and G/H, we get similar conclusions for them from the connectedness of H and G/H. Let K be a maximal compact subgroup of G. By Fact 5.14.2, $K \cap H$ is a maximal compact subgroup in H, and $\pi(K)$ is a maximal compact subgroup in G/H. The kernel of $\pi|_K$ is isomorphic to $K \cap H$, and the image is $\pi(K)$. Hence, $m = m_1 + m_2$. This gives us (1) recalling that $m + n = \dim(G)$, $m_1 + n_1 = \dim(H)$, $m_2 + n_2 = \dim(G/H)$, and $\dim(G) = \dim(H) + \dim(G/H)$.

We now prove (2). Since $Z(G) \cap H \leq Z(H)$, and $\pi(Z(G)) \leq Z(G/H)$, we have $h \leq h_1 + h_2$. It remains to show $h \geq h_1 + h_2$. As G is semisimple, H and G/H are semisimple by Fact 5.25. Take Iwasawa decompositions $G = KAN$, $H = K_1A_1N_1$, and $G/H = K_2A_2N_2$ as in Lemma 7.16. By the first isomorphism theorem for Lie groups (Fact 5.13.1), $1 \to K \cap H \to K \to K_2 \to 1$ is an exact sequence of Lie groups. We also have an exact sequence

$$1 \to K_1 \to K \to K_2' \to 1. \tag{7.3}$$

As $K_1 = (K \cap H)_0$, by the third isomorphism theorem, we have $K_2 = K/(K \cap H) =$

$(K/K_1)/((K\cap H)/K_1) = K_2'/((K\cap H)/K_1)$. Since $(K\cap H)/K_1$ is discrete, K_2' is a covering group of K_2. Let $\phi : K_2' \to K_2$ be the covering map. Note that ϕ has discrete kernel, and K_2, K_2' have the same dimension. Suppose S is a compact subgroup of K_2' with the maximum dimension. Then $\phi(S)$ is a compact subgroup of K_2, and S and $\phi(S)$ have the same dimension. This shows that the noncompact Lie dimension of K_2' is at least the noncompact Lie dimension of K_2. By (7.3) and Statement (1), the noncompact Lie dimension of K is the sum of noncompact Lie dimensions of K_1 and K_2', hence it is at least the sum of noncompact Lie dimensions of K_1 and K_2. It then follows from Proposition 7.8 that $h \geq h_1 + h_2$. □

Lemma 7.18. *Suppose* $1 \to H \to G \xrightarrow{\pi} (\mathbb{R}^{>0}, \times) \to 1$ *is an exact sequence of Lie groups, and* G *is connected. Then* H *is connected.*

Proof. Consider first the case when when G and H are Lie groups but are not necessarily connected. Let G_0 and H_0 be the identity components of G and H respectively. As Lie groups are locally path connected, G_0 is open in G. Hence, G_0 and G have the same noncompact Lie dimension by Corollary 7.14. Likewise, H_0 has the same noncompact Lie dimension as H. As G_0 is an open connected subgroup of G, the map $\pi|_{G_0}$ is continuous and open. Hence, its image $\pi(G_0)$ is an open connected subgroup of $(\mathbb{R}^{>0}, \times)$. Therefore, $\pi(G_0) = (\mathbb{R}^{>0}, \times)$, and $\pi|_{G_0}$ is a quotient map by the first isomorphism theorem (Fact 5.1.1). The kernel of $\pi|_{G_0}$ is $H \cap G_0$, so we get the exact sequence of Lie groups

$$1 \to H \cap G_0 \to G_0 \xrightarrow{\pi|_{G_0}} (\mathbb{R}^{>0}, \times) \to 1.$$

We claim that $H_0 = H \cap G_0$, which will bring us back to the known case where both G and H are connected. The forward inclusion is immediate by definition. By the third isomorphism theorem (Fact 5.1.3), we get the exact sequence of Lie groups

$$1 \to (H \cap G_0)/H_0 \to G_0/H_0 \to (\mathbb{R}^{>0}, \times) \to 1.$$

The group $(H \cap G_0)/H_0$ is discrete. Hence, G_0/H_0 is a Lie group with dimension 1. As G_0 is connected, the Lie group G_0/H_0 is also connected. Hence, G_0/H_0 is either isomorphic to $\mathbb{R}$ or $\mathbb{T}$. But since G_0/H_0 has $(\mathbb{R}^{>0}, \times)$ as a quotient, it cannot be compact, and therefore must be isomorphic to $\mathbb{R}$. This implies that $(H \cap G_0)/H_0$ is trivial, and hence $H_0 = H \cap G_0$. $\square$

The next proposition gives us a summability result of noncompact Lie dimensions along a short exact sequence of locally compact groups when the quotient group is $(\mathbb{R}^{>0}, \times)$.

Proposition 7.19. *Suppose* $1 \to H \to G \xrightarrow{\pi} (\mathbb{R}^{>0}, \times) \to 1$ *is an exact sequence of locally compact groups. Then we have the following:*

1. *If* n, n_1, *and* n_2 *are the noncompact Lie dimensions of* G, H, *and* $(\mathbb{R}^{>0}, \times)$ *respectively, then* $n = n_1 + n_2 = n_1 + 1$.

2. G *and* H *have the same helix dimension.*

Proof. First, we consider the case when G is a connected Lie group. Then by Lemma 7.18, H is also connected. Hence, (1) for this case is a consequence of Proposition 7.17.1.

We prove (2) for this special case. Let Q be the radical of G. We claim that $QH = G$, or equivalently, that $\pi(Q) = (\mathbb{R}^{>0}, \times)$. Suppose this is not true. Then $\pi(Q)$ is a connected subgroup of $(\mathbb{R}^{>0}, \times)$, so it must be $\{1\}$. Hence, $Q \subseteq H$. Then $(\mathbb{R}^{>0}, \times) = G/H$ which is isomorphic as a topological group to $(G/Q)/(H/Q)$ by the third isomorphism theorem (Fact 5.1.3). This is a contradiction, because $(G/Q)/(H/Q)$ is semisimple as a quotient of the semisimple group G/Q, while $(\mathbb{R}^{>0}, \times)$ is solvable.

We next show that $Q \cap H$ is the radical of H. The radical of H is a characteristic closed subgroup of H (by Fact 5.21), hence a connected solvable closed normal subgroup of G. Thus, the radical of H is a subgroup of $Q \cap H$. It is straightforward that $Q \cap H$ is solvable. We also have that $Q \cap H$ is second countable as both Q and H are second countable. From the preceding paragraph, $\pi(Q) = (\mathbb{R}^{>0}, \times)$. Using the first isomorphism theorem for Lie groups (Fact 5.13.1), we have the exact sequence

$$1 \to Q \cap H \to Q \to (\mathbb{R}^{>0}, \times) \to 1.$$

Applying Lemma 7.18, we learn that $Q \cap H$ is connected. This completes the proof that $Q \cap H$ is the radical of H.

Note that $QH = G$ and Q is a closed subgroup of G. Hence, by the second isomorphism theorem for Lie groups (Fact 5.13.2), $H/(Q \cap H)$ is isomorphic as a topological group to $HQ/Q = G/Q$. Therefore G and H have the same helix dimension.

Next, we address the slightly more general case where G is a Lie group but not necessarily connected. Let G_0 be the connected component of G. Then $\pi(G_0)$ is an open subgroup of $(\mathbb{R}^{>0}, \times)$, so $\pi(G_0) = (\mathbb{R}^{>0}, \times)$. By the first isomorphism theorem for Lie groups (Fact 5.13.1), we have the exact sequence

$$1 \to G_0 \cap H \to G_0 \to (\mathbb{R}^{>0}, \times) \to 1.$$

Applying Lemma 7.29 and the known case of the current lemma where the middle term of the exact sequence is a connected Lie group, we obtain both (1) and (2) for this more general case.

Using the Gleason–Yamabe theorem and a similar argument as in the preceding paragraph, we can reduce (1) and (2) for general locally compact groups to the case where we assume that G is an almost-Lie group. Hence, there is a compact normal subgroup K of G such that G/K is a Lie group. As K is compact, $\pi(K)$ is a compact subgroup of $(\mathbb{R}^{>0}, \times)$, so $\pi(K) = \{1\}$. Hence $K \lhd H$. By the third isomorphism theorem (Fact 5.1.3), we have the exact sequence $1 \to H/K \to G/K \to (\mathbb{R}^{>0}, \times) \to 1$. Applying Lemma 7.11 and the known case of the current lemma where the middle term of the exact sequence is a Lie group, we obtain both (1) and (2) for this remaining case. $\square$

We discuss the relationship between the noncompact Lie dimension and helix dimension of a locally compact group G.

Corollary 7.20. *Suppose G has noncompact Lie dimension n and helix dimension h. Then we have $h \leq n/3$.*

Proof. We first check the result for simple Lie groups. If $h = 0$, then the statement holds vacuously. Hence, using Fact 5.29, it suffices to consider the case where $h = 1$. Let $G = KAN$ be an Iwasawa decomposition. Then by Proposition 7.7, we have $n - 1 = \dim(AN) \geq 0$. Hence, $n > 0$ and G is not compact. From Fact 5.32, we have $\dim(AN) \geq 2$. Therefore $n \geq 3$. Hence, we get the desired conclusion for simple Lie groups.

When G is a connected semisimple Lie group which is not simple. Using induction on dimension, we can assume we have proven the statement for all connected semisimple Lie groups of smaller dimensions. Using Fact 5.23, we get an exact sequence of semisimple Lie groups $1 \to H \to G \to G/H \to 1$ with $0 < \dim(H) < \dim(G)$. Replacing H with its connected component if necessary, we can arrange that H is connected. The desired conclusion then follows from Proposition 7.17.2.

For a general locally compact group G, from Proposition 7.13, we may assume G is a Lie group. Corollary 7.14 and Fact 5.12 allow us to reduce the problem to connected Lie groups. By Lemma 7.10 and Lemma 7.9, the radical of G only contributes the noncompact Lie dimension of G. Using Fact 5.22 and Proposition 7.17.1, we reduce the problem to connected semisimple Lie groups. $\square$

7.3 Proof of Theorem 7.3

The constructions given in this section are open sets (hence all have positive measure), and the exact statement given in Theorem 7.3 (i.e., in the compact sets case) follows by the inner regularity of Haar measure.

We first prove the theorem when G is a unimodular Lie group.

Proof of Theorem 7.3, unimodular Lie group case. Since G is unimodular, without loss of generality we assume that $\mu = \nu$. Let d be the dimension of G. Let K be the maximal compact subgroup of G and let $m = \dim K$. Hence $n = d - m$ is the noncompact Lie dimension of G.

If $n = 0$, then the identity component G_0 of G is compact. Taking $X = G_0$, we have

$$\mu(X) = \mu(X^2).$$

Hence Theorem 7.3 holds in this case. In the rest of the proof we assume $n > 0$.

Since K is closed, G/K is a homogeneous (and smooth) manifold. Fix an arbitrary G-invariant (smooth) Riemannian metric on G/K (such a metric exists by first finding a K-invariant Riemannian metric at [id] and then extend it onto the whole G/K by the action of G). This metric induces a volume measure Vol on G/K.

Let π be the projection from G to G/K. For any Borel subset U of G/K, $\pi^{-1}(U)$ is also Borel and hence μ-measurable. For any $r > 0$, we use B_r to denote the (open) r-ball around [id] on G/K under the chosen metric and use D_r to denote $\pi^{-1}(B_r)$. We claim that:

(i) There exists a constant $b > 0$ only depending on the metric on G/K such that as Borel measures $\pi_*(\mu) = b \cdot \mathrm{Vol}$, and

(ii) For any $r > 0$, $D_r \cdot D_r \subseteq D_{2r}$.

We postpone the proofs of claims (i) and (ii) to the end of this proof and first show how they lead to Theorem 7.3. We can take X to be D_δ for a sufficiently small $\delta > 0$ (depending on ε) to be determined. Then by (i),

$$\mu(X) = \pi_*(\mu(B_\delta)) = b \cdot \mathrm{Vol}(B_\delta).$$

And by (ii), $X^2 \subseteq D_{2\delta}$ and hence as before, we get $\mu(X^2) \leq \mu(D_{2\delta}) = b \cdot \mathrm{Vol}(B_{2\delta})$. Note that the invariant metric on G/K is smooth and thus

$$\lim_{\delta\to 0}\frac{\mathrm{Vol}(B_{2\delta})}{\mathrm{Vol}(B_\delta)}=2^n.$$

Hence a sufficiently small δ can guarantee $\frac{\mu(X)^{\frac{1}{n}-\varepsilon}}{\mu(X^2)^{\frac{1}{n}-\varepsilon}}>\frac{1}{2}$ and we have proved Theorem 7.3 in this special case.

It remains to prove claims (i) and (ii). To see claim (i), note that Vol is G-invariant. We also see that $\pi_*(\mu)$ is G-invariant because $\mu(\pi^{-1}(U))=\mu(g\pi^{-1}(U))=\mu(\pi^{-1}(gU))$ for any $g\in G$ and any Borel subset $U\subseteq G/K$. Since the G-invariant Borel measure on G/K is unique up to a scalar (see Theorem 8.36 in [114]), Vol has to be a scalar multiple of $\pi_*(\mu)$.

Finally we verify claim (ii). Taking arbitrary $g_1,g_2\in D_r$ and it suffices to show $g_1g_2\in D_{2r}$. By definition, there is a piecewise smooth curve γ_j connecting [id] and $[g_j]$ such that the length of γ_j is strictly smaller than r (for $j=1,2$). Note that by the invariance of the metric, $[g_1]\gamma_2$ must have the same length as γ_2. Let γ be the curve formed by $[g_1]\gamma_2$ after γ_1. It is a curve connecting [id] and $[g_1g_2]$ and by the reasoning above has two pieces and each of them has length strictly smaller than r. Hence γ has length shorter than $2r$ and thus by definition $g_1g_2\in D_{2r}$. We have successfully verified (ii). □

Running the above proof with a little bit of extra effort, we have the following slightly stronger "stability" result. We will use it in the generalization to the nonunimodular Lie group case.

