1
Introduction

1.1 Introduction

As we described in the preface, the theory of approximate groups can be thought of as describing those subsets of groups that are 'approximately closed'. We start by presenting a preliminary notion of approximate closure: *small doubling*. Given two sets A, B inside a group G we define their *product set* $AB = \{ab : a \in A, b \in B\}$. We also write A^{-1} for the set of inverses of elements of A, and write A^n and A^{-n} to denote the *iterated product sets* defined recursively by $A^0 = \{1\}$, $A^n = AA^{n-1}$ and $A^{-n} = (A^{-1})^n$. The study of product sets began in the setting of abelian groups, where one traditionally uses additive notation. Thus, if G is abelian we define the *sum set* $A + B = \{a + b : a \in A, b \in B\}$ and the *difference set* $A - B = \{a - b : a \in A, b \in B\}$. We also write $-A$ for the set of inverses of elements of A, and write nA and $-nA$ in place of A^n and A^{-n}, respectively.

To say that a finite set A is closed under taking products is then to say that $A^2 = A$. One way to define 'approximate' closure is to say that A^2 is not too much larger than A. To get a feel for what this might mean in practice, let us consider for a moment what might be thought of as 'extremal' or 'typical' for the size of A^2. It is not difficult to see what the extremal possibilities for $|A^2|$ are in terms of $|A|$: it is clear that $|A| \leq |A^2| \leq |A|^2$, and in general neither bound can be improved. Indeed, if A is a finite subgroup of G then $|A^2| = |A|$, while if G is the free group generated by A then $|A^2| = |A|^2$.

It turns out that the quadratic upper bound on the size of A^2 is in fact typical in some sense. For example, we show in Section 2.1 that if A is a set of size k chosen uniformly at random from an interval $\{1, \dots, n\} \subset \mathbb{Z}$ with n much larger than k then $\mathbb{E}[|A^2|]$ is close to $k^2/2$. This suggests

that a 'generic' set A should have $|A^2|$ comparable to $|A|^2$, and so it is sets for which

$$|A^2| = o(|A|^2) \tag{1.1.1}$$

that we should view as being 'exceptional'.

The theory of approximate groups is essentially concerned with the extreme case of (1.1.1) in which $|A^2|$ is *linear* in $|A|$, in the sense that

$$|A^2| \leq K|A| \tag{1.1.2}$$

for some fixed $K \geq 1$. Since (1.1.2) represents 'non-random' behaviour, we can expect such sets to exhibit a certain amount of 'structure'. One of the principal aims of approximate-group theory, and of this book, is to describe this structure in as much detail as possible.

Of course, one type of structure satisfying (1.1.2) is a finite subgroup, for which we may even take $K = 1$. Another trivial example is if A itself has size at most K. Let us reassure ourselves, though, that the theory of sets satisfying (1.1.2) is more general than just the theory of finite subgroups and 'small' sets. Indeed, it is easy to see that the set $A = \{-n, \dots, n\} \subset \mathbb{Z}$ satisfies $|A + A| \leq 2|A|$, and so the group $\mathbb{Z}$ contains arbitrarily large finite sets of small doubling, even though it contains no non-trivial subgroups. We will develop and generalise this example in Chapter 3.

Since it is the key property that we will be investigating, we now give a name to those sets satisfying (1.1.2).

Definition 1.1.1 (small doubling) Given a finite subset A of a group we call the quantity $|A|^2/|A|$ the *doubling constant* of A. If the doubling constant of A is at most a given constant K then we often say simply that A is a *set of doubling at most* K, or even merely a *set of small doubling*.

As we shall explain in some detail in Chapter 2, in some contexts, and particularly in the case of non-abelian groups, it is convenient for technical reasons to replace Definition 1.1.1 with a slightly stronger definition, due to Tao, which gives its name both to this book and to the theory.

Definition 1.1.2 (approximate group) A subset A of a group G is said to be a *K-approximate subgroup of G*, or simply a *K-approximate group*, if $A^{-1} = A$ and $1 \in A$, and if there exists $X \subset G$ with $|X| \leq K$ such that $A^2 \subset XA$.

