A talk given at the Institute of Math., Chinese Academy of Sciences (2007-01-30).

COMBINATORIAL ASPECTS OF SZEMERÉDI'S THEOREM

Zhi-Wei Sun

Department of Mathematics
Nanjing University
Nanjing 210093, P. R. China
zwsun@nju.edu.cn
http://pweb.nju.edu.cn/zwsun

Abstract. In the first section we review the beautiful combinatorial theory of Ramsey as well as the history of Szemerédi's theorem. In Section 2 we give a sketch of Ruzsa's modern proof of Freiman's theorem on sumsets which plays an important role in Gowers' quantitative proof of Szemerédi's theorem. In the last section we deduce Roth's theorem (Szemerédi's theorem in the case k=3) from the Triangle Removal Lemma (in graph-theoretic language) which is an application of Szemerédi's Regularity Lemma (a powerful tool in graph theory), and show the Balog-Szemerédi-Gowers theorem by a graph-theoretic method.

1. Introduction to Ramsey's theory and Szemerédi's theorem

Pigeon-hole Principle or Dirichlet's Principle. *If we put at least $n+1$ objects into n drawers where $n \in \mathbb{Z}^+ = \{1, 2, \ldots\}$, then some drawer contains at least two objects.*

Note that a distribution of all elements of a set A into n drawers corresponds to an ordered partition $A = A_1 \cup \cdots \cup A_n$ with $A_1, \ldots, A_n$ pairwise disjoint. We may also call such a partition an n-coloring of A with elements in A_i colored the ith color.

Applying the pigeon-hole principle, Erdős and Szekeres proved that any sequence of real numbers with $mn+1$ terms contains a monotonic increasing subsequence of $m+1$ terms or a monotonic decreasing subsequence of $n+1$ terms.

An ordered partition $A_1\cup\cdots\cup A_n$ of a set A corresponds to an ordered partition $\mathcal{A}_1\cup\cdots\cup\mathcal{A}_n$ of $\mathcal{A}=\{\{x\}:x\in A\}$ where $\mathcal{A}_i=\{\{x\}:x\in A_i\}$. We call a subset of A with cardinality r an r-subset of A.

The following deep generalization of the Pigeon-hole principle was obtained by F. P. Ramsey in 1930 in his paper "*On a problem of formal logic*" in which he aimed to prove a logical result which is actually impossible by an undecidable result of Gödel.

Ramsey's Theorem (Ramsey, 1930)**.** *Let $r>0$ and $q_1,\ldots,q_n\geqslant r$ be integers. If S is a set with $|S|$ large enough and we arbitrarily distribute all r-subsets of S into n ordered drawers, then for some $1\leqslant i\leqslant n$ the set S has a q_i-subset whose all r-subsets lie in the ith drawer.*

A Sketch of the Proof. In the case $n=1$ it suffices to let $|S|\geqslant q_1$.

Now we handle the case $n=2$. If $r=1$ then it suffices to let $|S|\geqslant q_1+q_2-1$. If $r>1$ and the result holds for smaller values of r, then by induction on q_1+q_2 we can prove the desired result for r.

If the desired result holds for a fixed $n\geqslant 2$. Then we can prove the desired result for $n+1$ by using the induction hypothesis together with the case $n=2$ that we have handled. □

Given integers $q_1,\ldots,q_n\geqslant r\geqslant 1$, the *Ramsey number* $R_r(q_1,\ldots,q_n)$ is

the smallest positive integer such that however we distribute all r-subsets of a set S with $|S| \geqslant R_r(q_1, \ldots, q_n)$ into n ordered drawers there exists $1 \leqslant i \leqslant n$ such that S has a q_i-subset whose all r-subsets lie in the ith drawer.

Ramsey's theorem in the special case $r = 1$ and $q_1 = \cdots = q_n = 2$ reduces to the Pigeon-hole principle. It is easy to see that

$$R_1(q_1, \ldots, q_n) = q_1 + \cdots + q_n - n + 1.$$

It was P. Erdős who first recognized the importance of Ramsey's theorem. Roughly speaking, Ramsey's theorem indicates that *complete disorder is impossible.* In 1935 Erdős and Szekeres proved that if $q_1, q_2 \geqslant 2$ then

$$R_2(q_1, q_2) \leqslant \binom{q_1 + q_2 - 2}{q_1 - 1}.$$

In 1947 Erdős showed that $R_2(q, q) \geqslant 2^{m/2}$ and his probability method greatly influenced the later development of combinatorics.

It is difficult to determinate exact values of Ramsey numbers. It is known that

$$R_2(3,3) = 6,\ R_2(3,4) = 9,\ R_2(3,5) = 14,\ R_2(3,6) = 18,$$

$$R_2(3,7) = 23,\ R_2(3,8) = 28,\ R_2(3,9) = 36,\ R_2(4,4) = 18,\ R(4,5) = 25$$

and $c_1 n^2/\ln n \leqslant R_2(3,n) \leqslant c_2 n^2/\ln n$ (Shearer (1983) and Kim (1995)) Also,

$$R_2(3,3,\ldots,3) \leqslant \lfloor n!e \rfloor + 1 \quad \text{(Greenwood and Gleason).}$$

Erdős-Szekeres Theorem (Erdős-Szekeres, 1935). *Let $m \geqslant 3$ be a positive integer. Given sufficiently many points in a plane no three of which are co-linear, we can select m of them so that they generate a convex polygon with m sides.*

Proof. Suppose that we are given $n \geqslant R_4(5,m)$ points $P_1,\ldots,P_n$ in a plane no three of which are co-linear. Put a 4-subset $\{P_i,P_j,P_k,P_l\}$ of $S=\{P_1,\ldots,P_n\}$ into a "concave" drawer or a "convex" drawer according as P_i,P_j,P_k,P_l are vertices of a 4-sided concave polygon or a 4-sided convex polygon. As $n\geqslant R_4(5,m)$, either there is a 5-subset of S whose all 4-subsets are in the concave drawer, or there is an m-subset of S whose all 4-subsets are in the convex drawer. It can be easily shown that the former case cannot happen, and in the latter case the m points in S generate an m-sided polygon. □

Let $ES(m)$ denote the minimal number of points that are needed to guarantee the result in the above theorem. Erdős and Szekeres showed that

$$2^{m-2}+1\leqslant ES(m)\leqslant \binom{2m-4}{m-2}+1$$

and they conjectured that $ES(m)=2^{m-2}+1$. This famous conjecture remains open.