Proposition 7.21. *Given any unimodular Lie group G, let n be its noncompact Lie dimension. Let $\tilde{\varepsilon}>0$ be fixed. Then there exists precompact open subsets X and X_1 with $\mu(X)>0$ such that the closure $\overline{X}\subseteq X_1$ and $\mu(X_1{\cdot}X)<(2+\tilde{\varepsilon})^n\mu(X)$.*

Proof. This proof is very similar to the proof of the unimodular Lie case of Theorem 7.3 we just did. We continue to use notations in that proof and take $X=D_\delta=\pi^{-1}(B_\delta)$ and $X_1=D_{\delta_1}=\pi^{-1}(B_{\delta_1})$ where $0<\delta<\delta_1$ and both δ and δ_1 to be determined.

We see that X and X_1 are open because $X = \pi^{-1}(B_\delta)$, etc. B_δ and B_{δ_1} are precompact, by Lemma 5.4, X and X_1 are also precompact. Moreover we have $\overline{X} \subseteq \pi^{-1}(\overline{B_\delta}) \subseteq \pi^{-1}(B_{\delta_1}) = X_1$.

Now by the same reasoning as in the previous proof of Theorem 7.3 (unimodular Lie case), we see that $X_1 \cdot X \subseteq D_{\delta_1+\delta}$. Now,

$$\lim_{\delta\to 0} \frac{\mathrm{Vol}(B_{2\delta})}{\mathrm{Vol}(B_\delta)} = 2^n, \quad \text{and} \quad \lim_{\delta_1\to\delta} \frac{\mathrm{Vol}(B_{2\delta})}{\mathrm{Vol}(B_{\delta_1+\delta})} = 1.$$

Hence we can take δ sufficiently small, and then δ_1 sufficiently close to δ, such that we have all good properties in the last paragraph and

$$\frac{\mu(X_1 \cdot X)}{\mu(X)} \le \frac{\mu(D_{\delta_1+\delta})}{\mu(D_\delta)} = \frac{\mathrm{Vol}(B_{\delta_1+\delta})}{\mathrm{Vol}(B_\delta)} < (2+\tilde{\varepsilon})^n,$$

which proves the proposition. $\square$

Next we use Proposition 7.21 to prove Theorem 7.3 for general Lie groups.

Proof of Theorem 7.3, Lie group case. We have already proved the theorem when G is unimodular. In the rest of this proof, we assume G is nonunimodular. Let G_0 be the connected component of G. Since $\mu_G|_{G_0}$ is a left Haar measure on G_0, and same holds $\nu_G|_{G_0}$, we may assume without loss of generality that $G = G_0$. As the only connected subgroups of $(\mathbb{R}^{>0}, \times)$ is itself and $\{1\}$, and G is not unimodular, the modular function Δ_G must be surjective. Hence, Δ_G is a quotient map by the first isomorphism for Lie groups Fact 5.13.1.

Let H be the kernel of the modular function on G. By Proposition 7.19.1, the noncompact Lie dimension of H is $n-1$ where n is the noncompact Lie dimension of G. By Fact 5.6.1, H is unimodular. To avoid confusion, we will always use μ_G and ν_G for μ and ν below and use $\mu_H = \nu_H$ to denote a fixed Haar measure on H.

In light of Fact 5.7, we can fix a Haar measure $\mathrm{d}r$ on the multiplicative group $(\mathbb{R}^{>0}, \times) =$

G/H such that for any Borel function f on g,

$$\int_G f(x)\,\mathrm{d}\mu_G(x) = \int_{G/H}\int_H f(rh)\,\mathrm{d}\mu_H(h)\,\mathrm{d}r. \tag{7.4}$$

Let $\mathfrak{g}$ and $\mathfrak{h}$ be the Lie algebras of G and H, respectively. We fix an element $Z \in \mathfrak{g}$ such that $Z \notin \mathfrak{h}$. Note that $t \mapsto \Delta(\exp(tZ))$ is a nontrivial continuous group homomorphism from $(\mathbb{R}, +)$ to $(\mathbb{R}^{>0}, \times)$. As the only connected subset of $(\mathbb{R}^{>0}, \times)$ are points and intervals, this map must be surjective, and hence an isomorphism by the first isomorphism for Lie groups (Fact 5.13). In light of the quotient integral formula (7.4), we can choose an appropriate Haar measure $\mathrm{d}t$ on $\mathbb{R}$ such that for any Borel subset A of G, we have the Fubini-type measure formula

$$\mu_G(A) = \int_{\mathbb{R}} \mu_H((\exp(-tZ)A) \cap H)\,\mathrm{d}t. \tag{7.5}$$

Without loss of generality we assume $\mathrm{d}t$ is the standard Lebesgue measure (otherwise we multiply μ_G by a constant).

With the preliminary discussions above, we now construct X satisfying the inequality in Theorem 7.3.

Before going to details of the construction, we first describe the intuition behind it. We arrange our X to live very close to H so that μ and ν are almost proportional on X and X^2. We then realize that it suffices to choose our X to be like a thickened copy of the almost sharp example of Theorem 7.3 for (the unimodular group) H.

More precisely, let $\tilde{\varepsilon} > 0$ be a small number (depending on ε) to be determined. let $\tilde{X}$ and $\tilde{X}_1$ be the "X" and "X_1", respectively, in Proposition 7.21 where we replace "G" by "H". We now take $X = \{\exp(tZ)h : t \in [0, \tilde{\varepsilon}], h \in \tilde{X}\}$ and will show that X^2 is reasonably small when $\tilde{\varepsilon}$ is small enough.

By (7.5), we have

$$\mu_G(X^2) = \int_{\mathbb{R}} \mu_H((\exp(-tZ)X^2) \cap H)\,\mathrm{d}t. \tag{7.6}$$

Note that an arbitrary element in X^2 can be written as

$$\begin{aligned}&\exp(t_1Z)h_1\exp(t_2Z)h_2\\=\ &\exp((t_1+t_2)Z)(\exp(-t_2Z)h_1\exp(t_2Z)\cdot h_2)\in\exp((t_1+t_2)Z)H,\end{aligned}$$

where $t_1,t_2\in[0,\tilde{\varepsilon}]$ and $h_1,h_2\in H$. Hence (7.6) is reduced to

$$\mu_G(X^2)=\int_0^{2\tilde{\varepsilon}}\mu_H((\exp(-tZ)X^2)\cap H)\,\mathrm{d}t \tag{7.7}$$

and moreover for any $0\le t_0\le 2\tilde{\varepsilon}$, we see from the above discussion that

$$(\exp(-t_0Z)X^2)\cap H=\bigcup_{0\le t_1,t_2\le\tilde{\varepsilon},t_1+t_2=t_0}(\exp(-t_1Z)\tilde{X}\exp(t_1Z))\cdot\tilde{X}.$$

By Lemma 5.17 and Proposition 7.21, when $\tilde{\varepsilon}$ is sufficiently small, which we will always assume, we have the above union contained in $\tilde{X}_1\cdot\tilde{X}$. Now by (7.7),

$$\mu_G(X^2)\le\int_0^{2\tilde{\varepsilon}}\mu_H(\tilde{X}_1\cdot\tilde{X})\,\mathrm{d}t=2\tilde{\varepsilon}\mu_H(\tilde{X}_1\cdot\tilde{X}). \tag{7.8}$$

On the other hand, by (7.5) we have

$$\mu_G(X)=\tilde{\varepsilon}\mu_H(\tilde{X}). \tag{7.9}$$

Combining (7.8) and (7.9) and use the measure properties of $\tilde{X}$ and $\tilde{X}_1$ guaranteed by Proposition 7.21, we have

$$\frac{\mu_G(X)}{\mu_G(X^2)}\ge\frac{\mu_H(\tilde{X})}{2\mu_H(\tilde{X}_1\cdot\tilde{X})}>\frac{1}{2(2+\tilde{\varepsilon})^{n-1}}. \tag{7.10}$$

Recall that $\Delta(\exp(\cdot Z))$ is an isomorphism from $(\mathbb{R},+)$ to $(\mathbb{R}^{>0},\times)$. Hence there exists a constant $C>0$ only depending on Z such that on the support of X we have $e^{-C\tilde{\varepsilon}}<\Delta<e^{C\tilde{\varepsilon}}$

and on the support of X^2 we have $e^{-2C\tilde{\varepsilon}} < \Delta < e^{2C\tilde{\varepsilon}}$. Thus by Fact 5.6.4, we have

$$\frac{\nu_G(X)}{\mu_G(X)} > e^{-C\tilde{\varepsilon}}$$

and

$$\frac{\mu_G(X^2)}{\nu_G(X^2)} > e^{-2C\tilde{\varepsilon}}.$$

Combining the above inequalities with (7.10), we have

$$\frac{\nu_G(X)}{\nu_G(X^2)} > \frac{e^{-3C\tilde{\varepsilon}}}{2(2+\tilde{\varepsilon})^{n-1}}. \tag{7.11}$$

Hence for the X we constructed,

$$\frac{\nu_G(X)^{\frac{1}{n}-\varepsilon}}{\nu_G(X^2)^{\frac{1}{n}-\varepsilon}} + \frac{\mu_G(X)^{\frac{1}{n}-\varepsilon}}{\mu_G(X^2)^{\frac{1}{n}-\varepsilon}} > (1+e^{-3C\tilde{\varepsilon}(\frac{1}{n}-\varepsilon)})(2(2+\tilde{\varepsilon})^{n-1})^{-\frac{1}{n}+\varepsilon}. \tag{7.12}$$

It suffices to take $\tilde{\varepsilon}$ small enough such that the right hand side of (7.12) is > 1. $\square$

With the Gleason–Yamabe Theorem and the results developed in Section 7.2, we are able to pass our Lie group constructions to general locally compact groups.

Proof of Theorem 7.3. By Fact 5.10, there is open subgroup G' of G which is almost-Lie. Since $\mu_G|_{G'}$ is a left Haar measure on G', and same holds $\nu_G|_{G'}$, we may assume without loss of generality that G is almost-Lie.

With this assumption, there is a a short exact sequence $0 \to H \to G \xrightarrow{\pi} G/H \to 0$ where H is a compact subgroup, and G/H is a Lie group. Let X be a subset of G/H such that

$$\frac{\nu_{G/H}(X)^{\frac{1}{n}-\varepsilon}}{\nu_{G/H}(X^2)^{\frac{1}{n}-\varepsilon}} + \frac{\mu_{G/H}(X)^{\frac{1}{n}-\varepsilon}}{\mu_{G/H}(X^2)^{\frac{1}{n}-\varepsilon}} > 1, \tag{7.13}$$

where n is the noncompact Lie dimension of G/H. Thus by the quotient integral formula, we have

$$\begin{aligned}\mu_G(\pi^{-1}(X)) &= \int_{G/H} \mu_H(g^{-1}(\pi^{-1}(X)) \cap H)\, \mathrm{d}\mu_{G/H}(g) \\ &= \int_{G/H} \mathbb{1}_X(g)\, \mathrm{d}\mu_{G/H}(g) = \mu_{G/H}(X),\end{aligned}$$

and similarly $\nu_G(\pi^{-1}(X)) = \nu_{G/H}(X)$. Observe that $\pi^{-1}(X^2) = \pi^{-1}(X) \cdot \pi^{-1}(X)$. Thus the desired conclusion follows from (7.13) and Propositions 7.13. $\square$

7.4 Reduction to outer terms of certain short exact sequences

For $n \in \mathbb{Z}^{\geq 0}$ and $(x, y) \in \mathbb{R}^2$, we set

$$\big\|(x,y)\big\|_{1/n} = \begin{cases} (|x|^{1/n} + |y|^{1/n})^n & \text{if } n \neq 0, \\ \max\{|x|, |y|\} & \text{if } n = 0. \end{cases}$$

We say that the group G satisfies the **Brunn–Minkowski inequality with exponent** n, abbreviated as BM(n), if for all compact $X, Y \subseteq G$,

$$\left\|\Big(\frac{\nu(X)}{\nu(XY)}, \frac{\mu(Y)}{\mu(XY)}\Big)\right\|_{1/n} \leq 1.$$

When G is unimodular and $n \geq 1$, the above is equivalent to having the inequality

$$\mu(XY)^{1/n} \geq \mu(X)^{1/n} + \mu(Y)^{1/n}.$$

Note that $\frac{\nu(X)}{\nu(XY)} \leq 1$ and $\frac{\mu(Y)}{\mu(XY)} \leq 1$. Hence, every locally compact group G satisfies the Brunn–Minkowski inequality with exponent $n = 0$. Moreover, if $n < n'$ and G satisfies the Brunn–Minkowski inequality with exponent n', then it satisfies the Brunn–Minkowski inequality with exponent n.

Given a function $f : X \to \mathbb{R}$, for every $t \in \mathbb{R}$, define the *superlevel set* of f

$$L_f^+(t) := \{x \in X : f(x) \geq t\}.$$

We will use this notation at various points in the later proofs. We use the following simple consequence of Fubini concerning the superlevel sets:

Fact 7.22. *Let $f : G \to \mathbb{R}$ be a function. For every $r > 0$,*

$$\int_G f^r(x)\,\mathrm{d}x = \int_{\mathbb{R}^{\geq 0}} r x^{r-1} L_f^+(x)\,\mathrm{d}x.$$

The next proposition is the main result of this section. The current statement of the proposition is proved by McCrudden as the main result in [132]. We give a simpler (but essentially the same) proof here for the sake of the completeness.

Proposition 7.23. *Let G be a unimodular group, $n_1, n_2 \geq 0$ are integers, H is a closed normal subgroup of G satisfying* $\mathrm{BM}(n_1)$*, and the quotient group G/H is unimodular satisfying* $\mathrm{BM}(n_2)$*. Then G satisfies* $\mathrm{BM}(n_1 + n_2)$*.*

Proof. Suppose Ω is a compact subset of G. Let the "fiber length function" $f_\Omega : G/H \to \mathbb{R}^{\geq 0}$ be a measurable function such that for every $gH \in G/H$, $f_\Omega(gH) = \mu_H(g^{-1}\Omega \cap H)$. The case when both $n_1 = n_2 = 0$ holds trivially.