Note in particular that a finite K-approximate group has doubling at most K. The conditions $A^{-1} = A$ and $1 \in A$ are largely for notational convenience. On the one hand, assuming that $A^{-1} = A$ avoids the need to distinguish between positive and negative iterated products, allowing us to replace an untidy-looking expression such as $A^2A^{-3}AA^{-1}A^3 \cup A^{-4}A^3A^{-1}$ with the more succinct A^{10}, for example. On the other hand, assuming that $1 \in A$ means that we have the nesting $A \subset A^2 \subset A^3 \subset \cdots$, which is also convenient at times. The existence of $X \subset G$ with $|X| \leq K$ such that $A^2 \subset XA$ is more serious, however. Indeed, we shall see in Chapter 2 that one can construct sets of bounded doubling that fail to be K-approximate groups for arbitrarily large K, so being a finite approximate group is strictly stronger than having small doubling. However, when introducing the definition of approximate groups Tao showed that the study of sets of small doubling nonetheless essentially reduces to the study of approximate groups in a certain precise way; in Theorem 2.5.6 we present a strengthening of this reduction that follows from work of Petridis.

One specific advantage of Definition 1.1.2 over Definition 1.1.1 that is worth emphasising at this point is that it applies without modification to infinite subsets of groups. Indeed, there has recently begun to emerge a theory of infinite approximate groups in certain particular contexts (see [6], for example). Nonetheless, the theory of finite approximate groups is far more developed than the theory of infinite approximate groups, and is the focus of this book.

In Chapter 2 we motivate and develop Definitions 1.1.1 and 1.1.2 in more detail, in particular deriving some of their elementary properties. In Chapter 3 we look in detail at some specific examples of sets of small doubling and approximate groups. In the largest part of the book, comprising Chapters 4–10, we prove a number of results describing the structure of approximate subgroups in various classes of group. Finally, in Chapter 11 we present some applications of approximate groups to geometric group theory.

1.2 Historical Discussion

In this section we very briefly present the historical context of the material of this book. We stress that this is designed to give the reader an overall feel for the development of the theory, rather than to be a comprehensive history.

Much of the early progress on classifying sets of small doubling focused on abelian groups. The theory was initiated in the 1960s by Freiman [26], who in particular gave an essentially complete classification of sets of small doubling in the integers. The theory was subsequently developed considerably by Ruzsa, who amongst other things gave a simpler proof of Freiman's theorem [55]. Ruzsa's work was brought to the attention of a wider audience when Gowers [30, 31] applied it in his celebrated proof of a theorem of Szemerédi [61] concerning arithmetic progressions in dense sets of integers. In the mid 2000s, Green and Ruzsa [35] generalised Freiman's theorem to arbitrary abelian groups; we present their result in Chapter 4.

Another important early result on abelian groups was the so-called *sum–product theorem* of Bourgain, Katz and Tao [10]. This roughly states that a subset of $\mathbb{F}_p$ cannot simultaneously have small additive doubling and small multiplicative doubling, unless it is either very small or already almost all of $\mathbb{F}_p$. One of the tools used in the proof was a result from Gowers's work on Szemerédi's theorem, refining work of Balog and Szemerédi and now often known as the *Balog–Szemerédi–Gowers theorem*. We introduce this briefly as Theorem 2.1.5. We discuss sum–product theorems further in Section 9.2.

At around the same time as Green and Ruzsa's generalisation of Freiman's theorem, efforts began in earnest to generalise these concepts and results to non-abelian groups. Some of the first work in this direction was by Helfgott [39], who showed that a generating subset of A of $SL_2(\mathbb{Z}/p\mathbb{Z})$ does not even satisfy the weaker version $|A| \leq c|A|^{1+\varepsilon}$ of (1.1.1), unless it is already close to the whole of $SL_2(\mathbb{Z}/p\mathbb{Z})$. Amongst the tools used by Helfgott were aspects of Ruzsa's theory, the Bourgain–Katz–Tao sum–product theorem, and the Balog–Szemerédi–Gowers theorem. Helfgott's result is of particular interest because of its use by Bourgain and Gamburd [9] to construct so-called *expander graphs*, one of the most celebrated applications of the theory.

The first systematic account of the elementary theory of sets of small doubling in non-abelian groups was Tao's foundational work [62]. This work also introduced the notion of approximate groups and proved their essential equivalence to small doubling (although as we note in Remark 2.4.8 the definition of approximate groups was to some extent anticipated by Green and Ruzsa). We present much of this material in Chapter 2.