Schur's Theorem (Schur, 1916). *If we distribute $1,\ldots,\lfloor n!e\rfloor$ into n drawers, then some drawer contains certain x,y,z with $x+y=z$.*

Proof. For $k=1,\ldots,n$, let A_k be the set of integers contained in the kth

drawer and set

$$\mathcal{A}_k = \{\{i,j\} :\ 1 \leqslant i < j \leqslant \lfloor n!e \rfloor + 1 \text{ and } j - i \in A_k\}.$$

Then $\mathcal{A}_1, \ldots, A_n$ are pairwise disjoint and their union consists of all the 2-subsets of $S = \{1, \ldots, \lfloor n!e \rfloor + 1\}$. As $|S| \geqslant R_2(3, \ldots, 3)$, there is a 3-subset $\{a,b,c\}$ $(a < b < c)$ of S such that $\{a,b\}, \{b,c\}, \{a,c\}$ belong to the same $\mathcal{A}_k$. Thus

$$x = b - a \in A_k, \quad y = c - b \in A_k, \quad \text{and } z = c - a \in A_k.$$

Note that $1 \leqslant x, y, z \leqslant \lfloor n!e \rfloor$ and $x + y = z$. We are done. □

In 1927 van der Waerden established the following result conjectured by Schur, this contribution made him famous as a young mathematician.

van der Waerden Theorem. *For any positive integers k and m, if n is sufficiently large and we distribute $1, \ldots, n$ into k drawers, then some drawer contains an AP (arithmetic progression) with m terms.*

In 1933 R. Rado, one of Schur's students, proved the following theorem which includes both Schur's theorem and van der Waerden's theorem as special cases.

Rado's Theorem. *Let $A = (a_{ij})_{1 \leqslant i \leqslant m,\, 1 \leqslant j \leqslant n}$ be a matrix with $a_{ij} \in \mathbb{Z}$. Then the equation $A(x_1, \ldots, x_n)^T = 0$ is partition regular (i.e., however we distribute all positive integers into finitely many drawers the equation always has a solution with $x_1, \ldots, x_n$ in the same drawer), if and only if we can renumber the column vectors of A so that there are integers*

$1 \leqslant n_1 < n_2 < \cdots < n_l = n$ *for which the sum of the first* n_k $(1 \leqslant k \leqslant l)$ *column vectors is a rational linear combination of the first* n_{k-1} *column vectors, where we set* $n_0 = 0$.

For $a_1, \ldots, a_m \in \mathbb{Z}^+$ with $m \geqslant 2$, we define the two-color Rado number $R(a_1, \ldots, a_m)$ to be the least positive integer N such that for every 2-coloring of the set $[1, n] = \{1, \ldots, n\}$ with $n \geqslant N$ there exists a monochromatic solution to the equation $a_1x_1 + \cdots + a_mx_m = x_0$ with $x_0, \ldots, x_m \in [1, n]$.

In 2005 S. Guo and Z. W. Sun [J. Combin. Theory Ser. A, to appear] determined the exact value of $R(a_1, \ldots, a_m)$.

A Result of S. Guo and Z. W. Sun (conjectured by B. Hopkins and D. Schaal). *For any* $a_1, \ldots, a_m \in \mathbb{Z}^+$ *with* $m \geqslant 2$, *we have*

$$R(a_1, \ldots, a_m) = av^2 + v - a,$$

where

$$a = \min\{a_1, \ldots, a_m\} \quad \text{and} \quad v = a_1 + \cdots + a_m.$$

Shelah's Pigeon-hole Principle. *Let* $k, m, n \in \mathbb{Z}^+$ *and* $m \geqslant f(n, k)$, *where* $f(1, k) = k+1$ *and* $f(j+1, k) = k^{f(j,k)^{2j}} + 1$ *for* $j = 1, 2, \ldots$. *Then, for any* k*-colorings* $c_1, \ldots, c_n : [1, m]^{2n} \to [1, k]$, *there are* $1 \leqslant a_1 < b_1 \leqslant m, \ldots, 1 \leqslant a_n < b_n \leqslant m$ *such that*

$$\begin{aligned} &c_j(a_1, b_1, \ldots, a_{j-1}, b_{j-1}, a_j, a_j, a_{j+1}, b_{j+1}, \ldots, a_n, b_n) \\ =&c_j(a_1, b_1, \ldots, a_{j-1}, b_{j-1}, b_j, b_j, a_{j+1}, b_{j+1}, \ldots, a_n, b_n) \end{aligned} \tag{$*$}$$

for all $j = 1, \ldots, n$.

Proof. We use induction on n.

In the case $n = 1$, as $|\{(a,a) : a \in [1,m]\}| = m \geqslant f(1,k) > k$, by the usual pigeon-hole principle there are $1 \leqslant a < b \leqslant m$ such that $c_1(a,a) = c_1(b,b)$.

Now let $n \geqslant 1$ and assume the desired result for n. Let $m \geqslant f(n+1,k) > k^{f(k,n)^{2n}}$ and $c_1, \dots, c_{n+1}$ be mappings from $[1,m]^{2n+2}$ to $[1,k]$. For $a, b \in [1,m]$ we define $c_{a,b} : [1, f(n,k)]^{2n} \to [1,k]$ by

$$c_{a,b}(a_1, b_1, \dots, a_n, b_n) = c_{n+1}(a_1, b_1, \dots, a_n, b_n, a, b) \in [1,k].$$

Since

$$m > k^{f(n,k)^{2n}} = |\{f : [1, f(n,k)]^{2n} \to [1,k]\}|,$$

by the pigeon-hole principle there are $1 \leqslant a_{n+1} < b_{n+1} \leqslant m$ such that $c_{a_{n+1},a_{n+1}} = c_{b_{n+1},b_{n+1}}$, i.e., for any $a_1, b_1, \dots, a_n, b_n \in [1, f(n,k)]$ we have

$$c_{n+1}(a_1, b_1, \dots, a_n, b_n, a_{n+1}, a_{n+1}) = c_{n+1}(a_1, b_1, \dots, a_n, b_n, b_{n+1}, b_{n+1}).$$

For $j = 1, \dots, n$ define $c'_j : [1, f(n,k)]^{2n} \to [1,k]$ by

$$c'_j(x_1, \dots, x_{2n}) = c_j(x_1, \dots, x_{2n}, a_{n+1}, b_{n+1}).$$

By the induction hypothesis, there are $1 \leqslant a_1 < b_1 \leqslant N(n,k), \dots, 1 \leqslant a_n < b_n \leqslant f(n,k)$ such that whenever $1 \leqslant j \leqslant n$ we have

$$\begin{aligned} &c'_j(a_1, b_1, \dots, a_{j-1}, b_{j-1}, a_j, a_j, a_{j+1}, b_{j+1}, \dots, a_n, b_n) \\ =&c'_j(a_1, b_1, \dots, a_{j-1}, b_{j-1}, b_j, b_j, a_{j+1}, b_{j+1}, \dots, a_n, b_n), \end{aligned}$$

i.e.,

$$\begin{aligned}&c_j(a_1,b_1,\ldots,a_{j-1},b_{j-1},a_j,a_j,a_{j+1},b_{j+1},\ldots,a_{n+1},b_{n+1})\\=&c_j(a_1,b_1,\ldots,a_{j-1},b_{j-1},b_j,b_j,a_{j+1},b_{j+1},\ldots,a_{n+1},b_{n+1}).\end{aligned}$$