Now we split the proof into three cases.

Case 1. *When $n_1 \geq 1$ and $n_1 + n_2 \geq 2$.*

By the quotient integral formula (Fact 5.7), we have

$$\mu_G^{1/(n_1+n_2)}(\Omega) = \left(\int_{G/H} f_\Omega(x)\,\mathrm{d}\mu_{G/H}(x)\right)^{1/(n_1+n_2)}$$

$$= \left(\int_{\mathbb{R}^{>0}} n_1 t^{n_1 - 1} \mu_{G/H}\big(L^+_{f_\Omega}(t^{n_1})\big)\, \mathrm{d}t \right)^{1/(n_1+n_2)}. \tag{7.14}$$

Set $\alpha = \frac{n_1-1}{n_1+n_2-1}$, $\beta = \frac{n_2}{n_1+n_2-1}$, $\gamma = n_1+n_2-1$, and

$$F_\Omega(t) = t^\alpha \mu_{G/H}^{\beta/n_2}\left(L^+_{f_\Omega}(t^{n_1})\right),$$

for compact set Ω in G and $t > 0$ (Note that F_Ω is well-defined when $n_2 = 0$). Then (7.14) can be rewritten as

$$n_1^{-1/(\gamma+1)} \mu_G^{1/(\gamma+1)}(\Omega) = \left(\int_{\mathbb{R}^{>0}} F_\Omega^\gamma(t)\, \mathrm{d}t \right)^{1/(\gamma+1)} \tag{7.15}$$

Fix nonempty compact sets $X, Y \subseteq G$. By (7.15), we need to show that

$$\left(\int_{\mathbb{R}^{>0}} F_{XY}^\gamma(t)\, \mathrm{d}t \right)^{1/(\gamma+1)} \geq \left(\int_{\mathbb{R}^{>0}} F_X^\gamma(t)\, \mathrm{d}t \right)^{1/(\gamma+1)} + \left(\int_{\mathbb{R}^{>0}} F_Y^\gamma(t)\, \mathrm{d}t \right)^{1/(\gamma+1)} \tag{7.16}$$

We will do so in two steps. First, we will show the following convexity property

$$F_{XY}(t_1+t_2) \geq F_X(t_1) + F_Y(t_2). \tag{7.17}$$

For every $t_1, t_2 \in \mathbb{R}^{>0}$, since H satisfies $\mathrm{BM}(n_1)$, by definition we have

$$L^+_{f_X}(t_1^{n_1}) L^+_{f_Y}(t_2^{n_1}) \subseteq L^+_{f_{XY}}\left((t_1+t_2)^{n_1}\right).$$

Also, since G/H satisfies $\mathrm{BM}(n_2)$, we have

$$\mu_{G/H}^{1/n_2}\left(L^+_{f_X}(t_1^{n_1})\right) + \mu_{G/H}^{1/n_2}\left(L^+_{f_Y}(t_2^{n_1})\right) \leq \mu_{G/H}^{1/n_2}\left(L^+_{f_{XY}}\left((t_1+t_2)^{n_1}\right)\right). \tag{7.18}$$

By Hölder's inequality and (7.18), as well as the fact that $n_1, n_2 \geq 1$, we obtain

$$\begin{aligned}
&(t_1+t_2)^{n_1-1}\left(\mu_{G/H}^{1/n_2}\big(L^+_{f_{XY}}((t_1+t_2)^{n_1})\big)\right)^{n_2} \\
\geq &(t_1+t_2)^{n_1-1}\left(\mu_{G/H}^{1/n_2}\left(L^+_{f_X}(t_1^{n_1})\right)+\mu_{G/H}^{1/n_2}\left(L^+_{f_Y}(t_2^{n_1})\right)\right)^{n_2} \\
= &\left\|\,(t_1^\alpha, t_2^\alpha)\,\right\|_{1/\alpha}^{\gamma}\left\|\left(\mu_{G/H}^{\beta/n_2}\left(L^+_{f_X}(t_1^{n_1})\right), \mu_{G/H}^{\beta/n_2}\left(L^+_{f_Y}(t_2^{n_1})\right)\right)\right\|_{1/\beta}^{\gamma} \\
\geq &\left(t_1^\alpha\mu_{G/H}^{\beta/n_2}\left(L^+_{f_X}(t_1^{n_1})\right)+t_2^\alpha\mu_{G/H}^{\beta/n_2}\left(L^+_{f_Y}(t_2^{n_1})\right)\right)^{\gamma}.
\end{aligned}$$

We remark that the above inequalities also make sense when $n_2 = 0$. In that case $\|(a,b)\|_{1/n_2}$ is to be understood as $\max\{a,b\}$ for every $a, b \in \mathbb{R}^{\geq 0}$. The first line of the above inequality is $F^\gamma_{XY}(t_1+t_2)$ and the last line is $(F_X(t_1)+F_Y(t_2))^\gamma$. So we finished the first step.

We now prove (7.16). By the above convexity property (7.17) and Kneser's inequality [115] for $\mathbb{R}$ (i.e. the Brunn–Minkowski inequality for $\mathbb{R}$), we have

$$\mu_{\mathbb{R}}\big(L^+_{F_{XY}}(s_1+s_2)\big) \geq \mu_{\mathbb{R}}\big(L^+_{F_X}(s_1)\big)+\mu_{\mathbb{R}}\big(L^+_{F_Y}(s_2)\big). \tag{7.19}$$

Let $M_X = \operatorname{ess\,sup}_x F_X(x)$, $M_Y = \operatorname{ess\,sup}_x F_Y(x)$. By Hölder's inequality and (7.19), we have

$$\begin{aligned}
\int_{\mathbb{R}^{>0}} F^\gamma_{XY}(s)\,\mathrm{d}s &\geq \int_0^{M_X+M_Y}\gamma s^{\gamma-1}\mu_{\mathbb{R}}\big(L^+_{F_{XY}}(s)\big)\,\mathrm{d}s \\
&= (M_X+M_Y)^\gamma\int_0^1\gamma s^{\gamma-1}\mu_{\mathbb{R}}\big(L^+_{F_{XY}}(M_Xs+M_Ys)\big)\,\mathrm{d}s \\
&\geq (M_X+M_Y)^\gamma\int_0^1\gamma s^{\gamma-1}\mu_{\mathbb{R}}\big(L^+_{F_X}(M_Xs)\big)\,\mathrm{d}s \\
&\quad + (M_X+M_Y)^\gamma\int_0^1\gamma s^{\gamma-1}\mu_{\mathbb{R}}\big(L^+_{F_Y}(M_Ys)\big)\,\mathrm{d}s \\
&= (M_X+M_Y)^\gamma\left(\frac{1}{M_X^\gamma}\int_{\mathbb{R}^{>0}}F^\gamma_X(s)\,\mathrm{d}s+\frac{1}{M_Y^\gamma}\int_{\mathbb{R}^{>0}}F^\gamma_Y(s)\,\mathrm{d}s\right).
\end{aligned} \tag{7.20}$$

Finally, by (7.14), (7.20) and Hölder's inequality,

$$n_1^{-1/(\gamma+1)}\mu_G^{1/(\gamma+1)}(XY)$$

$$
\begin{aligned}
&= \left(\int_{\mathbb{R}^{>0}} F_{XY}^{\gamma}(t)\,\mathrm{d}t \right)^{1/(\gamma+1)} \\
&\geq \left(\left((M_X^{\gamma})^{1/\gamma} + (M_Y^{\gamma})^{1/\gamma} \right)^{\gamma/(\gamma+1)} \left(\frac{1}{M_X^{\gamma}} \int_{\mathbb{R}^{>0}} F_X^{\gamma}(t)\,\mathrm{d}t + \frac{1}{M_Y^{\gamma}} \int_{\mathbb{R}^{>0}} F_Y^{\gamma}(t)\,\mathrm{d}t \right) \right)^{1/(\gamma+1)} \\
&\geq \left(\int_{\mathbb{R}^{>0}} F_X^{\gamma}(t)\,\mathrm{d}t \right)^{1/(\gamma+1)} + \left(\int_{\mathbb{R}^{>0}} F_Y^{\gamma}(t)\,\mathrm{d}t \right)^{1/(\gamma+1)} \\
&= n_1^{-1/(\gamma+1)} \mu_G^{1/(\gamma+1)}(X) + n_1^{-1/(\gamma+1)} \mu_G^{1/(\gamma+1)}(Y),
\end{aligned}
$$

this proves the case when n_1 is at least 1.

Case 2. *When $n_1 = 1$ and $n_2 = 0$.*

In this case, the conclusion can be derived from (7.14) directly. In particular, using the fact that G/H satisfies BM(0) and H satisfies BM(1), we have

$$
\mu_{G/H}\left(L^+_{f_{XY}}(t_1+t_2)\right) \geq \max\left\{ \mu_{G/H}\left(L^+_{f_X}(t_1)\right), \mu_{G/H}\left(L^+_{f_Y}(t_2)\right) \right\}.
$$

Let $N_X = \sup_t f_X(t)$ and $N_Y = \sup_t f_Y(t)$. Therefore, by Hölder's inequality,

$$
\begin{aligned}
\mu_G(XY) &= \int_{\mathbb{R}^{>0}} \mu_{G/H}(L^+_{f_{XY}}(t))\,\mathrm{d}t \\
&= \int_{\mathbb{R}^{>0}} (N_X+N_Y)\mu_{G/H}(L^+_{f_{XY}}((N_X+N_Y)t))\,\mathrm{d}t \\
&\geq (N_X+N_Y) \max\left\{ \int_0^1 \mu_{G/H}(L^+_{f_X}(N_X t)\,\mathrm{d}t, \int_0^1 \mu_{G/H}(L^+_{f_Y}(N_Y t)\,\mathrm{d}t \right\} \\
&\geq N_X \int_0^1 \mu_{G/H}(L^+_{f_X}(N_X t))\,\mathrm{d}t + N_Y \int_0^1 \mu_{G/H}(L^+_{f_Y}(N_Y t))\,\mathrm{d}t \\
&= \mu_G(X) + \mu_G(Y).
\end{aligned}
$$

Thus G satisfies BM(1).

Case 3. *When $n_1 = 0$ and $n_2 \geq 1$.*

Applying Brunn–Minkowski inequality with exponent 0 on H, and the fact that G/H

satisfies BM(n_2), we obtain

$$\mu_{G/H}^{1/n_2}\left(L^+_{f_{XY}}(\max\{t_1,t_2\})\right) \geq \mu_{G/H}^{1/n_2}\left(L^+_{f_X}(t_1)\right) + \mu_{G/H}^{1/n_2}\left(L^+_{f_Y}(t_1)\right). \tag{7.21}$$

Given a compact set Ω in G, we define

$$E_\Omega(t) = \mu_{G/H}^{1/n_2}(L^+_{f_\Omega}(t)), t > 0.$$

Thus by (7.21), we have $E_{XY}(\max\{a_1,a_2\}) \geq E_X(a_1) + E_Y(a_2)$ for all a_1, a_2. This can be seen as a "convexity property" for E, but the maximum operator insider the function E prevent us from using the same argument as used in Case 1 for F. On the other hand, we observe that

$$\mu_{\mathbb{R}}(L^+_{E_{XY}}(s_1+s_2)) \geq \max\{\mu_{\mathbb{R}}(L^+_{E_X}(s_1)), \mu_{\mathbb{R}}(L^+_{E_Y}(s_2))\}. \tag{7.22}$$

Now we consider $\mu_G(XY)$. We have

$$\mu_G^{1/n_2}(XY) = \left(\int_{\mathbb{R}^{>0}} E_{XY}^{n_2}(s)\,\mathrm{d}s\right)^{1/n_2} = \left(\int_{\mathbb{R}^{>0}} n_2 s^{n_2-1}\mu_{\mathbb{R}}(L^+_{E_{XY}}(s))\,\mathrm{d}s\right)^{1/n_2} \tag{7.23}$$

Let $P_X = \operatorname{ess\,sup}_t E_X(t)$ and $P_Y = \operatorname{ess\,sup}_t E_Y(t)$. By (7.22) and (7.23) we see

$$\begin{aligned}
&n_2^{-1/n_2}\mu_G^{1/n_2}(XY)\\
\geq &\left((P_X+P_Y)^{n_2}\max\left\{\int_0^1 s^{n_2-1}\mu_{\mathbb{R}}(L^+_{E_X}(P_X s)\,\mathrm{d}s, \int_0^1 s^{n_2-1}\mu_{\mathbb{R}}(L^+_{E_Y}(P_Y s)\,\mathrm{d}s\right\}\right)^{1/n_2}\\
\geq &\left(P_X^{n_2}\int_0^1 s^{n_2-1}\mu_{\mathbb{R}}(L^+_{E_X}(P_X s))\,\mathrm{d}s\right)^{1/n_2} + \left(P_Y^{n_2}\int_0^1 s^{n_2-1}\mu_{\mathbb{R}}(L^+_{E_Y}(P_Y s))\,\mathrm{d}s\right)^{1/n_2}\\
= &n_2^{-1/n_2}\mu_G^{1/n_2}(X) + n_2^{-1/n_2}\mu_G^{1/n_2}(Y).
\end{aligned}$$

This proves the case when $n_1 = 0$, and hence finishes the proof of the proposition. □

Using a similar technique as used in the proof of Proposition 7.23, we are able to reduce the problem to open subgroups.

Proposition 7.24. *Let G be a unimodular group, and let G' be an open subgroup of G. Suppose G' satisfies* BM(n) *for some integer $n \geq 0$, then G satisfies* BM(n).

Proof. When $n = 0$, the conclusion follows from $\mu(XY) \geq \mu(Y)$. In the remaining time we assume $n \geq 1$.

Let μ_G be a Haar measure on G, and let $\mu_{G'}$ be the restricted Haar measure of μ_G on G'. By Fact 5.1.1, for every compact set Ω in G we have

$$\mu_G(\Omega) = \sum_{g \in G/G'} \mu_{G'}(g\Omega \cap G').$$

We similarly define $f_\Omega : G/G' \to \mathbb{R}^{\geq 0}$ such that $f_\Omega(g) = \mu_{G'}(g^{-1}\Omega \cap G')$.

