Chapter 1: Small Doubling

Adam Sheffer

March 31, 2016

A large portion of this chapter and a couple of smaller excerpts from following chapters are inspired by Lovett's survey [3].

1 Introduction

Let $\mathcal{G}$ be an abelian group with operation $+$, and let $A, B \subset \mathcal{G}$ be finite subsets. The *sum set* of A and B is defined as

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

Similarly, the *difference set* of A and B is defined as

$$A - B = \{a - b : a \in A \text{ and } b \in B\}.$$

To get some intuition, we begin with the group of real numbers $\mathbb{R}$ under addition. A trivial upper bound for the size of a sum set is $|A + A| \leq \frac{|A|^2+|A|}{2}$ (equality is obtained when every pair of elements of A have a distinct sum). If we build A by taking elements of $\mathbb{R}$ at random, then we expect $|A+A|$ to be very close to this upper bound, since the probability of $a_1+a_2 = a_3+a_4$ is very small (where $a_1, a_2, a_3, a_4 \in A$). On the other hand, if $A = \{1, 2, \ldots, n\}$ then $|A + A| = |\{2, 3, 4, \ldots, 2n\}| = 2|A| - 1$. The same bound holds whenever A is an arithmetic progression. One of the main problems of additive combinatorics is characterizing the finite sets A (in either $\mathbb{R}$ or some other group) for which $|A + A|$ is small with respect to $|A|$. As a warmup, we begin with the following easy claim.

Claim 1.1. *For every finite set $A \subset \mathbb{R}$ we have $|A + A| \geq 2|A| - 1$.*

Proof. Denote the elements of A as $a_1 < a_2 < \cdots < a_{|A|}$. Then $A + A$ contains the following $2|A| - 1$ distinct elements:

$$a_1 + a_1 < a_1 + a_2 < a_1 + a_3 < \cdots < a_1 + a_{|A|} < a_2 + a_{|A|} < a_3 + a_{|A|} < \cdots < a_{|A|} + a_{|A|}.$$

□

The claim establishes that, in $\mathbb{R}$, arithmetic progressions have minimum-sized sum sets. The following set is not an arithmetic progression, although it is defined in a similar way.

$$A = \big\{3k_1 + 100k_2 : k_1 \in \{1, 2, 3\} \text{ and } k_2 \in \{1, 2, 3, \ldots, n\}\big\}. \tag{1}$$

Notice that $|A| = 3n$ and $|A + A| = 5(2n - 1) = 10n - 5 < \frac{10}{3}|A|$. Although $\frac{10}{3}|A|$ may seem rather large compared to $2|A| - 1$, it is still relatively small with respect to most finite sets $A \subset \mathbb{R}$ (recall that a random set is expected to satisfy $|A+A| \approx |A|^2$). In general, we are interested in sets A that satisfy $|A+A| \leq k|A|$ for some constant k (as $|A|$ grows asymptotically). We say that such sets have *small doubling*. Note that property of a set A having small doubling does not depend on the size $|A + A|$ but rather on the ratio between $|A + A|$ and $|A|$.

A *generalized arithmetic progression* of dimension d is defined as

$$\left\{ a + \sum_{j=1}^{d} k_j b_j : a, b_1, \ldots, b_d \in \mathbb{R} \text{ and with integer } 0 \leq k_j \leq n_j - 1 \text{ for every } 1 \leq j \leq d \right\}.$$

An arithmetic progression is a generalized arithmetic progression of dimension 1, and the set in (1) is a generalized arithmetic progression of dimension 2. The size of a generalized arithmetic progression of dimension d is at most $n = n_1 n_2 \cdots n_d$, and it is not difficult to verify that it has a sumset of size smaller than $2^d n$. The following result characterizes the finite sets $A \subset \mathbb{R}$ that have a small doubling (e.g., see [6]).

Theorem 1.2 (Freiman's theorem over the reals). *Let $A \subset \mathbb{R}$ be a finite set with $|A + A| \leq k|A|$ for some constant k. Then A is contained in a generalized arithmetic progression of size at most cn and dimension at most d; both c and d depend on k but not on $|A|$.*

If we replace $\mathbb{R}$ with a finite group $\mathcal{G}$, smaller sum sets can be obtained. When $A = \mathcal{G}$ we obviously have $|A + A| = |A|$. More generally, we have $|A + A| = |A|$ when A is a subgroup or a coset of $\mathcal{G}$. For example, when working in $\mathbb{F}_9$, the set $A = \{1, 4, 7\}$ satisfies $|A + A| = |A|$.