After Tao's work there were a number of papers in fairly quick succession proving Freiman- or Helfgott-type results for various non-abelian

groups, such as soluble groups (Tao [65]), free groups (Razborov [50] and Safin [57]), torsion-free nilpotent groups (Breuillard–Green [12]), and various linear groups (Breuillard–Green [13, 14] and Gill–Helfgott [28]). We present some of these results in this book; for example, in Chapter 6 we generalise the result of [12] to arbitrary nilpotent groups, and in Chapter 9 we present the result of [13].

There has also been much subsequent work on generalising Helfgott's work and its applications to expansion, notably by Pyber and Szabó [49] and Breuillard, Green and Tao [16]. We describe this briefly in an appendix to Chapter 11, but Tao's book [66] already gives an excellent and comprehensive account of this work, so we refer the interested reader to that source for the details rather than repeating them here.

It turns out that many of the results discussed above are somewhat reminiscent of a phenomenon seen in the related context of *polynomial growth*. A subset A of a group exhibits *polynomial growth* if there exists a polynomial p such that $|A^n| \leq p(n)$ for all $n \in \mathbb{N}$. One slightly imprecise but intuitively useful way of comparing this to Definition 1.1.1 is that whilst Definition 1.1.1 says that A 'grows slowly' when it is multiplied by itself once, polynomial growth means that A 'grows slowly' when it is multiplied by itself any number of times. Moreover, a famous theorem of Gromov describing the structure of sets of polynomial growth exploits the easily checked fact that if A is such a set then there are infinitely many n for which A^n has small doubling. Gromov's theorem states that if A has polynomial growth then the group generated by A has a *nilpotent* subgroup of finite index (for readers unfamiliar with nilpotence, we give a detailed introduction in Chapter 5). As we will see in this book, many of the results listed above show that sets of small doubling in the groups under consideration also have a significant amount of nilpotent structure in some sense.

Helfgott and Lindenstrauss conjectured that these similarities between Gromov's theorem and results on sets of small doubling were not coincidental, and that in fact an arbitrary approximate subgroup should have a large amount of nilpotent structure in a precise sense. This was finally proved in 2011 by Breuillard, Green and Tao [18]. Their result, which we state in Chapter 7, essentially describes the structure of an arbitrary approximate group. It also leads to a refinement of Gromov's theorem, and in turn to various other applications to geometric group theory, some of which we describe in Chapter 11.

We end this historical note by emphasising that the history of approximate groups is still being written. In particular, the reader should not

interpret the existence of the Breuillard–Green–Tao result as meaning that the theory is complete. Indeed, whilst that result is very general, as we explain in Chapter 7 its conclusion is rather imprecise in a particular, quantitative sense. Indeed, even the optimal classification of sets of small doubling in abelian groups is not yet known, and, as we said in the preface, essentially all of the results of this type that we present in this book have room for improvement.

1.3 Bounds and Asymptotic Notation

The larger K is in Definitions 1.1.1 and 1.1.2, the weaker they become. We can therefore expect that the structure of a set satisfying Definition 1.1.1 or 1.1.2 that we are able to obtain should become 'rougher' as K increases. A big part of the results we present will be to quantify this increased 'roughness'. For example, in Theorem 2.2.1 we show that if A is a finite subset of a group satisfying Definition 1.1.1 with $K < \frac{3}{2}$ then there exists a subgroup H such that A lies in a coset of H and $|A|/|H| \geq 1/K$. Thus, A is a 'large' proportion of a coset of a subgroup, and the meaning of 'large' depends on K in a precise, quantified way.

At times, however, the precise expression we obtain in terms of K is less important than the overall form it takes. For example, if one result says that A is a subset of a certain structure H with $|A|/|H| \geq \exp(-15K^3+\log K)$, and another says the same thing but with $|A|/|H| \geq K^{-17}/100$, the fact that the first bound is exponential but the second is merely polynomial is far more important than the precise values of the constants or exponents in these expressions. In this specific setting, one might reasonably choose simply to say that there exist absolute constants $c, C > 0$ such that $|A|/|H| \geq cK^{-C}$ in the case of the first result or $|A|/|H| \geq \exp(-CK^C)$ in the case of the second (to say that a constant is *absolute* here means that it does not depend in any way on A or K). We therefore deploy the some standard shorthand notation to abbreviate bounds such as these in a way that emphasises the important 'shape' of the bound without the distraction of inconsequential constants and exponents, as follows.