Fix two compact sets X, Y in G. Using the fact that G' satisfies BM(n), we have

$$\left|L^+_{f_{XY}}\left((t_1+t_2)^n\right)\right| \geq \max\left\{\left|L^+_{f_X}(t_1^n)\right|, \left|L^+_{f_Y}(t_2^n)\right|\right\}$$

because if $f_X(g_1), \dots, f_X(g_k) \geq t_1^n$ and $f_Y(\tilde{g}) \geq t_2^n$ we have $f_{XY}(g_1\tilde{g}), \dots, f_{XY}(g_k\tilde{g}) \geq (t_1 + t_2)^n$.

Let $N_X = \sup_g f_X(g)$ and $N_Y = \sup_g f_Y(g)$. By the above inequality we deduce

$$\begin{aligned}
&n^{-1/n}\mu_G^{1/n}(XY) \\
={}& \left(\int_{\mathbb{R}^{>0}} t^{n-1}|L^+_{f_{XY}}(t^n)|\,\mathrm{d}t\right)^{1/n} \\
\geq{}& \left((N_X+N_Y)^n \max\left\{\int_0^1 t^{n-1}|L^+_{f_X}((N_Xt)^n|\,\mathrm{d}t, \int_0^1 t^{n-1}|L^+_{f_Y}((N_Yt)^n|\,\mathrm{d}t\right\}\right)^{1/n} \\
\geq{}& \left(N_X^n\int_0^1 t^{n-1}|L^+_{f_X}((N_Xt)^n)|\,\mathrm{d}t\right)^{1/n} + \left(N_Y^n\int_0^1 t^{n-1}|L^+_{f_Y}((N_Yt)^n)|\,\mathrm{d}t\right)^{1/n} \\
={}& n^{-1/n}\mu_G^{1/n}(X) + n^{-1/n}\mu_G^{1/n}(Y),
\end{aligned}$$

Thus G satisfies BM(n). ∎

7.5 Reduction to unimodular subgroups

The main result of this section allows us to obtain a Brunn–Minkowski inequality for a nonunimodular group from its certain unimodular normal subgroup. We use $\mu_{\mathbb{R}^\times}$ to denote a Haar measure on the multiplicative group $(\mathbb{R}^{>0}, \times)$. The next lemma concerns the case when the modular function on X and on Y are "sufficiently uniform".

Lemma 7.25. *Suppose the modular function $\Delta_G : G \to (\mathbb{R}^{>0}, \times)$ is a quotient map of topological groups. Let X, Y be compact subsets of G, and parameters $a, b, \varepsilon > 0$ and $n \geq 0$ an integer, such that for every $x \in X$, $\Delta_G(x) \in [a, a+\varepsilon)$ and for every $y \in Y$, $\Delta_G(y) \in [b, b+\varepsilon)$. Suppose $H = \ker(\Delta_G)$ satisfying* $\mathrm{BM}(n)$. *Then*

$$\frac{\nu_G(X)^{1/(n+1)}}{\nu_G(XY)^{1/(n+1)}} + \frac{\mu_G(Y)^{1/(n+1)}}{\mu_G(XY)^{1/(n+1)}} \leq 1 + f(\varepsilon),$$

where $f(\varepsilon)$ is an explicit function depending only on a, b, n and ε, and $f(\varepsilon) \to 0$ as $\varepsilon \to 0$. Moreover, this convergence is uniform when n is fixed and a and b vary over compact sets.

Proof. We first consider the case when $n \geq 1$. For every compact subset Ω of G, define two functions $\ell_\Omega, r_\Omega : (\mathbb{R}^{>0}, \times) \to \mathbb{R}^{\geq 0}$ such that

$$\ell_\Omega(g) = \mu_H(g^{-1}\Omega \cap H), \text{ and } r_\Omega(g) = \mu_H(\Omega g^{-1} \cap H).$$

Note that given g_1, g_2 in G, note that $(X \cap Hg_1)\cdot(Y \cap g_2 H)$ lies in

$$Hg_1g_2H = (g_1g_2)(g_1g_2)^{-1}H(g_1g_2)H = H(g_1g_2)H(g_1g_2)^{-1}(g_1g_2)$$

since H is normal. Now we fix Haar measures $\mu_H, \mu_{\mathbb{R}^\times}$ on H and on $(\mathbb{R}^{>0}, \times)$, and these two measures will also uniquely determine a left Haar measure μ_G on G and a right Haar measure ν_G on G via the quotient integral formula.

For every compact sets X_1, X_2 in H, and g_1, g_2 in G, by the above equality, $X_1g_1g_2X_2 \subseteq$

g_1g_2H. By Fact 5.8.2 and the fact that H satisfies BM(n), we have

$$\begin{aligned}\mu_H^{1/n}((g_1g_2)^{-1}X_1g_1g_2X_2) &\geq \mu_H^{1/n}((g_1g_2)^{-1}X_1g_1g_2)+\mu_H^{1/n}(X_2)\\ &= (\Delta_G(g_1)\Delta_G(g_2))^{-1/n}\mu_H^{1/n}(X_1)+\mu_H^{1/n}(X_2).\end{aligned}\tag{7.24}$$

In light of this, applying the Brunn–Minkowski inequality on $(\mathbb{R}^{>0},\times)$, we get

$$\mu_{\mathbb{R}^\times}\left(L^+_{\ell_{XY}}\left(\left(\inf_{x\in X,y\in Y}(\Delta_G(x)\Delta_G(y))^{-1/n}t_1+t_2\right)^n\right)\right)\geq \mu_{\mathbb{R}^\times}(L^+_{r_X}((t_1)^n))+\mu_{\mathbb{R}^\times}(L^+_{\ell_Y}(t_2^n)),$$

and similarly for right Haar measure on H, we have

$$\mu_{\mathbb{R}^\times}\left(L^+_{r_{XY}}\left(\left(t_1+\inf_{x\in X,y\in Y}(\Delta_G(x)\Delta_G(y))^{1/n}t_2\right)^n\right)\right)\geq \mu_{\mathbb{R}^\times}(L^+_{r_X}((t_1)^n))+\mu_{\mathbb{R}^\times}(L^+_{\ell_Y}(t_2^n)).$$

Let $M_X=\sup_x \mu_{\mathbb{R}^\times}(L^+_{r_X}(x))$ and $M_Y=\sup_y \mu_{\mathbb{R}^\times}(L^+_{\ell_Y}(y))$. By a change of variables and then by the first inequality above, we have

$$\begin{aligned}\mu_G(XY) &= \int_{\mathbb{R}^\times}\mu_H(g^{-1}XY\cap H)\,\mathrm{d}\mu_{\mathbb{R}^\times}(g)\\ &= \int_{\mathbb{R}^{>0}} nt^{n-1}\mu_{\mathbb{R}^\times}(L^+_{\ell_{XY}}(t^n))\,\mathrm{d}t\\ &\geq \left\|\left(\frac{1}{(a+\varepsilon)(b+\varepsilon)}M_X,M_Y\right)\right\|_{1/n}\\ &\quad\cdot\int_0^1 nt^{n-1}\mu_{\mathbb{R}^\times}\left(L^+_{\ell_{XY}}\left(\left(\left(\frac{1}{(a+\varepsilon)(b+\varepsilon)}M_X\right)^{1/n}t+M_Y^{1/n}t\right)^n\right)\right)\,\mathrm{d}t\\ &\geq \left\|\left(\frac{1}{(a+\varepsilon)(b+\varepsilon)}M_X,M_Y\right)\right\|_{1/n}\left(\frac{1}{M_X}\nu_G(X)+\frac{1}{M_Y}\mu_G(Y)\right).\end{aligned}$$

Thus by Hölder's inequality, we get

$$\mu_G^{1/(n+1)}(XY)\geq\left(\frac{1}{(a+\varepsilon)(b+\varepsilon)}\nu_G(X)\right)^{1/(n+1)}+\mu_G^{1/(n+1)}(Y).\tag{7.25}$$

Similarly, for $\nu_G(XY)$ we have

$$\begin{aligned}\nu_G(XY) &= \int_{\mathbb{R}^\times} \mu_H(XYg^{-1}\cap H)\,\mathrm{d}\mu_{\mathbb{R}^\times}(g)\\ &= \int_{\mathbb{R}^{>0}} nt^{n-1}\mu_{\mathbb{R}^\times}(L^+_{r_{XY}}(t^n))\,\mathrm{d}t\\ &\geq \left\|\Big(M_X, abM_Y\Big)\right\|_{1/n}\left(\frac{1}{M_X}\nu_G(X)+\frac{1}{M_Y}\mu_G(Y)\right),\end{aligned}$$

and we obtain

$$\nu_G^{1/(n+1)}(XY) \geq \nu_G^{1/(n+1)}(X) + \big(ab\mu_G(Y)\big)^{1/(n+1)}. \tag{7.26}$$

Therefore, combining (7.25) and (7.26), we conclude

$$\begin{aligned}&\frac{\nu_G^{1/(n+1)}(X)}{\nu_G^{1/(n+1)}(XY)} + \frac{\mu_G^{1/(n+1)}(Y)}{\mu_G^{1/(n+1)}(XY)}\\ \leq\, &\frac{1}{1+(Cab)^{1/(n+1)}} + \frac{1}{1+\big(\frac{1}{C(a+\varepsilon)(b+\varepsilon)}\big)^{1/(n+1)}}\\ \leq\, &1+\frac{(C(ab+\varepsilon(a+b+\varepsilon)))^{1/(n+1)}-(Cab)^{1/(n+1)}}{(1+(Cab)^{1/(n+1)})(1+(C(a+\varepsilon)(b+\varepsilon))^{1/(n+1)})}.\end{aligned}$$

where $C=\mu_G(Y)/\nu_G(X)$.

Hence

$$\frac{\nu_G^{1/(n+1)}(X)}{\nu_G^{1/(n+1)}(XY)} + \frac{\mu_G^{1/(n+1)}(Y)}{\mu_G^{1/(n+1)}(XY)} \leq 1+f(\varepsilon)$$

where

$$f(\varepsilon)=\sup_{r>0}\frac{(r(ab+\varepsilon(a+b+\varepsilon)))^{1/(n+1)}-(rab)^{1/(n+1)}}{(1+(rab)^{1/(n+1)})(1+(r(a+\varepsilon)(b+\varepsilon))^{1/(n+1)})}$$

depends only on a,b,n and ε and we see $\lim_{\varepsilon\to 0} f(\varepsilon)=0$ uniformly when a,b taken values in a compact set by an elementary computation.

The remaining case is when $n = 0$. Note that in this case, inequality (7.24) becomes

$$\mu_H((g_1g_2)^{-1}X_1g_1g_2X_2) \geq \max\{(\Delta_G(g_1)\Delta_G(g_2))^{-1}\mu_H(X_1), \mu_H(X_2)\}.$$

This implies for every t_1, t_2,

$$\mu_{\mathbb{R}^\times}\left(L^+_{\ell_{XY}} \max\left\{\inf_{x\in X, y\in Y}(\Delta_G(x)\Delta_G(y))^{-1}t_1, t_2\right\}\right) \geq \mu_{\mathbb{R}^\times}(L^+_{r_X}(t_1)) + \mu_{\mathbb{R}^\times}(L^+_{\ell_Y}(t_2)).$$

For any compact set Ω in G, define two functions $\Phi_\Omega, \Psi_\Omega : \mathbb{R} \to \mathbb{R}$, that

$$\Phi_\Omega(t) = \mu_{\mathbb{R}^\times}(L^+_{\ell_\Omega}(t)), \quad \text{and} \quad \Psi_\Omega(t) = \mu_{\mathbb{R}^\times}(L^+_{r_\Omega}(t)).$$

Thus we have

$$\mu_{\mathbb{R}}(L^+_{\Phi_{XY}}(t_1+t_2)) \geq \max\left\{\inf_{x\in X, y\in Y}(\Delta_G(x)\Delta_G(y))^{-1}\mu_{\mathbb{R}}(L^+_{\Psi_X}(t_1)), \mu_{\mathbb{R}}(L^+_{\Phi_Y}(t_2))\right\}.$$

Let $N_X = \sup_x \mu_{\mathbb{R}}(L^+_{\Psi_X}(x))$ and $N_Y = \sup_y \mu_{\mathbb{R}}(L^+_{\Phi_Y}(y))$. By a change of variable, for $\mu_G(XY)$ we have

$$\begin{aligned}
\mu_G(XY) &= \int_{\mathbb{R}^{>0}} \mu_{\mathbb{R}}(L^+_{\Phi_{XY}}(t))\,\mathrm{d}t \\
&\geq (N_X+N_Y)\max\left\{\frac{1}{(a+\varepsilon)(b+\varepsilon)}\frac{\nu_G(X)}{N_X}, \frac{\mu_G(Y)}{N_Y}\right\} \\
&\geq \frac{1}{(a+\varepsilon)(b+\varepsilon)}\nu_G(X) + \mu_G(Y). \qquad (7.27)
\end{aligned}$$

Similarly, for every t_1, t_2 we also have

$$\mu_{\mathbb{R}^\times}\left(L^+_{r_{XY}} \max\left\{t_1, \inf_{x\in X, y\in Y}\Delta_G(x)\Delta_G(y)t_2\right\}\right) \geq \mu_{\mathbb{R}^\times}(L^+_{r_X}(t_1)) + \mu_{\mathbb{R}^\times}(L^+_{\ell_Y}(t_2)),$$

which implies

$$\mu_{\mathbb{R}}(L^+_{\Psi_{XY}}(t_1+t_2)) \geq \max\left\{\mu_{\mathbb{R}}(L^+_{\Psi_X}(t_1)), \inf_{x\in X, y\in Y} \Delta_G(x)\Delta_G(y)\mu_{\mathbb{R}}(L^+_{\Phi_Y}(t_2))\right\}.$$

Therefore, for $\nu_G(XY)$ we get

$$\nu_G(XY) \geq \nu_G(X) + ab\mu_G(Y).$$

Together with (7.27), similarly as in the case when $n \geq 1$, we get

$$\begin{aligned}\frac{\nu_G(X)}{\nu_G(XY)} + \frac{\mu_G(Y)}{\mu_G(XY)} &\leq \frac{1}{1+Cab} + \frac{1}{1+\frac{1}{C(a+\varepsilon)(b+\varepsilon)}} \\ &\leq 1 + \frac{\varepsilon C(a+b+\varepsilon)}{(1+Cab)(1+C(a+\varepsilon)(b+\varepsilon))},\end{aligned}$$

where $C = \mu_G(Y)/\nu_G(X)$. The conclusion follows by taking

$$f(\varepsilon) = \sup_{r>0} \frac{\varepsilon r(a+b+\varepsilon)}{(1+rab)(1+r(a+\varepsilon)(b+\varepsilon))},$$

and we can see that $f(\varepsilon) \to 0$ as $\varepsilon \to 0$ uniformly when a, b taken values in a compact set by elementary computations. □

The next proposition is the main result of the section. As we mentioned in the introduction, the proof uses a discretized "spillover" method. We remark that one can always make the proof continuous like what we did in Section 7.4, but we give a discrete proof here since we believe this reflects our idea in a clearer way.