Note that $f(n,k)\leqslant f(n+1,k)\leqslant m$ ($f(n,k)=2$ if $k=1$). So the desired result for $n+1$ follows from the above. $\square$

For $n,k\in\mathbb{Z}^+$ we define the *Shelah number* $S(n,k)$ to be the smallest $m\in\mathbb{Z}^+$ such that for any $c_1,\ldots,c_n\in[1,m]^{2n}\to[1,k]$ there are $1\leqslant a_1<b_1\leqslant m,\ldots,1\leqslant a_n<b_n\leqslant m$ such that $(*)$ holds for all $j=1,\ldots,n$. By Lemma 2.1 and its proof, $S(n,k)\leqslant f(n,k)$; furthermore,

$$S(1,k)=k+1\ \text{ and }\ S(n+1,k)\leqslant k^{S(n,k)^{2n}}+1.$$

Let S be a finite nonempty set. A *combinatorial line* in S^n has the form

$$\begin{aligned}L=\{(x_1,\ldots,x_n)\in S^n:&\text{ all those }x_i\text{ with }i\in I\text{ are equal,}\\&\text{and those }x_j\text{ with }j\notin I\text{ are fixed}\},\end{aligned}$$

where I is a nonempty subset of $[1,n]$.

In 1963 Hales and Jewett established a Ramsey-type result which stripes the van der Waerden theorem of its unessential elements and reveals the heart of Ramsey Theory.

Hales-Jewett Theorem. *For any $m,k\in\mathbb{Z}^+$ if $n\in\mathbb{Z}^+$ is large enough then for every k-coloring of $[1,m]^n$, $[1,m]^n$ contains a monochromatic combinatorial line.*

Denote the smallest n in the Hales-Jewett theorem by $HJ(m,k)$. In 1988 Shelah used his pigeon-hole principle to show that

$$HJ(m+1,k) \leqslant HJ(m,k)S\left(HJ(m,k), k^{(m+1)^{HJ(m,k)}}\right),$$

which yields the first primitive upper bound for $HJ(m,k)$.

Proof of the van der Waerden Theorem from the Hales-Jewett Theorem. Let $h = HJ(m,k)$. For $x_1, \dots, x_h \in [1,m]$ define

$$F(x_1, \dots, x_h) = 1 + \sum_{i=1}^{h-1}(x_i - 1)m^i.$$

Then F is a one-to-one correspondence between $[1,m]^h$ and $[1,m^h]$. Any distribution of $1, \dots, m^h$ into k drawers corresponds to a distribution of k-coloring of $[1,m]^h$. By the Hales-Jewett, $[1,m]^h$ contains a monochromatic combinatorial line

$$\{(x_1, \dots, x_h) \in [1,m]^h : \text{those } x_i \text{ with } i \in I \text{ are equal, } x_j = a_j \text{ for } j \notin I\},$$

where $\emptyset \neq I \subseteq [1,h]$ and $a_j \in [1,m]$ for $j \in \bar{I} = [1,h] \setminus I$. Thus, those numbers

$$1 + \sum_{j \in \bar{I}}(a_j - 1)m^{j-1} + \sum_{i \in I}(x-1)m^{i-1} \quad (x = 1, \dots, m)$$

lie in the same drawer. In other words, some drawer contains the arithmetic progression $a, a+d, \dots, a+(m-1)d$, where

$$a = 1 + \sum_{j \in \bar{I}}(a_j - 1)m^{j-1} \quad \text{and} \quad d = \sum_{i \in I} m^{i-1}.$$

We are done. $\square$

Erdős-Graham Conjecture proved by Croot. *If we distribute all integers greater than one into n drawers, then some drawer contains integers $x_1,\ldots,x_m$ with $\sum_{k=1}^m 1/x_k=1$.*

Here we pose a general conjecture from which Croot's result follows immediately.

A Conjecture of Z. W. Sun (Jan. 28, 2007)**.** *If A is a subset of $\{2,3,\ldots\}$ with positive upper (asymptotic) density, then there are finitely many distinct elements $a_1<\cdots<a_m$ of A with $\sum_{k=1}^m 1/a_k=1$.*

The following deep result conjectured by P. Erdős and P. Turán in 1936, implies the van der Waerden theorem.

Szemerédi's Theorem. *Let $0<\delta\leqslant 1$ and $k\in\{3,4,\ldots\}$. Then there is $N(k,\delta)$ such that if $n\geqslant N(k,\delta)$ and $A\subseteq[1,n]$ with $|A|\geqslant\delta n$ then A contains an AP of length k.*

In 1956 K. Roth proved this result for $k=3$ by the circle method in analytic number theory. In 1969 E. Szemerédi handled the case $k=4$ by a combinatorial method. The case of general k was settled by Szemerédi in 1975 in a paper which was regarded as "*a masterpiece of combinatorial reasoning*" by R. L. Graham. In 1977 H. Furstenberg used ergodic theory to give a new proof of Szemerédi's theorem. In 2001 W. T. Gowers employed Fourier analysis and combinatorics (including Frieman's theorem on sumsets) to reprove the theorem with explicit bounds.

Here are the best known bounds for $N(k,\delta)$:

$$c^{\log(1/\delta)^{k-1}}\leqslant N(k,\delta)\leqslant 2^{2^{\delta^{-2^{2^{k+9}}}}},$$

where the lower bound is due to Behrend (for $k = 3$) and Rankin (1962), and the upper bound is due to Gowers (2001). In 1999 J. Bourgain showed that $N(3, \delta) \leqslant c^{\delta^{-2}\log(1/\delta)}$.

Szemerédi's theorem plays an important role in the proof of the following celebrated result.

Green-Tao Theorem. *There are arbitrarily long APs of primes.*

The following difficult conjecture includes both Szemerédi's theorem and the Green-Tao theorem as special cases.

Erdős-Turán Conjecture. *Let $a_1 < a_2 < \cdots$ be a sequence of positive integers with $\sum_{n=1}^{\infty} 1/a_n$ divergent. Then, for any $k = 3, 4, \ldots$ the sequence has a subsequence which is an AP of length k.*

In my opinion this conjecture might be too strong to hold. I'd like to modify this conjecture as follows: If $a_1 < a_2 < \cdots$ is a sequence of positive integers with $\sum_{n=1}^{\infty} 1/a_n = \infty$ and $\sum_{i\in I} 1/a_i \notin \mathbb{Z}^+$ for any finite subset I of $\mathbb{Z}^+$, then the sequence contains arbitrarily long APs.