The following argument is taken from [2].

Lemma 1.3. *Consider a set $A \subset \mathcal{G}$ such that $|A - A| < 3|A|/2$. Then $A - A$ is a subgroup of $\mathcal{G}$.*

Proof. Let $x = a - a'$ for some $a, a' \in A$ (that is, $x \in A - A$). We have

$$|A \cap (A + x)| = |(A - a) \cap (A - a')|,$$

since subtracting a from both A and $A+x$ does not change the size of their intersection. Notice that $|A - a| = |A - a'| = |A|$ and that every element of $A - a$ or $A - a'$ is in $A - A$. Thus, we obtain

$$\begin{aligned} |(A - a) \cap (A - a')| &= |A - a| + |A - a'| - |(A - a) \cup (A - a')| \\ &\geq |A - a| + |A - a'| - |A - A| > |A| + |A| - 3|A|/2 = |A|/2. \end{aligned}$$

The above implies that for every $x \in A - A$, we have $|A \cap (A + x)| > |A|/2$. This implies that for any $x, y \in A-A$, the intersection $(A+x)\cap(A+y)$ is non-empty. That is, there exist $a, a' \in A$ such that $a+x = a'+y$, or equivalently $x-y = a'-a \in A-A$. Since this holds for any $x, y \in A - A$, we get that $A - A$ is closed under subtraction. It is not difficult to verify that this closure property implies that $A - A$ is a subgroup of $\mathcal{G}$ (for example, see the preliminaries chapter). □

To see that Lemma 1.3 is tight, consider the set $A = \{0, 1\}$ in $\mathbb{Z}$. In this case $|A - A| = 3|A|/2$ although $A - A$ is not a subgroup.

Our final example is about sets in $\mathbb{R}^n$. The following result is by Freiman [1], one of the founding fathers of the field.

Lemma 1.4. *Consider a finite set $A \subset \mathbb{R}^n$ with $|A + A| \leq k|A|$, for some constant k. Then A is fully contained in a subspace of dimension $2k$.*

Proof. Let $M(A)$ be the set of midpoints of pairs of points of A:

$$M(A) = \left\{ p \in \mathbb{R}^n : p = \frac{a + b}{2} \text{ for } a, b \in A \right\}.$$

In this definition of $M(A)$ we allow $a = b$, which implies that $A \subseteq M(A)$. Since $M(A)$ is obtained by dividing every element of $A+A$ by 2, we have $|M(A)| = |A+A| \leq k|A|$. We denote by $C(A)$ the convex hull of A, and let d denote the dimension of $C(A)$. We prove by induction on $|A| + d$ that

$$|M(A)| \geq (d + 1)|A| - \binom{d + 1}{2}. \tag{2}$$

For the induction basis, the claim is clear when $|A| \leq 2$ (in this case $d = |A| - 1$). For the induction step we consider an arbitrary point $a \in A$ that is a vertex of $C(A)$, and set $A' = A \setminus \{a\}$. If $C(A')$ is of dimension $d-1$, then the set of midpoints $(a+A)/2$ consists of $|A|$ distinct points. In this case, by the induction hypothesis we have

$$|M(A)| = |A| + |M(A')| \geq |A| + d|A'| - \binom{d}{2} = |A|(d+1) - \binom{d+1}{2}.$$

Next, consider the case where $C(A')$ is of dimension d. In this case, on the boundary of $C(A)$ the vertex a is connected by an edge to at least d other vertices. Each of these vertices forms a distinct midpoint with a. The induction hypothesis implies

$$|M(A)| \geq (d+1) + |M(A')| \geq (d+1) + |A'|(d+1) - \binom{d+1}{2} = |A|(d+1) - \binom{d+1}{2}.$$

This completes the induction step, and thus the proof of (2). We now return to proving the lemma. If $C(A)$ is of dimension at least $2k$ then $|A| > 2k$. By combining this with (2) we get

$$|A+A| = |M(A)| \geq |A|(2k+1) - \binom{2k+1}{2} > |A|(2k+1) - |A|\frac{2k+1}{2} > |A|k.$$

Since this contradicts the assumption of the lemma, the dimension of $C(A)$ is at most $2k-1$. This in turn implies that A is contained in a subspace of dimension $2k$, as asserted. □

The above examples seem to hint of a general principle: If $|A+A| \leq k|A|$ (for some constant k) then we know a lot about the general structure of A. Below and in the following chapters we will keep studying this principle.