Proposition 7.26. *Suppose G is a locally compact group with $H = \ker(\Delta_G)$ satisfying* BM(n). *Suppose the map $\Delta_G : G \to (\mathbb{R}^{>0}, \times)$ is a quotient map of topological groups, then G satisfies* BM($n+1$).

Proof. Since X and Y are compact, there are a_1, a_2, b_1 and $b_2 > 0$, such that

$$a_1 = \inf_{x \in X} \Delta_G(x),\ a_2 = \sup_{x \in X} \Delta_G(x),\ b_1 = \inf_{y \in Y} \Delta_G(y),\ b_2 = \sup_{y \in Y} \Delta_G(y).$$

We fix μ_G and ν_G as in the proof of Lemma 7.25, and let $\varepsilon > 0$ be a sufficient small number (depending on a_1, a_2, b_1 and b_2). Then by Fact 5.7 and familiar properties of integrable functions on $\mathbb{R}$, there is an $N > 0$, such that we can partition $[a_1, a_2]$ and $[b_1, b_2]$ into N subintervals, that is

$$[a_1, a_2] = \bigcup_{i=1}^{N} A_i, \quad [b_1, b_2] = \bigcup_{i=1}^{N} B_i,$$

such that each subinterval has length at most ε, and the intersection of X with $\bigcup_{g \in A_i} Hg$ has ν_G-measure $\nu_G(X)/N$, the intersection of Y with $\bigcup_{g \in B_i} gH$ has μ_G-measure $\mu_G(Y)/N$, for every $1 \le i \le N$.

Let $X_i = X \cap HA_i$ and let $Y_i = Y \cap B_i H$. Then $\nu_G(X) = \sum_{i=1}^{N} \nu_G(X_i)$ and $\mu_G(Y) = \sum_{i=1}^{N} \mu_G(Y_i)$. In particular, we have

$$\mu_G(XY) \ge \sum_{i=1}^{N} \mu_G(X_i Y_i) \quad \text{and} \quad \nu_G(XY) \ge \sum_{i=1}^{N} \nu_G(X_i Y_i).$$

Observe that given $1 \le i, j \le N$ and $i \ne j$, $X_i Y_i$ and $X_j Y_j$ are disjoint. Indeed, the modulus of every element in $X_i Y_i$ lies in $A_i B_i$ and the modulus of every element in $X_j Y_j$ lies in $A_j B_j$. But $A_i B_i$ and $A_j B_j$ are disjoint subsets of $\mathbb{R}^{>0}$ when $i \ne j$.

By Lemma 7.25, for every $1 \le i \le N$, there is a function $f_i(\varepsilon)$, such that $f_i(\varepsilon) \to 0$ when $\varepsilon \to 0$ uniformly, and

$$\frac{\nu_G^{1/(n+1)}(X_i)}{\nu_G^{1/(n+1)}(X_i Y_i)} + \frac{\mu_G^{1/(n+1)}(Y_i)}{\mu_G^{1/(n+1)}(X_i Y_i)} \le 1 + f_i(\varepsilon).$$

Take $\tilde{f}(\varepsilon) = \sup_i f_i(\varepsilon)$, hence $\tilde{f}(\varepsilon) \to 0$ as $\varepsilon \to 0$. Therefore, for every $1 \le t \le N$,

$$\frac{\nu_G^{1/(n+1)}(X)}{\nu_G^{1/(n+1)}(XY)} + \frac{\mu_G^{1/(n+1)}(Y)}{\mu_G^{1/(n+1)}(XY)} \le \left(\frac{N\nu_G(X_t)}{\sum_{i=1}^N \nu_G(X_iY_i)} \right)^{\frac{1}{n+1}} + \left(\frac{N\mu_G(Y_t)}{\sum_{i=1}^N \mu_G(X_iY_i)} \right)^{\frac{1}{n+1}}. \quad (7.28)$$

Also by Hölder's inequality, we observe that for every t,

$$\left(\sum_{i=1}^N \left(\frac{\nu_G(X_i)}{\nu_G(X_iY_i)} \right)^{\frac{1}{n+2} \cdot \frac{n+2}{n+1}} \right)^{\frac{n+1}{n+2}} \left(\sum_{i=1}^N \nu_G(X_iY_i) \right)^{\frac{1}{n+2}} \ge N\nu_G^{\frac{1}{n+2}}(X_t). \quad (7.29)$$

Averaging (7.28) over all t and using inequality (7.29), we have

$$\begin{aligned} &\frac{\nu_G^{1/(n+1)}(X)}{\nu_G^{1/(n+1)}(XY)} + \frac{\mu_G^{1/(n+1)}(Y)}{\mu_G^{1/(n+1)}(XY)} \\ \le &\frac{1}{N} \sum_{i=1}^N \left(\frac{\nu_G(X_i)}{\nu_G(X_iY_i)} \right)^{1/(n+1)} + \frac{1}{N} \sum_{i=1}^N \left(\frac{\mu_G(X_i)}{\mu_G(X_iY_i)} \right)^{1/(n+1)} \le 1 + \tilde{f}(\varepsilon). \end{aligned}$$

The desired conclusion follows by taking $\varepsilon \to 0$. □

7.6 Reduction to cocompact and codiscrete subgroups

The main results in this section will help us to reduce the problem to cocompact subgroups or open normal subgroups. We make use of the following integral formula, see [113, Proposition 5.26, Consequence 1].

Fact 7.27. *Let G be a connected unimodular Lie group. Suppose S, T are closed subgroups of G, such that $G = ST$, and the intersection $S \cap T$ is compact. Then there is a left Haar measure μ_S on S and a right Haar measure ν_T on T, such that*

$$\int_G f(x)\, \mathrm{d}\mu_G(x) = \int_{S\times T} f(st)\, \mathrm{d}\mu_S(s)\, \mathrm{d}\nu_T(t),$$

for every $f \in C_c(G)$.

The next proposition allows us to reduce the problem to closed cocompact subgroups with the same noncomapct Lie dimension.

Proposition 7.28. *Suppose G is connected unimodular Lie group, H is a connected closed subgroup of G satisfying* $\mathrm{BM}(n)$*, K is a connected unimodular subgroup of G, such that $G = KH$ and $K \cap H$ is compact. Then G satisfies* $\mathrm{BM}(n)$.

Proof. We assume $n \geq 1$, otherwise the result is trivial. Note that both G and K are unimodular. In light of this we will not be using ν_G, ν_K, etc. and only use $\mu_G = \nu_G$ and $\mu_K = \nu_K$ below.

We fix a Haar measure μ_K on K, and a Haar measure μ_G on G. These measures will also uniquely determine a left Haar measure μ_H and a right Haar measure ν_H on H such that we have the integral formula in Fact 7.27 and another similar formula involving $\mathrm{d}\mu_H(h)\,\mathrm{d}\mu_K(k)$. For a compact subset Ω of G, we define two functions $r_\Omega, \ell_\Omega : K \to \mathbb{R}^{\geq 0}$, such that

$$r_\Omega(k) := \nu_H(k\Omega \cap H), \quad \ell_\Omega(k) := \mu_H(\Omega k \cap H),$$

for every $k \in K$. We also define two bivariate functions $R_\Omega, L_\Omega : K \times K \to \mathbb{R}^{\geq 0}$ that for every k_1, k_2 in K,

$$R_\Omega(k_1, k_2) := \nu_H(k_1 \Omega k_2 \cap H), \quad L_\Omega(k_1, k_2) := \mu_H(k_1 \Omega k_2 \cap H).$$

Thus Fact 7.27 gives us

$$\mu_G(\Omega) = \int_K \nu_H(k^{-1}\Omega \cap H)\,\mathrm{d}\mu_K(k) = \int_K \mu_H(\Omega k^{-1} \cap H)\,\mathrm{d}\mu_K(k).$$

We define two probability measures p_X and p_Y on K in the following way:

$$\mathrm{dp}_X = \frac{r_X\,\mathrm{d}\mu_K}{\mu_G(X)}, \quad \mathrm{dp}_Y = \frac{\ell_Y\,\mathrm{d}\mu_K}{\mu_G(Y)}.$$

Now, we choose a left coset k_1H of H in G randomly with respect to the probability measure p_X, and choose a right coset Hk_2 of H in G randomly with respect to the probability measure p_Y. By the fact that H satisfies BM(n), we get

$$\left(\frac{r_X(k_1)}{R_{XY}(k_1,k_2)}\right)^{1/n} + \left(\frac{\ell_Y(k_2)}{L_{XY}(k_1,k_2)}\right)^{1/n} \leq 1.$$

This implies

$$\mathbb{E}_{\mathrm{p}_X(k_1)}\mathbb{E}_{\mathrm{p}_Y(k_2)}\left[\left(\frac{r_X(k_1)}{R_{XY}(k_1,k_2)}\right)^{1/n} + \left(\frac{\ell_Y(k_2)}{L_{XY}(k_1,k_2)}\right)^{1/n}\right] \leq 1. \tag{7.30}$$

On the other hand, by Hölder's inequality, Fact 7.27 and the fact that G is unimodular,

$$\begin{aligned}
&\mathbb{E}_{\mathrm{p}_X(k_1)}\left(\frac{r_X(k_1)}{R_{XY}(k_1,k_2)}\right)^{\frac{1}{n}}\\
={}&\frac{1}{\mu_G(X)}\int_K r_X^{\frac{n+1}{n}}(k_1)R_{XY}^{-\frac{1}{n}}(k_1,k_2)\,\mathrm{d}\mu_K(k_1)\\
\geq{}&\frac{1}{\mu_G(X)}\left(\int_K r_X(k_1)\,\mathrm{d}\mu_K(k_1)\cdot\left(\int_K R_{XY}(k_1,k_2)\,\mathrm{d}\mu_K(k_1)\right)^{-\frac{1}{n+1}}\right)^{\frac{n+1}{n}}\\
={}&\left(\frac{\mu_G(X)}{\mu_G(XY)}\right)^{\frac{1}{n}}.
\end{aligned}$$

We have a similar inequality concerning $\mathbb{E}_{\mathrm{p}_Y(k_2)}\left(\frac{\ell_Y(k_2)}{L_{XY}(k_1,k_2)}\right)^{\frac{1}{n}}$. Combining both inequalities with (7.30), we get

$$\left(\frac{\mu_G(X)}{\mu_G(XY)}\right)^{\frac{1}{n}} + \left(\frac{\mu_G(Y)}{\mu_G(XY)}\right)^{\frac{1}{n}} \leq 1,$$

and hence G satisfies BM(n). $\square$

Using the proportionated averaging trick in a similar fashion, the next result allows us to reduce the problem to certain open subgroups.

Proposition 7.29. *Let G be a locally compact group, and let G' be an open normal unimodular subgroup of G. Suppose G' satisfies* $\mathrm{BM}(n)$ *for some integer $n \geq 1$. Then G satisfies* $\mathrm{BM}(n)$.

Proof. Let $\mu_{G'}$ be a left (and hence right) Haar measure on G'. By Fact 5.8.2, there is a left Haar measure μ_G and a right Haar measure ν_G on G, such that for every compact set Ω in G we have

$$\nu_G(\Omega) = \sum_{g \in G/G'} \Delta_G(g^{-1})\mu_{G'}(g^{-1}\Omega \cap G'), \quad \mu_G(\Omega) = \sum_{g \in G'\backslash G} \Delta_G(g)\mu_{G'}(\Omega g^{-1} \cap G').$$

Now we fix two compact sets X, Y in G. For every $g \in G/G'$, let $X_g = g^{-1}X \cap G'$, and we similarly define $Y_h = Xh^{-1} \cap G'$ for every $h \in G'\backslash G$. Since G' satisfies $\mathrm{BM}(n)$, we have that

$$\left(\frac{\mu_{G'}(X_g)}{\mu_{G'}(X_gY_h)}\right)^{1/n} + \left(\frac{\mu_{G'}(Y_h)}{\mu_{G'}(X_gY_h)}\right)^{1/n} \leq 1. \tag{7.31}$$

Now we choose g from G/G' randomly with probability $\mathrm{p}_X(g) = \frac{\Delta_G(g^{-1})\mu_{G'}(X_g)}{\nu_G(X)}$. Therefore by Hölder's inequality,

$$\begin{aligned}\mathbb{E}_{\mathrm{p}_X(g)}\left(\frac{\mu_{G'}(X_g)}{\mu_{G'}(X_gY_h)}\right)^{\frac{1}{n}} &= \frac{1}{\nu_G(X)}\sum_{g \in G/G'} \frac{(\mu_{G'}(X_g)\Delta_G(g^{-1}))^{\frac{n+1}{n}}}{(\mu_{G'}(X_gY_h)\Delta_G(g^{-1}))^{\frac{1}{n}}} \\ &\geq \left(\frac{\nu_G(X)}{\nu_G(XYh)}\right)^{\frac{1}{n}} = \left(\frac{\nu_G(X)}{\nu_G(XY)}\right)^{\frac{1}{n}}.\end{aligned}$$

Similarly, we choose h from $G'\backslash G$ randomly with probability $\mathrm{p}_Y(h) = \frac{\Delta_G(h)\mu_{G'}(Y_h)}{\mu_G(Y)}$. Again using Hölder's inequality, we conclude that

$$\mathbb{E}_{\mathrm{p}_Y(h)}\left(\frac{\mu_{G'}(Y_h)}{\mu_{G'}(X_gY_h)}\right)^{\frac{1}{n}} \geq \left(\frac{\mu_G(Y)}{\mu_G(XY)}\right)^{\frac{1}{n}}.$$

Hence by (7.31),

$$\begin{aligned}&\left(\frac{\nu_G(X)}{\nu_G(XY)}\right)^{\frac{1}{n}}+\left(\frac{\mu_G(Y)}{\mu_G(XY)}\right)^{\frac{1}{n}}\\ \le\,&\mathbb{E}_{\mathrm{p}_X(g)}\mathbb{E}_{\mathrm{p}_Y(h)}\left[\left(\frac{\mu_{G'}(X_g)}{\mu_{G'}(X_gY_h)}\right)^{1/n}+\left(\frac{\mu_{G'}(Y_h)}{\mu_{G'}(X_gY_h)}\right)^{1/n}\right]\le 1,\end{aligned}$$

and thus G also satisfies BM(n). $\square$

7.7 Proof of Theorems 7.1, 7.2, and 7.5

7.7.1 A dichotomy lemma

In this subsection, we prove a dichotomy result for the kernel of a continuous homomorphism to $(\mathbb{R}^{>0},\times)$.