2. Ruzsa's approach to Freiman's theorem on sumsets

Let $A_1, \ldots, A_n$ be subsets of an abelian group. We define the *sumset*

$$A_1 + \cdots + A_n = \{a_1 + \cdots + a_n : \ a_1 \in A_1, \ldots, a_n \in A_n\}$$

and denote it by nA if $A_1 = \cdots = A_n = A$. It can be easily showed that

$$|A_1 + \cdots + A_n| \geqslant \sum_{i=1}^{n} |A_i| - n + 1,$$

and equality holds if and only if $A_1,\ldots,A_n$ are arithmetic progressions with the same common difference.

The following deep result of G. Freiman on sumsets first appeared in a paper of Freiman in 1964 and then in his 1966 monograph.

Freiman's Theorem. *Let A be a finite nonempty subset of $\mathbb{Z}$ with $|2A|\leqslant c|A|$. Then A is contained in an n-dimensional AP*

$$Q=Q(a;q_1,\ldots,q_n;l_1,\ldots,l_n)=\{a+x_1q_1+\cdots+x_nq_n:\ 0\leqslant x_i<l_i\}$$

with $|Q|\leqslant c'|A|$, where c' and n only depend on c.

The n-dimensional AP Q mentioned above is said to have length $l(Q)=l_1\cdots l_n$, and it is called *proper* if $|Q|=l(Q)$.

Since Freiman's theorem plays a crucial role in Gowers' quantitative proof of Szemerédi's theorem, in this section we will present a sketch of Rusza's modern approach to Freiman's theorem.

Given a finite sequence $(A_i)_{i=1}^n$ of sets, if $a_1\in A_1,\ldots,a_n\in A_n$ and $a_i\neq a_j$ for all $1\leqslant i<j\leqslant n$, then we call the sequence $(a_i)_{i=1}^n$ an SDR (*system of distinct representatives*) of $(A_i)_{i=1}^n$.

Hall's Theorem (P. Hall, 1935)**.** *Let $A_1,\ldots,A_n$ be sets. Then $(A_i)_{i=1}^n$ has an SDR if and only if $|\bigcup_{i\in I}A_i|\geqslant|I|$ for all $I\subseteq[1,n]$.*

A natural induction proof of Hall's theorem was given by Z. W. Sun in 2001. Hall's theorem is very important in discrete mathematics, it reveals the fundamental combinatorial min-max relation.

Defect Form of Hall's Theorem. *Let $A_1, \dots, A_n$ be sets. Then*

$$\max\{|J| : J \subseteq [1,n] \text{ and } (A_j)_{j\in J} \text{ has an SDR}\} = \min_{I\subseteq[1,n]} \left(\left| \bigcup_{i\in I} A_i \right| + n - |I| \right).$$

An undirected graph G consists of the vertex set $V(G)$ and the edge set $E(G)$ (with $V(G)\cap E(G) = \emptyset$), and each edge in $E(G)$ is an unordered pair $\{u, v\}$ with $u, v \in V(G)$. To make the structure intuitive we use points in the plane to represent vertices in $V(G)$ and join u and v if $\{u, v\}$ is an edge.

For an undirected graph $G = (V, E)$, if there are disjoint subsets X and Y of $V = V(G)$ such that $X \cup Y = V$ and each edge $e \in E = E(G)$ has one endpoint in X and another endpoint in Y, then we call G a *bipartite graph* with vertex classes X and Y. A *matching* from X to Y in such a bipartite is a set of $|X|$ disjoint edges.

Let $A_1, \dots, A_n$ be finite sets and write $\bigcup_{i=1}^n A_i = \{a_1, \dots, a_m\}$. We make a bipartite graph G with vertex classes $V_1 = \{A_1, \dots, A_n\}$ and $V_2 = \{a_1, \dots, a_m\}$ by joining A_i and a_j if $a_j \in A_i$. Thus we can reformulate Hall's theorem as follows.

Hall's Matching Theorem. *Let G be any bipartite graph with vertex classes X and Y. Then G has a matching from X to Y if and only if $|\Gamma(S)| \geqslant |S|$ for all $S \subseteq X$, where $\Gamma(S) = \{y \in Y : (x, y) \in E(G)$ for some $x \in S\}$.*

In a directed graph, each edge has a direction. For two distinct vertices u, v in a directed graph G, a *path* from u to v is a sequence of distinct

vertices $v_0 = u, v_1, \ldots, v_l = v$ with $(v_0, v_1), \ldots, (v_{l-1}, v_l) \in E(G)$, and l is called the length of the path. Two paths from u to v are *independent* if no vertex other than u and v occurs in both paths.

In 1927 Menger discovered the following fundamental result in graph theory.

Menger's Theorem. *Let G be a directed graph, and let $a, b \in V(G)$ be distinct vertices with $(a, b) \notin E(G)$. Then the maximal number of independent paths from a to b is the smallest positive integer l such that there is a set $S \subseteq V \setminus \{a, b\}$ with $|S| = l$ which separates a and b (i.e., every path from a to b contains at least one vertex in S).*

Menger's theorem is actually equivalent to Hall's theorem.

Proof of Hall's Matching Theorem from Menger's Theorem. Join a new vertex a to all elements of X and a new vertex b to all elements of Y to form a new graph G'. Suppose that $S \subseteq V(G') \setminus \{a, b\} = X \cup Y$ separates a and b, and $|\Gamma(T)| \geqslant |T|$ for all $T \subseteq X$. Then

$$|S \cap Y| \geqslant |\Gamma(X \setminus S)| \geqslant |X \setminus S|$$

and hence

$$|S| = |S \cap X| + |S \cap Y| \geqslant |S \cap X| + |X \setminus S| = |X|.$$

Thus, by Menger's theorem, there are $|X|$ independent paths from a to b. This yields a matching in G. $\square$

Let $G = (V, E)$ be a directed graph. Suppose that there is a partition $V = V_0 \cup V_1 \cup \cdots \cup V_h$ so that $E \subseteq \bigcup_{i=1}^{h} (V_{i-1} \times V_i)$. Then we call G a

directed graph of level h. Such a graph G is called a *Plünnecke graph of level h* if it has the following properties:

(a) Suppose that $1 \leqslant i < h$, $u \in V_{i-1}$, $v \in V_i$, $w_1, \dots, w_k \in V_{i+1}$ are distinct, and $(u,v),(v,w_1),\dots,(v,w_k) \in E$. Then there are distinct $v_1,\dots,v_k \in V_i$ such that $(u,v_i),(v_i,w_i) \in E$ for all $i = 1,\dots,k$.