2 Ruzsa Calculus

The following lemma presents a simple useful tool.

Lemma 2.1 (Ruzsa's triangle inequality). *For any abelain group $\mathcal{G}$ and $A, B, C \subset \mathcal{G}$, we have*

$$|A||B-C| \leq |A+B||A+C|.$$

Proof. For every $x \in B - C$ we arbitrarily fix a representation $x = b - c$ with $b \in B$ and $c \in C$. We define a map $f : A \times (B-C) \to (A+B) \times (A+C)$ as follows. For any $a \in A$ and $x \in B-C$ with fixed representation $x = b-c$, we set $f(a, x) = (a+b, a+c)$. If $f(a, x) = f(a', x') = (m, n)$ then we have $x = x' = m - n$. Since we fixed a specific representation $x = b - c$, we know what b and c are, and have $a = a' = m - b$. That is, f is injective, which in turn implies that the size of the domain of f is at most the size of the image of f. In other words, $|A||B - C| \leq |A + B||A + C|$. □

We now present a couple of simple applications of Ruzsa's triangle inequality. The first application considers two distinct sets with a small sum.

Corollary 2.2. *Let A and B be subsets of an abelian group $\mathcal{G}$ such that $|A + B| \leq k\sqrt{|A||B|}$. Then $|A - A| \leq k^2|A|$.*

Proof. By Ruzsa's triangle inequality we have

$$|B||A - A| \leq |A + B||A + B| \leq k^2|A||B|.$$

The assertion of the corollary is obtained by cancelling $|B|$ from both sides of this inequality. □

The second application considers sets that are obtained by summing an arbitrary subset with a subgroup.

Corollary 2.3. *Let H be a subgroup of an abelian group $\mathcal{G}$, let A be any subset of $\mathcal{G}$, and let $B = A + H$. Then*

$$\frac{|B + B|}{|B|} \leq \frac{|A + A + A|}{|A|}.$$

Proof. First, since H is a subgroup we have $H = -H$ and $B+B = A+H+A+H = A + A + H$. By Ruzsa's triangle inequality we have

$$|A||B+B| = |A||(A+A)+H| \leq |A+A+A||A-H| = |A+A+A||A+H| = |A+A+A||B|.$$

The assertion of the corollary is obtained by rearranging this inequality. □

3 Plünnecke's inequality

We now consider sums of several elements from a set. For a positive integer m, we set

$$mA = \{a_1 + \cdots + a_m : a \in a_1, \ldots, a_m \in A\}.$$

Similarly, for positive integers m and ℓ, we set

$$mA - \ell A = \{a_1 + \cdots + a_m - a_{m+1} - \cdots - a_{m+\ell} : a_1, \ldots, a_{m+\ell} \in A\}.$$

For intuition, we first consider the case of $\mathbb{R}$. If A is a random set of real numbers, then we expect $|mA| \approx |A|^m$. Similarly, in this case we expect $|mA - \ell A| \approx |A|^{m+\ell}$. On the other hand, when $A \subset \mathbb{R}$ is an arithmetic progression we have $|mA| = m|A| - m + 1$ and $|mA - \ell A| < (m + \ell)|A|$. A similar situation occurs for other types of sets with small doubling: generalized arithmetic progressions in $\mathbb{R}$, cosets of any group $\mathcal{G}$, and sets in a small subspace of $\mathbb{R}^n$. The following theorem shows that this is not a coincidence.

Theorem 3.1 (Plünnecke's inequality). *Let A and B be finite subsets of an abelian group $\mathcal{G}$, such that $|A| = |B|$ and $|A + B| \leq k|A|$. Then for any two positive integers m and ℓ, we have $|mA - \ell A| \leq k^{m+\ell}|A|$.*

This result was originally proven by Plünnecke and later rediscovered by Ruzsa. That is why it is sometimes referred to as the *Plünnecke-Ruzsa inequality*. Here we present a more recent proof from [4].