We follow the standard convention that if X, Y are real, variable quantities then $X \ll Y$ and $Y \gg X$ each mean that there exists a constant $C > 0$ such that X is always at most CY. Thus, for example, one may write $10n^2 \ll n^3$ for $n \in \mathbb{N}$ because, for example, $10n^2 \leq 10n^3$ for every

$n \in \mathbb{N}$. We call C the constant *implicit in* or *implied by* the $\ll$ or $\gg$ notation.

The notation $O(Y)$ denotes a quantity that is at most a certain constant multiple of Y, while $\Omega(X)$ denotes a quantity that is at least a certain positive constant multiple of X. Thus, for example, we write $A \subset B^{O(1)}$ to mean that there exists a constant C and a number $m \leq C$ such that $A \subset B^m$, or say that a subgroup H is of index $O(m)$ in G to mean that there exists a constant C such that $[G : H] \leq Cm$. Technically the O and Ω notation could be used to replace the $\ll$ and $\gg$ notation, but we tend to opt for $\ll$ and $\gg$ where possible.

In the $\ll, \gg, O, \Omega$ notation, if the constant in question depends on some other variable z then we indicate this with a subscript, for example $X \ll_z Y$ or $O_z(Y)$.

The reader may find it a useful exercise to check that he or she has understood the above notation by verifying that

$$K^K \leq \exp(K^{O(1)})$$

for $K > 0$, a bound that we use frequently in the book without explicit mention.

Despite the importance of the bounds in many of the theorems we prove, in a number of cases where we have the option to simplify an argument at the expense of making the bounds worse we opt to do so, a trade-off one would usually not make in a research paper, but which suits the pedagogical aims of this book. Nonetheless, we always provide references to arguments giving the best bounds the author is aware of.

1.4 General Notation

We assume familiarity with the basic concepts, definitions and results from group theory that can be found in a book such as Hall [38] or Robinson [51]. In particular, we assume familiarity with the definition of a *free group* as given in [38, §7.1], for example.

Here is a list of specific notation and definitions that we use in this book.

- We write
$$\begin{aligned}\mathbb{N} &= \{1, 2, \dots\},\\ \mathbb{N}_0 &= \mathbb{N} \cup \{0\},\\ [n] &= \{1, \dots, n\},\\ [n]_0 &= \{0, \dots, n\},\\ [n]^{\pm} &= \{-n, \dots, n\}.\end{aligned}$$
- We write $\mathbb{C}^\times$ for the set of non-zero complex numbers. Given a prime p, we also write $(\mathbb{Z}/p\mathbb{Z})^\times$ for the set of non-zero elements of $\mathbb{Z}/p\mathbb{Z}$. In each case these sets form groups under the operation of multiplication.
- Given a subset A of an abelian group and $n \in \mathbb{N}$ we define the *dilate* $n \cdot A$ via $n \cdot A = \{na : a \in A\}$.
- Given a subset A of a set X, we write $1_A : X \to \{0, 1\}$ for the indicator function of A defined via
$$1_A(x) = \begin{cases} 1 \text{ if } x \in A \\ 0 \text{ if } x \notin A. \end{cases}$$
Given a function $f : X \to Y$ into some other set Y, we write $f|_A : A \to Y$ for the restriction of f to A.
- Given $x > 0$ and $c \in \mathbb{R}$ we write $\log^c x$ to mean $(\log x)^c$.
- We use expectation notation to write averages over finite sets. Specifically, given a finite set X and a function $f : X \to \mathbb{C}$ we define
$$\mathbb{E}_{x \in X} f(x) = \frac{1}{|X|} \sum_{x \in X} f(x).$$
- In general we write 1 for the identity element of any group. The main exception to this is that we normally write abelian groups additively, in which case we write 0 for the identity element. When we occasionally use alternative symbols we always state this explicitly.
- Given two sets A, B, we write $A \subset B$ to mean that A is a subset of B. *This allows the possibility that* $A = B$. Given groups G, H, we write $H < G$ to mean that H is a subgroup of G, and $H \lhd G$ to mean that H is a normal subgroup of G, *again in each case allowing the possibility that* $A = B$. To indicate that $A \subset B$ with no possibility of equality we write $A \subsetneq B$.