(b) Suppose that $1 \leqslant i < h$, $u_1,\dots,u_k \in V_{i-1}$ are distinct, $v \in V_i$, $w \in V_{i+1}$, and $(u_1,v),(u_2,v),\dots,(u_k,v),(v,w) \in E$. Then there are distinct $v_1,\dots,v_k \in V_i$ such that $(u_i,v_i),(v_i,w) \in E$ for all $i = 1,\dots,k$.

In 1969 Plünnecke obtained the following important result.

Plünnecke's Inequality. *Let G be a Plünnecke graph of level $h \geqslant 1$. Then we have*

$$D_1 \geqslant D_2^{1/2} \geqslant \cdots \geqslant D_h^{1/h},$$

where D_i is the ith magnification ratio

$$D_i(G) = \inf_{\emptyset \neq X \subseteq V_0} \frac{|\mathrm{im}_i(X)|}{|X|}$$

with

$$\mathrm{im}_i(X) = \{v \in V_i : \ G \text{ contains a path from some } x \in X \text{ to } v\}.$$

The proof of Plünnecke's inequality involves Menger's theorem and the following technical lemma.

A Lemma for Plünnecke's Inequality. (i) *If G and H are directed graphs of level h, then $D_i(G \times H) = D_i(G)D_i(H)$ for $i = 1,\dots,h$, where the product graph $G \times H$ is defined in a natural way.*

(ii) *Let G be a Plünnecke graph of level h with the partition $V=V_0\cup V_1\cup\cdots\cup V_h$. If $(u,v)\in E(G)$, then $d^+(u)\geqslant d^+(v)$ and $d^-(u)\leqslant d^-(v)$, where*

$$d^+(v)=|\{w\in V(G):\ (v,w)\in E(G)\}|$$

and

$$d^-(v)=|\{w\in V(G):\ (w,v)\in E(G)\}|.$$

If $D_h\geqslant 1$ then there are $|V_0|$ disjoint paths from vertices in V_0 to vertices in V_h.

Plünnecke-Ruzsa Theorem. *Let A and B be finite nonempty subsets of an abelian group with $|A+B|\leqslant c|A|$. Then for any $k,l\in\mathbb{N}=\{0,1,2,\ldots\}$ we have*

$$|kB-lB|\leqslant c^{k+l}|A|,$$

where we regard $0B$ as $\{0\}$.

Proof. As $0B-0B=\{0\}$, the desired result is trivial in the case $k=l=0$. Without loss of generality, below we assume that $k\leqslant l$ and $l\geqslant 1$.

Define a directed graph $G=G_{A,B}$ of level l as follows: $V(G)=\bigcup_{i=0}^l V_i$ with $V_i=A+iB$ (actually we should let $V_i=\{i\}\times(A+iB)$ since those $\{i\}\times(A+iB)$ $(i=0,\ldots,l)$ are pairwise disjoint), and

$$E(G)=\{(v_i,v_{i+1}):\ 0\leqslant i<h,\ v_i\in V_i,\ v_{i+1}\in V_{i+1}\text{ and }v_{i+1}-v_i\in B\}.$$

If $u\in V_{i-1}$, $v\in V_i$, $w_1,\ldots,w_m\in V_{i+1}$, and $(u,v),(v,w_1),\ldots,(v,w_m)\in E(G)$, then, for any $1\leqslant j\leqslant m$, we have

$$v_j:=u+(w_j-v)\in V_{i-1}+B=V_i\ \text{ and }\ (u,v_j),(v_j,w_j)\in E(G)$$

since $v_j - u = w_j - v \in B$ and $w_j - v_j = v - u \in B$. Similarly, if $u_1, \dots, u_m \in V_{i-1}$, $v \in V_i$, $w \in V_{i+1}$, and $(u_1, v), \dots, (u_m, v), (v, w) \in E(G)$, then, for any $1 \leqslant j \leqslant m$, we have

$$v_j := u_j + (w - v) \in V_{i-1} + B = V_i \quad \text{and} \quad (u_j, v_j), (v_j, w) \in E(G)$$

since $v_j - u_j = w - v \in B$ and $w - v_j = v - u_j \in B$. Thus G is a Plünnecke graph of level l.

If $k \geqslant 1$, then by Plünnecke's inequality, there is $\emptyset \neq A' \subseteq V_0 = A$ such that

$$\frac{|\mathrm{im}_k(A')|}{|A'|} = D_k(G) \leqslant D_1(G)^k \leqslant \left(\frac{|\mathrm{im}_1(A)|}{|A|}\right)^k = \frac{|A+B|^k}{|A|^k} \leqslant c^k$$

and thus $|A' + kB| \leqslant c^k |A'|$. If $k = 0$, then we take $A' = A$, and there is $\emptyset \neq A'' \subseteq A' = A$ with $|A'' + lB| \leqslant c^l |A''|$ by the same argument. Similarly, when $1 \leqslant k \leqslant l$ there is $\emptyset \neq A'' \subseteq A'$ such that

$$\frac{|A'' + lB|}{|A''|} = D_l(G_{A',B}) \leqslant D_k(G_{A',B})^{l/k} \leqslant \left(\frac{|A' + kB|}{|A'|}\right)^{l/k} \leqslant (c^k)^{l/k}.$$

Let R, S, T be finite nonempty subsets of an abelian group. Each $d \in S - T$ can be written as $s(d) - t(d)$ with $s(d) \in S$ and $t(d) \in T$. If $(r, d), (r', d') \in R \times (S - T)$ and $(r + s(d), r + t(d)) = (r' + s(d'), r' + t(d'))$, then $d = r + s(d) - (r + t(d)) = r' + s(d') - (r' + t(d')) = d'$ and $r = r' + s(d') - s(d) = r'$. Therefore

$$|R| \cdot |S - T| \leqslant |\{(r + s(d), r + t(d)) : r \in R \text{ and } d \in S - T\}| \leqslant |R + S| \cdot |R + T|.$$

In view of the above, we have

$$|A''| \cdot |kB - lB| \leqslant |A'' + kB| \cdot |A'' + lB| \leqslant |A' + kB| \cdot |A'' + lB| \leqslant c^k |A'| c^l |A''|$$

and hence $|kB - lB| \leqslant c^{k+l} |A'| \leqslant c^{k+l} |A|$. $\square$

Ruzsa's Analogue of Freiman's Theorem for Torsion Groups. *Let G be a torsion abelian group every element of which has order not exceeding r. Let A and B be finite nonempty subsets of G with $|A+B|\leqslant c|A|$. Then B is contained in a subgroup of G whose order does not exceed $c^2r^{c^4|A|/|B|}|A|$.*

Proof. By the Plünnecke-Ruzsa theorem, we have $|B-B|\leqslant c^2|A|$ and $|2B-2B|\leqslant c^4|A|$.