Proof of Theorem 3.1. Let B' be a non-empty subset of B that minimizes $k' = \frac{|A+B'|}{|B'|}$. That is, B' is the subset of B that has the "best additive behavior" with A. We first prove the following lemma.

Lemma 3.2. *For any finite $C \subset \mathcal{G}$ we have*

$$|A + B' + C| \leq k'|B' + C|.$$

Proof. The proof is by induction on $|C|$. For the induction basis we consider the case of $|C| = 1$. In this case

$$|A + B' + C| = |A + B'| = k'|B'| = k'|B' + C|.$$

For the induction step we assume that $|C| \geq 2$. Consider an arbitrary element $c \in C$. We set $C' = C \setminus \{c\}$ and

$$B'_c = \{b \in B' : A + b + c \subset A + B' + C'\}.$$

Using B'_c, we can define $A + B' + C$ as

$$A + B' + C = (A + B' + C') \bigcup \big((A + B' + c) \setminus (A + B'_c + c)\big).$$

Since $A + B'_c + c \subset A + B' + c$, we have

$$|A + B' + C| \le |A + B' + C'| + |A + B' + c| - |A + B'_c + c|.$$

By the induction hypothesis, we have $|A + B' + C'| \le k'|B' + C'|$. By definition, $|A + B' + c| = |A + B'| = k'|B'|$. By the minimality of B', we have $|A + B'_c + c| = |A + B'_c| \ge k'|B'_c|$. Combining all of the above, we obtain

$$|A + B' + C| \le k'(|B' + C'| + |B'| - |B'_c|). \tag{3}$$

We next define

$$B'' = \{b \in B' : b + c \in B' + C'\}.$$

Notice that if $b \in B''$ then $A + b + c \subset A + B' + C'$, so $b \in B'_c$. That is, $B'' \subset B'_c$. We use B'' to express $B' + C$ as

$$B' + C = (B' + C') \bigcup ((B' + c) \setminus (B'' + c)).$$

Since this is a disjoint union, we have

$$\begin{aligned} |B' + C| = |B' + C'| + |B' + c| - |B'' + c| &= |B' + C'| + |B'| - |B''| \\ &\ge |B' + C'| + |B'| - |B'_c|. \end{aligned} \tag{4}$$

By combining (3) and (4), we get the assertion of the lemma. □

We now use Lemma 3.2 to prove that $|mA + B'| \le k^m|B'|$. This is done by induction on m. For $m = 1$ we have $|A + B'| = k'|B'| \le k|B'|$ (by the minimality of k'). For $m \ge 2$, by applying Lemma 3.2 with $C = (m - 1)A$ and then the induction hypothesis, we get

$$|A + B' + (m - 1)A| \le k'|B' + (m - 1)A| \le k'k^{m-1}|A| \le k^m|B'|.$$

This completes the induction step, and we can now compete the proof of the theorem. By Ruzsa's triangle inequality (Lemma 2.1), we have

$$|B'||mA - \ell A| \le |mA + B'||\ell A + B'| \le k^{m+\ell}|B'|^2 \le k^{m+\ell}|A||B'|.$$

Cancelling $|B'|$ on both sides yields the assertion of the theorem. □

One question of the 2012 *International Mathematics Competition* asked to prove Theorem 3.1 without the $-\ell A$ part. None of the 316 participant managed to receive more than 2 point for this question (out of 10).

4 A variant of Freiman's theorem

Freiman's Theorem (Theorem 1.2) characterizes the sets of real numbers that have a small sum set. We conclude this chapter by proving a variant of Freiman's Theorem for groups with elements of a bounded order.

Theorem 4.1 (Ruzsa [5]). *Let $\mathcal{G}$ be an abelian group such that every element of $\mathcal{G}$ is of order at most r. Consider a subset $A \subset \mathcal{G}$ such that $|A+A| \leq k|A|$. Then A is contained in a coset of size at most $k^2 r^{k^4}|A|$.*

Proof. We consider an arbitrary $a \in A$ and set $A' = A - a$. We have $|A| = |A'|$ and $|A+A| = |A'+A'|$. The reason for considering A' rather than A is to force 0 to be in our set. We will prove that A' is contained in a subgroup of $\mathcal{G}$, which would imply that A is contained in a coset.