The following lemma records a fact on open maps between locally compact groups.

Lemma 7.30. *Suppose G, H are locally compact groups, $\phi: G\to H$ is a continuous and surjective group homomorphism, and there is an open subgroup G' of G such that $\phi|_{G'}$ is open. Then $\phi: G\to H$ is a quotient map of locally compact groups.*

Proof. By the first isomorphism theorem (Fact 5.1.1), it suffices to check that ϕ is open. Suppose U is an open subset of G. Then $U=\bigcup_{a\in G} U\cap aG'$. For each $a\in G$, we have

$$\phi(U\cap aG')=\phi(a)\phi|_{G'}(a^{-1}U\cap G').$$

As $\phi|_{G'}$ is open, $\phi(U\cap aG')$ is open for each $a\in G$. Hence, $\phi(U)=\bigcup_{a\in G}\phi(U\cap aG')$ is open in H, which is the desired conclusion. $\square$

In the next lemma we present our main dichotomy result.

Lemma 7.31. *If G is a locally compact group, and $\pi: G\to(\mathbb{R}^{>0},\times)$ is a continuous group homomorphism. Then exactly one of the following holds:*

1. we have the short exact sequence of locally compact groups

$$1 \to \ker \pi \to G \xrightarrow{\pi} (\mathbb{R}^{>0}, \times) \to 1;$$

2. $\ker \pi$ *is an open subgroup of* G*.*

Proof. It is easy to see that (1) and (2) are mutually disjoint, so we need to prove that we are always either in (1) or (2). Consider first the case when G is a Lie group. Let G_0 be the identity component of G. Then G_0 is open by Fact 5.12. Hence $\pi(G_0)$ is a connected subgroup of $(\mathbb{R}^{>0}, \times)$. As the only connected subsets of $(\mathbb{R}^{>0}, \times)$ are points and intervals, we deduce that $\pi(G_0)$ can only be $\{1\}$ or $(\mathbb{R}^{>0}, \times)$. In the former case, $\ker \pi$ is open as a union of translations of G_0. Now suppose $\pi(G_0) = (\mathbb{R}^{>0}, \times)$. Since G_0 is a connected Lie group. Using the first isomorphism theorem for Lie group (Fact 5.13.1), we get $\pi|_{G_0}$ is open. Applying Lemma 7.30, we get that π is a quotient map as desired.

We now deal with the general situation where G is locally compact. Using the Gleason–Yamabe Theorem (Fact 5.10.1), we obtain an almost-Lie open subgroup G' of G. Since G' is open, the natural embedding of $i : G' \to G$ induces a continuous homomorphism $\pi|_{G'} : G' \to (\mathbb{R}^{>0}, \times)$. Note that there is a compact normal subgroup H of G' such that G'/H is a Lie group. Then $H \leq \ker(\pi|_{G'})$ since $\pi|_{G'}(H)$ is a compact subgroup of $(\mathbb{R}^{>0}, \times)$. Let $\phi : G' \to G'/H$ be the quotient map. Hence the homomorphisms induce a continuous group homomorphism ψ from G'/H to $(\mathbb{R}^{>0}, \times)$.

$$\begin{tikzcd} G' \arrow[r, "\phi"] \arrow[d, "i"] \arrow[dr, "\pi|_{G'}"] & G'/H \arrow[d, dashed, "\psi"] \\ G \arrow[r, "\pi"] & (\mathbb{R}^{>0}, \times) \end{tikzcd}$$

Note that the above diagram commutes. By the proven special case for Lie groups, we then either have the exact sequence

$$1 \to \ker \psi \to G'/H \to (\mathbb{R}^{>0}, \times) \to 1$$

or $\ker\psi$ is open in G'/H. In the former case, $\pi|_{G'}$ is open as a composition of open maps. By Lemma 7.30, we conclude that π is a quotient map in this case. In the latter case, $\ker(\pi|_{G'})$ is open in G'. Thus, here we have $\ker\pi$ is open in G because $\ker\pi$ is a union of translations of $\ker(\pi|_{G'})$. □

The modular function $\Delta_G : G \to (\mathbb{R}^{>0}, \times)$ is a continuous group homomorphism by Fact 5.6.2, but generally not a quotient map. It is easy to construct examples where $G/(\ker\Delta_G)$ is discrete. The above proposition claims that these are the only two possibilities, which will be used in the later proofs.

7.7.2 Proofs of the main theorems

In this subsection, we prove Theorems 7.1 and 7.2. For the reader's convenience, Proposition 7.32 gathers together all the induction steps we can do using the earlier results with the exception of Proposition 7.28, which will be used in the proof of Theorem 7.1 directly.

Proposition 7.32. *Let G be a locally compact group with noncompact Lie dimension n and helix dimension h. Let $\Delta_G : G \to (\mathbb{R}^{>0}, \times)$ be the modular function of G. Then G satisfies* $\mathrm{BM}(n-h)$ *if one of the following assumptions holds:*

1. *The locally compact group $\ker\Delta_G$ has noncompact Lie dimension n' and helix dimension h', and $\ker\Delta_G$ satisfies* $\mathrm{BM}(n'-h')$.

2. *G is unimodular, G' is an open subgroup of G such that G' has noncompact Lie dimension n' and helix dimension h' and satisfies* $\mathrm{BM}(n'-h')$.

3. *G is unimodular, H is a compact normal subgroup of G, the quotient G/H has noncompact Lie dimension n' and helix dimension h' and satisfies* $\mathrm{BM}(n'-h')$.

4. *There is an exact sequence of connected semisimple Lie groups*

$$1 \to H \to G \to G/H \to 1$$

such that H has non compact Lie dimension n_1 and helix dimension h_1, and satisfies $\mathrm{BM}(n_1 - h_1)$*, and G/H has noncompact Lie dimension n_2 and helix dimension h_2, and satisfies* $\mathrm{BM}(n_2 - h_2)$*.*

5. *There is an exact sequence of connected unimodular Lie groups*

$$1 \to H \to G \to G/H \to 1$$

such that H has noncompact Lie dimension n_1 and helix dimension 0*, and satisfies* $\mathrm{BM}(n_1)$*, and G/H has noncompact Lie dimension n_2 and helix dimension h_2 with $h_2 = h$, and satisfies* $\mathrm{BM}(n_2 - h)$*.*

Proof. We first prove (1). Note that by Fact 5.6.1, $\ker \Delta_G$ is unimodular. By Lemma 7.31, we either have the exact sequence of locally compact groups

$$1 \to \ker \Delta_G \to G \to (\mathbb{R}^{>0}, \times) \to 1$$

or $\ker \Delta_G$ is open in G. In the former case, by Proposition 7.19, we have $n = n' + 1$ and $h = h'$. Hence, in this case G satisfies $\mathrm{BM}(n - h)$ by Proposition 7.26. In the latter case, by Corollary 7.14, $n = n'$ and $h = h'$. Here, we have G satisfies $\mathrm{BM}(n - h)$ by Proposition 7.29.

Next we prove (2). By Corollary 7.14, we have $n = n'$ and $h = h'$. The desired conclusion then follows from Proposition 7.24.

We now prove (3). By Corollary 7.15, we have $n = n'$ and $h = h'$. Also by Corollary 7.15, the compact group H has noncompact Lie dimension and helix dimension 0. Hence, using Proposition 7.23, we obtain the conclusion that we want.

We prove (4). By Proposition 7.17.1 and Proposition 7.17.2 respectively, we have $n = n_1 + n_2$ and $h = h_1 + h_2$. Recall that semisimple groups are unimodular. Using Proposition 7.23, we learn that G satisfies $\mathrm{BM}(n - h)$.

Finally, we prove (5). By Proposition 7.17.1, we have $n = n_1 + n_2$. Since the helix dimension of H is 0, and the helix dimension of G/H is h, by Proposition 7.23, G satisfies $\mathrm{BM}(n-h)$. $\square$

The following corollary says that when G is a Lie group, we can further reduce the problem to connected unimodular groups.

Corollary 7.33. *Let G be a Lie group with noncompact Lie dimension n and helix dimension h. Let $\Delta_G : G \to (\mathbb{R}^{>0}, \times)$ be the modular function of G. Let $G' = (\ker \Delta_G)_0$ be the identity component of $\ker \Delta_G$ with noncompact Lie dimension n' and helix dimension h'. Then G' is connected and unimodular, and if G' satisfies $\mathrm{BM}(n'-h')$, G satisfies $\mathrm{BM}(n-h)$.*

Proof. Note that $(\ker \Delta_G)_0$ is open in $\ker \Delta_G$ by Fact 5.12. The desired conclusion is then a consequence of Proposition 7.32.1 and Proposition 7.32.2. $\square$

Now we are able to prove the main inequality (7.2) for Lie groups. As mentioned earlier, the main strategy is induction on dimension.

Proof of Theorem 7.1. Consider first the case where G is a solvable Lie group. Using Corollary 7.33, we can also assume that G is connected and unimodular. Recall that d is the topological dimension of G. The case when $d = 0$ or 1 is trivial, as every group satisfies $\mathrm{BM}(0)$, and the one dimensional solvable Lie group is either $\mathbb{T}$ or $\mathbb{R}$ by Fact 5.15.1. If G is abelian, then it is isomorphic to $\mathbb{T}^m \times \mathbb{R}^{d-m}$. We get a desired conclusion applying Proposition 7.32.5 repeatedly. Otherwise, from the solvability of G we get the exact sequence

$$1 \to [G, G] \to G \to G/[G, G] \to 1$$

with both $[G, G]$ and $G/[G, G]$ connected, solvable and having smaller dimensions than G. Note that $G/[G, G]$ is abelian, and hence unimodular. Applying Proposition 7.32.5, and the statement for of the theorem for abelian Lie groups, we get desired conclusion for this case.

Consider next the case where G is connected and semisimple. We may further assume that G is a connected simple Lie group, otherwise by Fact 5.23, we can always find a connected group $H \triangleleft G$ such that both H and G/H are connected semisimple Lie groups with lower dimension; by Proposition 7.32.4, the Brunn–Minkowski inequality on G can be obtained from the Brunn–Minkowski inequalities on H and G/H. Now we write $G = KAN$ as in Fact 5.31. We first consider the case when G has a finite center, and then K is compact. Let n be the noncompact Lie dimension of G. Hence, n is the dimension of the solvable Lie group $Q = AN$. Note that A and N are simply connected by Fact 5.31. Hence their noncompact Lie dimensions are the same as their dimensions by Fact 5.15.2. By Proposition 7.17.1 and Fact 5.31, the noncompact Lie dimension of Q is n, and hence Q satisfies $\mathrm{BM}(n)$ from the solvable Lie case. We obtain the desired conclusion for G by applying Proposition 7.28.

Suppose the connected simple Lie group G has a center of rank $h \geq 1$. Apply Proposition 7.28 again, and we obtain an inequality (7.2) for G with exponent $\dim(AN)$. By Proposition 7.8, we have $\dim(AN) = n - h$. The desired conclusion for the connected semisimple Lie groups follows similarly from Fact 5.23 and Proposition 7.32.4.

Finally, we show the statement for an arbitrary Lie group G. Using Corollary 7.33 again, we can assume that G is connected and unimodular. Then by Fact 5.22 we obtain an exact sequence

$$1 \to Q \to G \to S \to 1,$$

where Q is a connected unimodular solvable group and S is a connected semisimple Lie group. We then apply Proposition 7.32.5 and the earlier two cases to get the desired conclusion. □

Finally, we prove the inequality (7.2) for all locally compact groups.

Proof of Theorem 7.2. By Proposition 7.32.1 we can assume that G is unimodular. By the Gleason–Yamabe Theorem (Fact 5.10.1), G has an almost-Lie open subgroup. Now using Proposition 7.32.2, we can further assume that G is a unimodular almost-Lie group. Then we can choose a compact subgroup K of G such that G/K is a unimodular Lie group. The desired conclusion then follows from Theorem 7.1 and Proposition 7.32.3. □

We briefly discuss Theorem 7.5, which is a consequence of the proof of Theorem 7.2.

Proof of Theorem 7.5. Repeating the arguments in the proofs of Proposition 7.32, Corollary 7.33, Theorem 7.1, Theorem 7.2, and Fact 5.23 while ignoring the helix dimension, it suffices to show the theorem when G is a simple Lie group.