Let $W=\{w_1,\ldots,w_k\}$ be a maximal subset of $2B-B$ such that $w_1-B,\ldots,w_k-B$ are pairwise disjoint. Then

$$k|B|=\sum_{i=1}^k|w_i-B|=\left|\bigcup_{i=1}^k(w_i-B)\right|\leqslant|(2B-B)-B|\leqslant c^4|A|$$

and hence $k\leqslant c^4|A|/|B|$.

For any $w\in 2B-B$, there is $1\leqslant i\leqslant k$ such that $(w-B)\cap(w_i-B)\neq\emptyset$ and hence $w\in w_i-B+B\subseteq W+B-B$. So $2B-B\subseteq W+B-B$. It follows that

$$3B-B\subseteq W+2B-B\subseteq 2W+B-B,\ \ 4B-B\subseteq 2W+2B-B\subseteq 3W+B-B$$

and so on. Thus, for any $l\in\mathbb{Z}^+$, we have $lB-B\subseteq(l-1)W+B-B\subseteq H(W)+B-B$, where

$$\begin{aligned}H(W)=&\{x_1w_1+\cdots+x_kw_k:\ x_1,\ldots,x_k\in\mathbb{Z}\}\\=&\{x_1w_1+\cdots+x_kw_k:\ 0\leqslant x_i<r_i\leqslant r\text{ for }i=1,\ldots,k\}\end{aligned}$$

is the subgroup of G generated by W (and r_i is the order of w_i). Therefore $H(B)=\bigcup_{l=1}^\infty(lB-B)\subseteq H(W)+(B-B)$ and hence

$$|H(B)|\leqslant|H(W)|\cdot|B-B|\leqslant r^kc^2|A|\leqslant r^{c^4|A|/|B|}c^2|A|.$$

We are done. □

For any real number x, we define

$$\|x\| = \min_{a\in\mathbb{Z}} |x-a| = \begin{cases} \{x\} & \text{if } \{x\} \leqslant 1/2, \\ 1-\{x\} & \text{otherwise,} \end{cases}$$

where $\{x\}$ is the fractional part of x. For $m \in \mathbb{Z}^+$, $r_1, \dots, r_k \in [0, m-1]$ and $\varepsilon > 0$, we call

$$B_m(r_1, \dots, r_n; \varepsilon) = \left\{ a + m\mathbb{Z} : \ \left\| \frac{ar_i}{m} \right\| \leqslant \varepsilon \text{ for all } i = 1, \dots, n \right\}$$

a *Bohr set*.

In 1939 N. N. Bogolyubov established the following result via roots of unity.

Bogolyubov's Theorem. *Let $m \geqslant 2$ be an integer and $\emptyset \neq A \subseteq \mathbb{Z}_m = \mathbb{Z}/m\mathbb{Z}$. Then there are distinct $r_1, \dots, r_n \in [0, m-1]$ with $r_1 = 0$ and $n \leqslant (m/|A|)^2$ such that*

$$B_m\left(r_1, \dots, r_n; \frac{1}{4}\right) \subseteq 2A - 2A.$$

Proof. Write $A = \{a_1 + m\mathbb{Z}, \dots, a_k + m\mathbb{Z}\}$ where $a_1, \dots, a_k \in [0, m-1]$ are distinct. For $r \in \mathbb{Z}$ set

$$S_A(r) = \sum_{s=1}^{k} e^{2\pi i a_s r/m}.$$

For any $g \in \mathbb{Z}$, we assert that

$$g + m\mathbb{Z} \in 2A - 2A \iff \sum_{r=0}^{m-1} |S_A(r)|^4 \chi_r(g) \neq 0,$$

where $\chi_r(g)=e^{2\pi igr/m}$. In fact, if we set $A_0=\{a_1,\ldots,a_k\}$ then

$$\begin{aligned}&\sum_{r=0}^{m-1}S_A(r)^2\,\overline{S_A(r)}^2\chi_r(g)\\=&\sum_{r=0}^{m-1}\sum_{a,b,c,d\in A_0}e^{2\pi i(g-a-b+c+d)r/m}\\=&\sum_{a,b,c,d\in A_0}\sum_{r=0}^{m-1}e^{2\pi i(g-a-b+c+d)r/m}\\=&m|\{(a,b,c,d)\in A_0^4:\ g\equiv a+b-c-d\ (\mathrm{mod}\ m)\}|.\end{aligned}$$

Let $\lambda=|A|/m\in(0,1]$, and set

$$R=\{r\in[0,m-1]:|S_A(r)|\geqslant\sqrt{\lambda}|A|\}$$

and

$$R'=\{r\in[0,m-1]:|S_A(r)|<\sqrt{\lambda}|A|\}.$$

As $S_A(0)=|A|\geqslant\sqrt{\lambda}|A|$, we have $0\in R$ and $|R'|<m$. Observe that

$$\begin{aligned}\left|\sum_{r\in R'}|S_A(r)|^4\chi_r(g)\right|\leqslant&\sum_{r\in R'}(\sqrt{\lambda}|A|)^2|S_A(r)|^2\\<&\lambda|A|^2\sum_{r=0}^{m-1}|S_A(r)|^2=\lambda|A|^2\sum_{r=0}^{m-1}\sum_{a,a'\in A_0}e^{2\pi i(a-a')r/m}\\=&\lambda|A|^2\sum_{a\in A_0}m=|A|^4.\end{aligned}$$

Thus $|A|^4+\Re(\sum_{r\in R'}|S_A(r)|^4\chi_r(g))>0$ since $\Re(z)\geqslant-|z|$ for any complex number z.

For $x\in\mathbb{R}$, clearly

$$\|x\|\leqslant\frac{1}{4}\iff\cos(2\pi x)\geqslant0\iff\Re(2\pi ix)\geqslant0.$$

If $r \in R$ and $g \in \mathbb{Z}$ then $\|gr/m\| \leqslant 1/4 \iff \Re(\chi_r(g)) \geqslant 0$. Let $r_1, \dots, r_n$ be all the elements of R with $r_1 = 0$, and let g be an integer with $g + m\mathbb{Z} \in B_m(r_1, \dots, r_n; 1/4)$. Then $\Re(\chi_{r_j}(g)) \geqslant 0$ for all $j = 1, \dots, n$. Therefore

$$\begin{aligned}
&\Re\Big(\sum_{r=0}^{m-1} |S_A(r)|^4 \chi_r(g)\Big) \\
&= \sum_{1<j\leqslant n} |S_A(r)|^4 \Re(\chi_{r_j}(g)) + |S_A(0)|^4 + \Re\Big(\sum_{r\in R'} |S_A(r)|^4 \chi_r(g)\Big) \\
&\geqslant |A|^4 + \Re\Big(\sum_{r\in R'} |S_A(r)|^4 \chi_r(g)\Big) > 0.
\end{aligned}$$

So $\sum_{r=0}^{m-1} |S_A(r)|^4 \chi_r(g) \neq 0$ and hence $g + m\mathbb{Z} \in 2A - 2A$.