Let B be a maximum sized subset of $2A' - A'$ such that for every $b, b' \in B$ we have $(b - A') \cap (b' - A') = \emptyset$. Since $b - A' \subset 2A' - 2A'$, we have $|B| \leq \frac{|2A'-2A'|}{|A'|}$. By Plünnecke's inequality (Theorem 3.1), we get

$$|B| \leq \frac{|2A' - 2A'|}{|A'|} \leq k^4.$$

Consider $x \in 2A' - A'$. By the maximality of B, there exists $b \in B$ such that $(x - A') \cap (b - A') \neq \emptyset$. That is, there exist $a, a' \in A'$ such that $x - a = b - a'$. Since this can rearranged as $x = b + a - a' \in B + A' - A'$, we obtain that $2A' - A' \subset B + A' - A'$. We now show that for $k \geq 1$, we have

$$kA' - A' \subset (k-1)B + A' - A'. \tag{5}$$

We prove (5) by induction on k. The case of $k = 1$ is trivial, and the case of $k = 2$ was proved in the previous paragraph. For $k \geq 3$, by the induction hypothesis we get

$$kA' - A' = A' + ((k-1)A' - A') \subset A' + ((k-2)B + A' - A').$$

The case of $k = 2$ states that $2A' - A' \subset B + A' - A'$, so

$$kA' - A' \subset A' + ((k-2)B + A' - A') \subset (k-1)B + A' - A'.$$

This completes the induction step, and the proof of (5).

We denote by $\mathrm{Span}(A')$ the subgroup of $\mathcal{G}$ that is spanned by A'. Equivalently, $\mathrm{Span}(A') = \bigcup_{k \geq 1} kA'$. Since $0 \in A'$ and by (5), we have

$$\mathrm{Span}(A') = \bigcup_{k \geq 1} kA' \subset \bigcup_{k \geq 1} (kA' - A') \subset \bigcup_{k \geq 1} (k-1)B + A' - A' = \mathrm{Span}(B) + A' - A'.$$

Combining this with Plünnecke's inequality (Theorem 3.1) implies

$$|\mathrm{Span}(A')| \leq |\mathrm{Span}(B)||A' - A'| \leq |\mathrm{Span}(B)| \cdot k^2|A'|.$$

Finally, since $|B| \leq k^4$, we have $|\mathrm{Span}(B)| \leq r^{k^4}$. That is, A' is contained in the subgroup $\mathrm{Span}(A')$ of size at most $r^{k^4}k^2|A'| = r^{k^4}k^2|A|$. □

In Theorem 4.1 the dependence in k is very bad — the bound is super-exponential in k. Other variants of the theorem (such as Theorem 1.2) also have a bad dependence in k. The *polynomial Freiman-Ruzsa conjecture* suggests that polynomial dependence in k should be possible under certain restrictions. This is one of the main open problems of additive combinatorics and has a large number of applications (e.g., see [3]). We now present a variant of this conjecture in $\mathbb{F}_2^n$.

Conjecture 4.2 (Polynomial Freiman-Ruzsa over $\mathbb{F}_2^n$). *Consider a set $A \subset \mathbb{F}_2^n$ with $|A + A| \leq k|A|$. Then there exists a subset $A' \subset A$ of size $|A'| \geq k^{-c}|A|$ such that $|\mathrm{Span}(A')| \leq k^c|A|$ (for some constant $c > 0$).*

The conjecture states that a set A with a small doubling contains a large subset that is fully contained in a subspace of size at most $|A|$. In Chapter 5 of these lecture notes we will study the current best bound for this problem.

References

[1] G. A. Freiman, *Foundations of a structural theory of set addition*, American Mathematical Soc., 1973.

[2] I. Łaba, Fugledes conjecture for a union of two intervals, *Proceedings of the American Mathematical Society* **129** (2001), 2965–2972.

[3] S. Lovett, Additive Combinatorics and its Applications in Theoretical Computer Science, Manuscript.

[4] G. Petridis, Plünnecke's inequality, *Combinatorics, Probability and Computing* **20** (2011): 921–938.

[5] I. Z. Ruzsa, An analog of Freimans theorem in groups, in *Structure Theory of Set Addition*, Astérisque, 323–326, 1999.

[6] I. Z. Ruzsa, Sumsets and structure, in *Combinatorial number theory and additive group theory* (2009), 87–210.