From the hypothesis, we already have the desired conclusion under the further assumption that our simple Lie group G is also simply connected. We now consider the general case. If G has finite center, the result is a special case of Theorem 7.1. So suppose the center $Z(G)$ of G is infinite. Let $\widetilde{G}$ be the universal cover of G, $Z(\widetilde{G})$ its center, and $\rho : \widetilde{G} \to G$ the covering map. Then $\ker \rho$ is a subgroup of $Z(\widetilde{G})$ by Fact 5.27. Using Fact 5.29, the center $Z(\widetilde{G})$ have rank at most 1. By the earlier assumption, the center $Z(G)$ also has rank at least 1. Hence, by Fact 5.27, both $Z(\widetilde{G})$ and $Z(G)$ must have rank 1, and $\ker \rho$ is finite. Therefore, the desired conclusion for G can be reduced to that of $\widetilde{G}$ by taking the inverse image under ρ, which we already know from the hypothesis. $\square$

References

[1] Noga Alon, József Balogh, Peter Keevash, and Benny Sudakov, *The number of edge colorings with no monochromatic cliques*, J. London Math. Soc. (2) **70** (2004), no. 2, 273–288.

[2] Noga Alon and Gregory A. Freiman, *On sums of subsets of a set of integers*, Combinatorica **8** (1988), no. 4, 297–306.

[3] Noga Alon and Daniel Kleitman, *Sum-free subsets*, A tribute to Paul Erdős, Cambridge Univ. Press, Cambridge, 1990, pp. 13–26.

[4] Béla Bajnok, *Additive combinatorics*, Discrete Mathematics and its Applications (Boca Raton), CRC Press, Boca Raton, FL, 2018, A menu of research problems.

[5] Béla Bajnok and Ryan Matzke, *The maximum size of (k, l)-sum free sets in cyclic groups*, Bull. Aust. Math. Soc. **99** (2019), no. 2, 184–194.

[6] József Balogh, *A remark on the number of edge colorings of graphs*, European J. Combin. **27** (2006), no. 4, 565–573.

[7] József Balogh and Lina Li, *The typical structure of Gallai colorings and their extremal graphs*, SIAM J. Discrete Math. **33** (2019), no. 4, 2416–2443.

[8] József Balogh and József Solymosi, *On the number of points in general position in the plane*, Discrete Anal. (2018), Paper No. 16, 20.

[9] József Balogh and Adam Zsolt Wagner, *Further applications of the container method*, Recent trends in combinatorics, IMA Vol. Math. Appl., vol. 159, Springer, 2016, pp. 191–213.

[10] Elías Baro, Eric Jaligot, and Margarita Otero, *Commutators in groups definable in o-minimal structures*, Proc. Amer. Math. Soc. **140** (2012), no. 10, 3629–3643.

[11] Felix A. Behrend, *On sets of integers which contain no three terms in arithmetical progression*, Proc. Nat. Acad. Sci. U.S.A. **32** (1946), 331–332.

[12] Fabrício S. Benevides, Carlos Hoppen, and Rudini M. Sampaio, *Edge-colorings of graphs avoiding complete graphs with a prescribed coloring*, Discrete Math. **340** (2017), no. 9, 2143–2160.

[13] Yves Benoist and Nicolas de Saxcé, *A spectral gap theorem in simple Lie groups*, Invent. Math. **205** (2016), no. 2, 337–361.

[14] Vitaly Bergelson and Alexander Leibman, *Polynomial extensions of van der Waerden's and Szemerédi's theorems*, J. Amer. Math. Soc. **9** (1996), no. 3, 725–753.

[15] Vitaly Bergelson and Terence Tao, *Multiple recurrence in quasirandom groups*, Geom. Funct. Anal. **24** (2014), no. 1, 1–48.

[16] Yuri Bilu, *The $(\alpha + 2\beta)$-inequality on a torus*, J. London Math. Soc. (2) **57** (1998), no. 3, 513–528.

[17] Michael Björklund, *Small product sets in compact groups*, Fund. Math. **238** (2017), no. 1, 1–27.

[18] Michael Björklund and Alexander Fish, *Approximate invariance for ergodic actions of amenable groups*, Discrete Anal. (2019), Paper No. 6, 56.

[19] Michael Björklund and Tobias Hartnick, *Approximate lattices*, Duke Math. J. **167** (2018), no. 15, 2903–2964.

[20] Thomas F. Bloom, *A quantitative improvement for Roth's theorem on arithmetic progressions*, J. Lond. Math. Soc. (2) **93** (2016), no. 3, 643–663.

[21] Thomas F. Bloom and Olof Sisask, *Logarithmic bounds for Roth's theorem via almost-periodicity*, Discrete Anal. (2019), Paper No. 4, 20.

[22] ———, *Breaking the logarithmic barrier in Roth's theorem on arithmetic progressions*, arXiv:2007.03528 (2020).

[23] Vladimir I. Bogachev, *Measure theory. Vol. II*, Springer-Verlag, Berlin, 2007.

[24] Nicolas Bourbaki, *Lie groups and Lie algebras. Chapters 1–3*, Elements of Mathematics (Berlin), Springer-Verlag, Berlin, 1989, Translated from the French, Reprint of the 1975 edition.

[25] ———, *General topology. Chapters 1–4*, Elements of Mathematics (Berlin), Springer-Verlag, Berlin, 1998, Translated from the French, Reprint of the 1989 English translation.

[26] ———, *Integration. II. Chapters 7–9*, Elements of Mathematics (Berlin), Springer-Verlag, Berlin, 2004, Translated from the 1963 and 1969 French originals by Sterling K. Berberian.

[27] Jean Bourgain, *On the behavior of the constant in the Littlewood-Paley inequality*, Geometric aspects of functional analysis (1987–88), Lecture Notes in Math., vol. 1376, Springer, Berlin, 1989, pp. 202–208.

[28] ———, *Estimates related to sumfree subsets of sets of integers*, Israel J. Math. **97** (1997), 71–92.

[29] ———, *On triples in arithmetic progression*, Geom. Funct. Anal. **9** (1999), no. 5, 968–984.

[30] Jean Bourgain and Alex Gamburd, *On the spectral gap for finitely-generated subgroups of* SU(2), Invent. Math. **171** (2008), no. 1, 83–121.

[31] ———, *A spectral gap theorem in* SU(d), J. Eur. Math. Soc. (JEMS) **14** (2012), no. 5, 1455–1511.

[32] Jean Bourgain, Alex Gamburd, and Peter Sarnak, *Affine linear sieve, expanders, and sum-product*, Invent. Math. **179** (2010), no. 3, 559–644.

[33] Jean Bourgain and Péter P. Varjú, *Expansion in $SL_d(\mathbf{Z}/q\mathbf{Z})$, q arbitrary*, Invent. Math. **188** (2012), no. 1, 151–173.

[34] Rémi Boutonnet, Adrian Ioana, and Alireza Salehi Golsefidy, *Local spectral gap in simple Lie groups and applications*, Invent. Math. **208** (2017), no. 3, 715–802.

[35] Emmanuel Breuillard, *Geometry of locally compact groups of polynomial growth and shape of large balls*, Groups Geom. Dyn. **8** (2014), no. 3, 669–732.

[36] Emmanuel Breuillard, Ben Green, and Terence Tao, *The structure of approximate groups*, Publ. Math. Inst. Hautes Études Sci. **116** (2012), 115–221.

[37] Emmanuel Breuillard and Alexander Lubotzky, *Expansion in simple groups*, arXiv:1807.03879 (2018).

[38] Hermann Brunn, *Über Ovale und Eiflächen*, Inaugural Dissertation, München: Akademische Buchdruckerei von F. Straub. (1887).

[39] Pablo Candela and Anne De Roton, *On sets with small sumset in the circle*, Q. J. Math. **70** (2019), no. 1, 49–69.

[40] Pietro Kreitlon Carolino, *The Structure of Locally Compact Approximate Groups*, ProQuest LLC, Ann Arbor, MI, 2015, Thesis (Ph.D.)–University of California, Los Angeles.

[41] Yangyang Cheng, Yifan Jing, Lina Li, Guanghui Wang, and Wenling Zhou, *Integer colorings with forbidden rainbow sums*, arXiv:2005.14384, submitted (2020).

[42] Michael Christ, *Near equality in the Brunn-Minkowski inequality*, arXiv:1207.5062 (2012).

[43] Michael Christ and Marina Iliopoulou, *Inequalities of Riesz-Sobolev type for compact connected abelian groups*, arXiv:1808.08368, to appear at Amer. J. Math. (2021).

[44] Dennis Clemens, Shagnik Das, and Tuan Tran, *Colourings without monochromatic disjoint pairs*, European J. Combin. **70** (2018), 99–124.

[45] Gabriel Conant and Anand Pillay, *Approximate subgroups with bounded VC-dimension*, arXiv:2004.05666 (2020).

[46] Gabriel Conant, Anand Pillay, and Caroline Terry, *Structure and regularity for subsets of groups with finite VC-dimension*, to appear at J. Eur. Math. Soc. (JEMS) (2021).

[47] Ernie Croot and Olof Sisask, *A probabilistic technique for finding almost-periods of convolutions*, Geom. Funct. Anal. **20** (2010), no. 6, 1367–1396.

[48] Shagnik Das, Roman Glebov, Benny Sudakov, and Tuan Tran, *Colouring set families without monochromatic k-chains*, J. Combin. Theory Ser. A **168** (2019), 84–119.

[49] Nicolas de Saxcé, *A product theorem in simple Lie groups*, Geom. Funct. Anal. **25** (2015), no. 3, 915–941.

[50] Anton Deitmar and Siegfried Echterhoff, *Principles of harmonic analysis*, Universitext, Springer, New York, 2009.

[51] Jean-Marc Deshouillers and Gregory A. Freiman, *A step beyond Kneser's theorem for abelian finite groups*, Proc. London Math. Soc. (3) **86** (2003), no. 1, 1–28.

[52] Jean-Marc Deshouillers, Gregory A. Freiman, Vera Sós, and Mikhail Temkin, *On the structure of sum-free sets. II*, no. 258, 1999, Structure theory of set addition, pp. xii, 149–161.

[53] Matt DeVos, *The structure of critical product sets*, arXiv:1301.0096 (2013).

[54] Mauro Di Nasso, Isaac Goldbring, and Martino Lupini, *Nonstandard methods in Ramsey theory and combinatorial number theory*, Lecture Notes in Mathematics, vol. 2239, Springer, Cham, 2019.

[55] Palahenedi Hewage Dianada and Hian Poh Yap, *Maximal sum-free sets of elements of finite groups*, Proc. Japan Acad. **45** (1969), 1–5.

[56] Sean Eberhard, *Følner sequences and sum-free sets*, Bull. Lond. Math. Soc. **47** (2015), no. 1, 21–28.

[57] Sean Eberhard, Ben Green, and Freddie Manners, *Sets of integers with no large sum-free subset*, Ann. of Math. (2) **180** (2014), no. 2, 621–652.

[58] György Elekes and Lajos Rónyai, *A combinatorial problem on polynomials and rational functions*, J. Combin. Theory Ser. A **89** (2000), no. 1, 1–20.

[59] Paul Erdős, *Extremal problems in number theory*, Proc. Sympos. Pure Math., Vol. VIII, Amer. Math. Soc., Providence, R.I., 1965, pp. 181–189.

[60] ———, *Some new applications of probability methods to combinatorial analysis and graph theory*, Proceedings of the Fifth Southeastern Conference on Combinatorics, Graph Theory and Computing (Florida Atlantic Univ., Boca Raton, Fla., 1974), 1974, pp. 39–51. Congressus Numerantium, No. X.

[61] Paul Erdős and Gregory Freiman, *On two additive problems*, J. Number Theory **34** (1990), no. 1, 1–12.

[62] Paul Erdős and Endre Szemerédi, *On sums and products of integers*, Studies in pure mathematics, Birkhäuser, Basel, 1983, pp. 213–218.

[63] Paul Erdős and Paul Turán, *On some sequences of integers*, J. London Math. Soc. **11** (1936), no. 4, 261–264.

[64] Victor Falgas-Ravry, Kelly O'Connell, and Andrew Uzzell, *Multicolor containers, extremal entropy, and counting*, Random Structures & Algorithms **54** (2019), no. 4, 676–720.

[65] Alessio Figalli and David Jerison, *Quantitative stability for sumsets in* $\mathbb{R}^n$, J. Eur. Math. Soc. (JEMS) **17** (2015), no. 5, 1079–1106.

[66] ———, *Quantitative stability for the Brunn-Minkowski inequality*, Adv. Math. **314** (2017), 1–47.

[67] Gerald B. Folland, *A course in abstract harmonic analysis*, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1995.

[68] Gregory A. Freĭman, *The addition of finite sets. I*, Izv. Vysš. Učebn. Zaved. Matematika **1959** (1959), no. 6 (13), 202–213.

[69] ———, *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.

[70] Hillel Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256.

[71] Hillel Furstenberg and Yitzhak Katznelson, *A density version of the Hales-Jewett theorem*, J. Anal. Math. **57** (1991), 64–119.

[72] Richard J. Gardner, *The Brunn-Minkowski inequality*, Bull. Amer. Math. Soc. (N.S.) **39** (2002), no. 3, 355–405.

[73] Andrew M. Gleason, *Groups without small subgroups*, Ann. of Math. (2) **56** (1952), 193–212.

[74] William Timothy Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), no. 3, 465–588.

[75] ———, *Hypergraph regularity and the multidimensional Szemerédi theorem*, Ann. of Math. (2) **166** (2007), no. 3, 897–946.

[76] Ben Green, *A Szemerédi-type regularity lemma in abelian groups, with applications*, Geom. Funct. Anal. **15** (2005), no. 2, 340–376.

[77] Ben Green and Sergei Konyagin, *On the Littlewood problem modulo a prime*, Canad. J. Math. **61** (2009), no. 1, 141–164.

[78] Ben Green and Imre Z. Ruzsa, *Sum-free sets in abelian groups*, Israel J. Math. **147** (2005), 157–188.

[79] ———, *Sets with small sumset and rectification*, Bull. London Math. Soc. **38** (2006), no. 1, 43–52.

[80] ———, *Freiman's theorem in an arbitrary abelian group*, J. Lond. Math. Soc. (2) **75** (2007), no. 1, 163–175.

[81] John T. Griesmer, *An inverse theorem: when the measure of the sumset is the sum of the measures in a locally compact abelian group*, Trans. Amer. Math. Soc. **366** (2014), no. 4, 1797–1827.