Finally we observe that

$$n(\sqrt{\lambda}|A|)^2 \leqslant \sum_{r\in R} |S_A(r)|^2 \leqslant \sum_{r=0}^{m-1} |S_A(r)|^2 = \frac{|A|^2}{\lambda}$$

and so $n \leqslant \lambda^{-2} = m^2/|A|^2$. This concludes the proof. □

A Lemma obtained by Minkowski's Second Theorem. *Let $m \geqslant 2$ be an integer, and let $r_1, \dots, r_n \in [0, m-1]$ with $\gcd(r_1, \dots, r_n, m) = 1$. Then there is a proper n-dimensional AP $Q \subseteq \mathbb{Z}_m$ such that*

$$Q \subseteq B_m\left(r_1, \dots, r_n; \frac{1}{4}\right) \quad \text{and} \quad |Q| > \frac{m}{(4n)^n}.$$

Let G and H be abelian groups, and let $h \geqslant 2$ be an integer. Let $A \subseteq G$ and $B \subseteq H$. A map $\phi : A \to B$ is called a *Freiman h-homomorphism* if we have

$$\phi(a_1) + \cdots + \phi(a_h) = \phi(a'_1) + \cdots + \phi(a'_h)$$

whenever $a_1, a_1', \ldots, a_h, a_h' \in A$ and $a_1 + \cdots + a_h = a_1' + \cdots + a_h'$. If $\phi : A \to B$ is surjective and for any $a_1, a_1', \ldots, a_h, a_h' \in A$ we have

$$a_1+\cdots+a_h = a_1'+\cdots+a_h' \iff \phi(a_1)+\cdots+\phi(a_h) = \phi(a_1')+\cdots+\phi(a_h'),$$

then we say that A is Freiman h-isomorphic to B via the Freiman h-isomorphism ϕ. Note that in this case ϕ is also injective since

$$\begin{aligned}\phi(a) = \phi(a') \Rightarrow\ & \phi(a) + (h-1)\phi(a) = \phi(a') + (h-1)\phi(a)\\ & \Rightarrow a + (h-1)a = a' + (h-1)a \Rightarrow a = a'.\end{aligned}$$

When $\phi : A \to B$ is a Freiman h-isomorphism, it can be shown that A is a proper n-dimensional AP if and only if B is a proper n-dimensional AP. This is the reason why Freiman introduced the concept of Freiman h-isomorphism.

Ruzsa's Reduction Lemma. *Let A be a finite nonempty subset of $\mathbb{Z}$, and let $h \geqslant 2$ be an integer. Then, for any $m \geqslant |hA - hA|$, there is $A' \subseteq A$ with $|A'| \geqslant |A|/h$ such that A' is Freiman h-isomorphic to a subset of $\mathbb{Z}_m$.*

Proof. Let p be a prime greater than $\max hA - \min hA$. For each $d \in (hA - hA) \setminus \{0\}$, we have $p \nmid d$ and hence

$$|\{q \in [1, p-1] :\ m | \{dq\}_p\}| = |\{r \in [1, p-1] :\ m \mid r\}| \leqslant \frac{p-1}{m},$$

where $\{a\}_p$ refers to the least nonnegative integer of a mod p. Thus

$$\begin{aligned}&|\{q \in [1, p-1] :\ m | \{dq\}_p \text{ for some } d \in (hA - hA) \setminus \{0\}\}|\\ &\qquad < |hA - hA| \frac{p-1}{m} \leqslant p - 1\end{aligned}$$

and hence there exists $q \in [1, p-1]$ such that $m \nmid \{dq\}_p$ for every $d \in (hA - hA) \setminus \{0\}$.

Define $\phi : \mathbb{Z} \to \mathbb{Z}_m$ by $\phi(x) = \{qx\}_p + m\mathbb{Z}$. For $j = 1, \dots, h$ let

$$S_j = \left\{ x \in A : \ \frac{j-1}{h} p \leqslant \{qx\}_p < \frac{j}{h} p \right\}.$$

Clearly $\sum_{j=1}^{h} |S_j| = |\bigcup_{j=1}^{h} S_j| = |A|$, so for some $1 \leqslant j \leqslant h$ we have $|S_j| \geqslant |A|/h$. We denote this S_j by A'.

Let $a_1, \dots, a_h \in A'$. Since

$$\frac{j-1}{h} p \leqslant \{qa_1\}_p, \dots, \{qa_h\}_p < \frac{j}{h} p,$$

we have

$$\{qa_1\}_p + \cdots + \{qa_h\}_p = (j-1)p + \{q(a_1 + \cdots + a_h)\}_p$$

and hence

$$\phi(a_1) + \cdots + \phi(a_h) = (j-1)p + \{q(a_1 + \cdots + a_h)\}_p + m\mathbb{Z}.$$

If $a_1', \dots, a_h' \in A'$ and $\{q(a_1 + \cdots + a_h)\}_p \geqslant \{q(a_1' + \cdots + a_h')\}_p$, then

$$\begin{aligned} & \phi(a_1) + \cdots + \phi(a_h) = \phi(a_1') + \cdots + \phi(a_h') \\ \iff & \{q(a_1 + \cdots + a_h - a_1' - \cdots - a_h')\}_p \\ & \quad = \{q(a_1 + \cdots + a_h)\}_p - \{q(a_1' + \cdots + a_h')\}_p \in m\mathbb{Z} \\ \iff & a_1 + \cdots + a_h - a_1' - \cdots - a_h' = 0. \end{aligned}$$

Therefore the restriction of ϕ on A' is Freiman h-isomorphic to $\phi(A') \subseteq \mathbb{Z}_m$. $\square$

Now we are ready to establish Ruzsa's extension of Freiman's theorem.