- We define the *rank* of a finitely generated group to be the size of the smallest or joint-smallest generating set.
- Given a group G and a subset $X \subset G$ we write $\langle X \rangle$ for the subgroup of G generated by X. If X is written with braces then we drop the braces when using the $\langle \cdot \rangle$ notation, for example writing $\langle x_1, \dots, x_r \rangle$ instead of $\langle \{x_1, \dots, x_r\} \rangle$.
- Given a group G with a subgroup $H < G$, we write H^G for the *normal closure* of H in G, that is the smallest normal subgroup of G containing H.
- We define the *commutator* $[x, y]$ of two elements in a group G via $[x, y] = x^{-1}y^{-1}xy$. We also indicate conjugation using exponents, defining $x^y = y^{-1}xy$, and more generally $Y^x = \{x^{-1}yx : y \in Y\}$ for a subset $Y \subset G$.
- Let G be a group. Given a subgroup $H < G$, we denote by $N_G(H)$ the normaliser of H in G; thus

$$N_G(H) = \{g \in G : H^g = H\}.$$

  Given a subset $X \subset G$, we denote by $C_G(X)$ the centraliser of X in G; thus

$$C_G(H) = \{g \in G : [g, x] = 1 \text{ for every } x \in X\}.$$

  Given, in addition, a normal subgroup $N \lhd G$, we write

$$C_{G/N}(X) = \{g \in G : [g, x] \subset N \text{ for every } x \in X\}.$$

1.5 Miscellaneous Results

Here are some standard results that are too general to belong in any particular chapter of this book, but useful to be able to refer to. Some proofs are left as exercises, and some are outsourced to standard texts.

Theorem 1.5.1 (fundamental theorem of finitely generated abelian groups [51, 4.2.10]) *Let G be a finitely generated abelian group. Then there exist $r \in \mathbb{N}_0$, primes $p_1, \dots, p_r$, and $m_0, \dots, m_r \in \mathbb{N}_0$ such that $G \cong \mathbb{Z}^{m_0} \oplus \mathbb{Z}/p_1^{m_1}\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/p_r^{m_r}\mathbb{Z}$.*

Recall that a subgroup C of a group G is *characteristic* if $\psi(C) = C$ for every $\psi \in \mathrm{Aut}(G)$.

Lemma 1.5.2 *Let $C \lhd N \lhd G$ be groups such that N is normal in G and C is characteristic in N. Then C is normal in G.*

Lemma 1.5.3 *Let $N, H \lhd G$ be normal subgroups of a group G. Then $C_{G/N}(H)$ is also normal in G.*

Given a finite set X and functions $f, g : X \to \mathbb{C}$, translating the Cauchy–Schwarz inequality into the expectation notation described in the preface gives

$$|\mathbb{E}_{x\in X} f(x)g(x)|^2 \leq (\mathbb{E}_{x\in X}|f(x)|^2)(\mathbb{E}_{x\in X}|g(x)|^2). \tag{1.5.1}$$

We also have $|\sum_{x\in X} f(x)|^2 = |\sum_{x\in X} 1_X(x)f(x)|^2$, and so the usual Cauchy–Schwarz inequality gives

$$\left|\sum_{x\in X} f(x)\right|^2 \leq |X| \sum_{x\in X} |f(x)|^2. \tag{1.5.2}$$

Theorem 1.5.4 (Fubini's theorem [5, Theorem 18.3]) *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a measurable function, and suppose that*

$$\int_{x\in\mathbb{R}^d} |F(x)|\,dx < \infty.$$

Then, viewing $\mathbb{R}^d$ as $\mathbb{R}^m \times \mathbb{R}^{d-m}$, we have

$$\begin{aligned}\int_{x\in\mathbb{R}^d} F(x)\,dx &= \int_{x_1\in\mathbb{R}^m}\int_{x_2\in\mathbb{R}^{d-m}} F(x_1,x_2)\,dx_2\,dx_1\\ &= \int_{x_2\in\mathbb{R}^{d-m}}\int_{x_1\in\mathbb{R}^m} F(x_1,x_2)\,dx_1\,dx_2.\end{aligned}$$