[82] ———, *Semicontinuity of structure for small sumsets in compact abelian groups*, Discrete Anal. (2019), Paper No. 18, 46.

[83] Mikhael Gromov, *Groups of polynomial growth and expanding maps*, Publ. Math. Inst. Hautes Études Sci. (1981), no. 53, 53–73.

[84] ———, *Isoperimetry of waists and concentration of maps*, Geom. Funct. Anal. **13** (2003), no. 1, 178–215.

[85] Karsten Grove, Hermann Karcher, and Ernst A. Ruh, *Group actions and curvature*, Invent. Math. **23** (1974), 31–48.

[86] David S. Gunderson and Vojtěch Rödl, *Extremal problems for affine cubes of integers*, Combin. Probab. Comput. **7** (1998), no. 1, 65–79.

[87] Hiep Hàn and Andrea Jiménez, *Maximum number of sum-free colorings in finite abelian groups*, Israel J. Math. **226** (2018), no. 2, 505–534.

[88] Godfrey Harold Hardy and John Edensor Littlewood, *A new proof of a theorem on rearrangements*, J. London Math. Soc. **23** (1948), 163–168.

[89] Harald Helfgott, *Growth and generation in* $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$, Ann. of Math. (2) **167** (2008), no. 2, 601–623.

[90] Ralph Henstock and Murray Macbeath, *On the measure of sum-sets. I. The theorems of Brunn, Minkowski, and Lusternik*, Proc. London Math. Soc. (3) **3** (1953), 182–194.

[91] Joachim Hilgert and Karl-Hermann Neeb, *Structure and geometry of Lie groups*, Springer Monographs in Mathematics, Springer, New York, 2012.

[92] Carlos Hoppen, Yoshiharu Kohayakawa, and Hanno Lefmann, *Edge colourings of graphs avoiding monochromatic matchings of a given size*, Combin. Probab. Comput. **21** (2012), no. 1-2, 203–218.

[93] ———, *Hypergraphs with many Kneser colorings*, European J. Combin. **33** (2012), no. 5, 816–843.

[94] ———, *Edge-colorings of graphs avoiding fixed monochromatic subgraphs with linear Turán number*, European J. Combin. **35** (2014), 354–373.

[95] ———, *Edge-colorings of uniform hypergraphs avoiding monochromatic matchings*, Discrete Math. **338** (2015), no. 2, 262–271.

[96] Carlos Hoppen and Hanno Lefmann, *Edge-colorings avoiding a fixed matching with a prescribed color pattern*, European J. Combin. **47** (2015), 75–94.

[97] Carlos Hoppen, Hanno Lefmann, and Knut Odermann, *A coloring problem for intersecting vector spaces*, Discrete Math. **339** (2016), no. 12, 2941–2954.

[98] ———, *On graphs with a large number of edge-colorings avoiding a rainbow triangle*, European J. Combin. **66** (2017), 168–190.

[99] ———, *A rainbow Erdős-Rothschild problem*, SIAM J. Discrete Math. **31** (2017), no. 4, 2647–2674.

[100] Bernard Host, *A short proof of a conjecture of Erdős proved by Moreira, Richter and Robertson*, Discrete Anal. (2019), Paper No. 19, 10.

[101] Ehud Hrushovski, *personal communication*.

[102] ———, *Stable group theory and approximate subgroups*, J. Amer. Math. Soc. **25** (2012), no. 1, 189–243.

[103] ———, *Beyond the lascar group*, arXiv:2011.12009 (2020).

[104] Yifan Jing, Souktik Roy, and Chieu-Minh Tran, *Semialgebraic methods and generalized sum-product phenomena*, arXiv:1910.04904, submitted (2019).

[105] Yifan Jing and Chieu-Minh Tran, *Minimal and nearly minimal measure expansions in connected unimodular groups*, arXiv:2006.01824, submitted (2021).

[106] Yifan Jing, Chieu-Minh Tran, and Ruixiang Zhang, *A nonabelian Brunn-Minkowski inequality*, arXiv:2101.07782, submitted (2021).

[107] Yifan Jing and Shukun Wu, *A note on the largest sum-free sets of integers*, arXiv:2011.09963, submitted (2020).

[108] ———, *The largest (k, ℓ)-sum-free subsets*, Trans. Amer. Math. Soc. **374** (2021), no. 7, 5163–5189.

[109] David Kazhdan, *On ε-representations*, Israel J. Math. **43** (1982), no. 4, 315–323.

[110] Alexander S. Kechris, *Classical descriptive set theory*, Graduate Texts in Mathematics, vol. 156, Springer-Verlag, New York, 1995.

[111] Johannes Kemperman, *On small sumsets in an abelian group*, Acta Math. **103** (1960), 63–88.

[112] ———, *On products of sets in a locally compact group*, Fund. Math. **56** (1964), 51–68.

[113] Anthony W. Knapp, *Representation theory of semisimple groups*, Princeton Mathematical Series, vol. 36, Princeton University Press, Princeton, NJ, 1986, An overview based on examples.

[114] ———, *Lie groups beyond an introduction*, second ed., Progress in Mathematics, vol. 140, Birkhäuser Boston, Inc., Boston, MA, 2002.

[115] Martin Kneser, *Summenmengen in lokalkompakten abelschen Gruppen*, Math. Z. **66** (1956), 88–110.

[116] Sergei Konyagin, *On the Littlewood problem*, Izv. Akad. Nauk SSSR Ser. Mat. **45** (1981), no. 2, 243–265, 463.

[117] Daniel Král, Oriol Serra, and Lluís Vena, *A combinatorial proof of the removal lemma for groups*, J. Combin. Theory Ser. A **116** (2009), no. 4, 971–978.

[118] Noah Kravitz, *The largest (k, ℓ)-sum-free sets in compact abelian groups*, Integers **20** (2020), Paper No. A57, 8.

[119] Hanno Lefmann and Yury Person, *Exact results on the number of restricted edge colorings for some families of linear hypergraphs*, J. Graph Theory **73** (2013), no. 1, 1–31.

[120] Hanno Lefmann, Yury Person, Vojtěch Rödl, and Mathias Schacht, *On colourings of hypergraphs without monochromatic Fano planes*, Combin. Probab. Comput. **18** (2009), no. 5, 803–818.

[121] Hanno Lefmann, Yury Person, and Mathias Schacht, *A structural result for hypergraphs with many restricted edge colorings*, J. Comb. **1** (2010), no. 3-4, 441–475.

[122] Gian Paolo Leonardi and Simon Masnou, *On the isoperimetric problem in the Heisenberg group $\mathbb{H}^n$*, Ann. Mat. Pura Appl. (4) **184** (2005), no. 4, 533–553.

[123] Vsevolod Lev, *Small doubling in cyclic groups*, arXiv:2010.03410 (2020).

[124] Hong Liu, Maryam Sharifzadeh, and Katherine Staden, *On the maximum number of integer colourings with forbidden monochromatic sums*, Electron. J. Combin. **28** (2021), no. 1, Research Paper 59.

[125] Peter A. Loeb, *Conversion from nonstandard to standard measure spaces and applications in probability theory*, Trans. Amer. Math. Soc. **211** (1975), 113–122.

[126] Jason Long and Adam Zsolt Wagner, *The largest projective cube-free subsets of $\mathbb{Z}_{2^n}$*, European J. Combin. **81** (2019), 156–171.

[127] Tomasz Łuczak and Tomasz Schoen, *On infinite sum-free sets of natural numbers*, J. Number Theory **66** (1997), no. 2, 211–224.

[128] Lazar' A. Lyusternik, *Die Brunn–Minkowskische ungleichnung für beliebige messbare mengen*, Comptes Rendus de l'Académie des Sciences de l'URSS. Nouvelle Série. III (1935), 55–58.

[129] Alexander M. Macbeath, *On measure of sum sets. II. The sum-theorem for the torus*, Proc. Cambridge Philos. Soc. **49** (1953), 40–43.

[130] ______, *On the measure of product sets in a topological group*, J. London Math. Soc. **35** (1960), 403–407.

[131] Jean-Cyrille Massicot and Frank O. Wagner, *Approximate subgroups*, J. Éc. polytech. Math. **2** (2015), 55–64.

[132] Michael McCrudden, *On the Brunn-Minkowski coefficient of a locally compact unimodular group*, Proc. Cambridge Philos. Soc. **65** (1969), 33–45.

[133] ______, *On the Brunn-Minkowski coefficient of a locally compact unimodular group*, Proc. Cambridge Philos. Soc. **65** (1969), 33–45.

[134] ______, *On critical pairs of product sets in a certain matrix group*, Proc. Cambridge Philos. Soc. **67** (1970), 569–581.

[135] O. Carruth McGehee, Louis Pigno, and Brent Smith, *Hardy's inequality and the L^1-norm of exponential sums*, Ann. of Math. (2) **113** (1981), no. 3, 613–618.

[136] Hermann Minkowski, *Geometrie der Zahlen*, Leipzig: Teubner. (1896).

[137] Deane Montgomery and Leo Zippin, *Small subgroups of finite-dimensional groups*, Ann. of Math. (2) **56** (1952), 213–241.

[138] Roberto Monti, *Brunn-Minkowski and isoperimetric inequality in the Heisenberg group*, Ann. Acad. Sci. Fenn. Math. **28** (2003), no. 1, 99–109.

[139] Joel Moreira, Florian K. Richter, and Donald Robertson, *A proof of a sumset conjecture of Erdős*, Ann. of Math. (2) **189** (2019), no. 2, 605–652.

[140] Brendan Murphy and Jonathan Pakianathan, *Kakeya configurations in Lie groups and homogeneous spaces*, Topology Appl. **180** (2015), 1–15.

[141] Brendan Nagle, Vojtěch Rödl, and Mathias Schacht, *The counting lemma for regular k-uniform hypergraphs*, Random Structures & Algorithms **28** (2006), no. 2, 113–179.

[142] Oleg Pikhurko and Zelealem B. Yilma, *The maximum number of K_3-free and K_4-free edge 4-colorings*, J. Lond. Math. Soc. (2) **85** (2012), no. 3, 593–615.

[143] László Pyber and Endre Szabó, *Growth in finite simple groups of Lie type*, J. Amer. Math. Soc. **29** (2016), no. 1, 95–146.

[144] Dmitrii Raikov, *On the addition of point-sets in the sense of Schnirelmann*, Rec. Math. [Mat. Sbornik] N.S. **5(47)** (1939), 425–440.

[145] Frank Plumpton Ramsey, *On a problem of formal logic*, Proc. London Math. Soc. (2) **30** (1929), no. 4, 264–286.

[146] Vojtěch Rödl and Jozef Skokan, *Regularity lemma for k-uniform hypergraphs*, Random Structures & Algorithms **25** (2004), no. 1, 1–42.

[147] Christian Rosendal, *Continuity of universally measurable homomorphisms*, Forum Math. Pi **7** (2019), e5, 20.

[148] Klaus Friedrich Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 104–109.

[149] Imre Z. Ruzsa and Endre Szemerédi, *Triple systems with no six points carrying three triangles*, Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II, Colloq. Math. Soc. János Bolyai, vol. 18, North-Holland, Amsterdam-New York, 1978, pp. 939–945.

[150] Tom Sanders, *The Littlewood-Gowers problem*, J. Anal. Math. **101** (2007), 123–162.

[151] ______, *On a nonabelian Balog-Szemerédi-type lemma*, J. Aust. Math. Soc. **89** (2010), no. 1, 127–132.

[152] ______, *On Roth's theorem on progressions*, Ann. of Math. (2) **174** (2011), no. 1, 619–636.

[153] David Saxton and Andrew Thomason, *Hypergraph containers*, Invent. Math. **201** (2015), no. 3, 925–992.

[154] Tomasz Schoen, *On the Littlewood conjecture in $\mathbb{Z}/p\mathbb{Z}$*, Mosc. J. Comb. Number Theory **7** (2017), no. 3, 66–72.

[155] ______, *Improved bound in Roth's theorem on arithmetic progressions*, arXiv:2005.01145 (2020).

[156] A. I. Sirota, *Centers of non-compact simple Lie groups*, Soviet Math. Dokl. **1** (1960), 1021–1024.

[157] Katherine Staden, *A note on the number of additive triples in subsets of integers*, preprint (2017).

[158] Endre Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245.

[159] Terence Tao, *Product set estimates for non-commutative groups*, Combinatorica **28** (2008), no. 5, 547–594.

[160] ———, *The Brunn-Minkowski inequality for nilpotent groups*, available at `https://terrytao.wordpress.com/2011/09/16/` (2011).

[161] ———, *Expansion in finite simple groups of Lie type*, Graduate Studies in Mathematics, vol. 164, American Mathematical Society, Providence, RI, 2015.

[162] ———, *An inverse theorem for an inequality of Kneser*, Proc. Steklov Inst. Math. **303** (2018), no. 1, 193–219, Published in Russian in Tr. Mat. Inst. Steklova **3**03 (2018), 209–238.

[163] Terence Tao and Van Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, vol. 105, Cambridge University Press, Cambridge, 2010.

[164] ———, *Sum-free sets in groups: a survey*, J. Comb. **8** (2017), no. 3, 541–552.

[165] Bartel Leendert van der Waerden, *Beweis einer Baudetschen Vermutung*, Nieuw. Arch. Wisk. **15** (1927), 212–216.

[166] A. G. Vosper, *The critical pairs of subsets of a group of prime order*, J. London Math. Soc. **31** (1956), 200–205.

[167] Ján Špakula and Pavol Zlatoš, *Almost homomorphisms of compact groups*, Illinois J. Math. **48** (2004), no. 4, 1183–1189.

[168] André Weil, *L'intégration dans les groupes topologiques et ses applications*, Actual. Sci. Ind., no. 869, Hermann et Cie., Paris, 1940.

[169] Hidehiko Yamabe, *A generalization of a theorem of Gleason*, Ann. of Math. (2) **58** (1953), 351–365.

[170] Raphael Yuster, *The number of edge colorings with no monochromatic triangle*, J. Graph Theory **21** (1996), no. 4, 441–452.

[171] Antoni Szczepan Zygmund, *Trigonometric series. Vol. I, II*, third ed., Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2002.