Freiman-Ruzsa Theorem. *Let A and B be finite nonempty subsets of $\mathbb{Z}$ with $|A+B|\leqslant c|A|$ and $|A|^2/|B|^2=\lambda$. Then B is contained in an n-dimensional AP Q with*

$$n\leqslant 2^8c^{32}\lambda+c^{-11}(2^{10}c^{32}\lambda)^{2^8c^{32}\lambda}\quad \textit{and}\quad l(Q)\leqslant 2^nc^4|A|.$$

Proof. Note that $c\geqslant 1$ since $|A+B|\geqslant |A|$. By the Plünnecke-Ruzsa theorem, we have $|B-B|\leqslant c^2|A|$ and $|8B-8B|\leqslant c^{16}|A|$. Choose a prime $p\in(c^{16}|A|,2c^{16}|A|]$. As $p>c^{16}|A|\geqslant |8B-8B|$, by Ruzsa's reduction lemma, there is $B'\subseteq B$ with $|B'|\geqslant |B|/8$ which is Freiman 8-isomorphic to a subset S of $\mathbb{Z}_p$. By Bogolyubov's theorem, there are distinct $r_1,\ldots,r_{n_1}\in[0,p-1]$ with $r_1=0$ such that

$$n_1\leqslant\left(\frac{p}{|S|}\right)^2=\left(\frac{p}{|B'|}\right)^2\leqslant\left(\frac{2c^{16}|A|}{|B|/8}\right)^2=2^8c^{32}\lambda$$

and $B_p(r_1,\ldots,r_{n_1};1/4)\subseteq 2S-2S$. In view of the lemma obtained by Minkowski's second theorem, if $n_1>1$ then there is a proper n_1-dimensional AP $Q'\subseteq\mathbb{Z}_p$ such that

$$Q'\subseteq B_p\left(r_1,\ldots,r_{n_1};\frac{1}{4}\right)\subseteq 2S-2S\ \text{ and }\ |Q'|\geqslant\frac{p}{(4n_1)^{n_1}}>\frac{c^{16}|A|}{(4n_1)^{n_1}}.$$

Note that this is also valid when $n_1=1$.

Let $\phi:B'\to S$ be a Freiman 8-isomorphism. For $a,b,c,d,a',b',c',d'\in$

B', clearly

$$\begin{aligned}
&\phi(a)+\phi(b)-\phi(c)-\phi(d)=\phi(a')+\phi(b')-\phi(c')-\phi(d')\\
\iff &\phi(a)+\phi(b)+\phi(c')+\phi(d')+4\phi(a)\\
&\quad=\phi(a')+\phi(b')+\phi(c)+\phi(d)+4\phi(a)\\
\iff &a+b+c'+d'+4a=a'+b'+c+d+4a\\
\iff &a+b-c-d=a'+b'-c'-d'.
\end{aligned}$$

Thus we can introduce a well defined surjective map $\psi : 2B' - 2B' \to 2S - 2S$ by letting $\psi(a+b-c-d) = \phi(a)+\phi(b)-\phi(c)-\phi(d)$ for $a, b, c, d \in B'$. For $a_1, b_1, \ldots, a_4, b_4, a_1', b_1', \ldots, a_4', b_4' \in S$,

$$\begin{aligned}
&\psi(a_1+a_2-a_3-a_4)+\psi(b_1+b_2-b_3-b_4)\\
&\quad=\psi(a_1'+a_2'-a_3'-a_4')+\psi(b_1'+b_2'-b_3'-b_4')\\
\iff &\phi(a_1)+\phi(a_2)-\phi(a_3)-\phi(a_4)+\phi(b_1)+\phi(b_2)-\phi(b_3)-\phi(b_4)\\
&\quad=\phi(a_1')+\phi(a_2')-\phi(a_3')-\phi(a_4')+\phi(b_1')+\phi(b_2')-\phi(b_3')-\phi(b_4')\\
\iff &\phi(a_1)+\phi(a_2)+\phi(b_1)+\phi(b_2)+\phi(a_3')+\phi(a_4')+\phi(b_3')+\phi(b_4')\\
&\quad=\phi(a_1')+\phi(a_2')+\phi(b_1')+\phi(b_2')+\phi(a_3)+\phi(a_4)+\phi(b_3)+\phi(b_4)\\
\iff &a_1+a_2+b_1+b_2+a_3'+a_4'+b_3'+b_4'\\
&\quad=a_1'+a_2'+b_1'+b_2'+a_3+a_4+b_3+b_4\\
\iff &(a_1+a_2-a_3-a_4)+(b_1+b_2-b_3-b_4)\\
&\quad=(a_1'+a_2'-a_3'-a_4')+(b_1'+b_2'-b_3'-b_4').
\end{aligned}$$

Therefore ψ is a Freiman 2-isomorphism from $2B' - 2B'$ to $2S - 2S$,

and hence $Q_1=\psi^{-1}(Q')\subseteq 2B'-2B'\subseteq 2B-2B$ is also a proper n_1-dimensional AP with

$$l(Q_1)=|Q_1|=|Q'|>\frac{c^{16}|A|}{(4n_1)^{n_1}}\geqslant\frac{c^{16}|A|}{(2^{10}c^{32}\lambda)^{2^8c^{32}\lambda}}.$$

.

Let $B^*=\{b_1,\ldots,b_{n_2}\}$ be a maximal subset of B with $b_1+Q_1,\ldots,b_{n_2}+Q_1$ pairwise disjoint. Then

$$n_2|Q_1|=\sum_{i=1}^{n_2}|b_i+Q_1|=|B^*+Q_1|\leqslant|B+(2B-2B)|=|3B-2B|\leqslant c^5|A|$$

and hence

$$n_2\leqslant\frac{c^5|A|}{|Q_1|}<c^{-11}(2^{10}c^{32}\lambda)^{2^8c^{32}\lambda}.$$

For each $b\in B$, there is $1\leqslant i\leqslant n_2$ such that $(b+Q_1)\cap(b_i+Q_1)\neq\emptyset$ and hence $b\in b_i+Q_1-Q_1\subseteq B^*+Q_1-Q_1\subseteq Q_2+Q_1-Q_1$, where

$$Q_2=\{\delta_1b_1+\cdots+\delta_{n_2}b_{n_2}:\ \delta_1,\ldots,\delta_{n_2}\in\{0,1\}\}$$

is an n_2-dimensional AP with $l(Q_2)=2^{n_2}$. Note that Q_1-Q_1 is an n_1-dimensional AP with

$$l(Q_1-Q_1)<2^{n_1}l(Q_1)=2^{n_1}|Q_1|\leqslant 2^{n_1}|2B-2B|\leqslant 2^{n_1}c^4|A|.$$

Thus B is contained in an n-dimensional AP $Q=Q_2+(Q_1-Q_1)$ with

$$n=n_1+n_2\leqslant 2^8c^{32}\lambda+c^{-11}(2^{10}c^{32}\lambda)^{2^8c^{32}\lambda}$$

and

$$|Q|\leqslant l(Q)=l(Q_1-Q_1)l(Q_2)\leqslant 2^{n_1}c^4|A|2^{n_2}=2^nc^4|A|.$$

This concludes the proof. □

3. Triangle Removal Lemma and the Balog-Szemerédi-Gowers theorem

The most important technique used by Szemerédi in his combinatorial proof of Szemerédi's theorem is his powerful regularity lemma in graph-theoretic language.

Let $G = (V, E)$ be an undirected graph (without multiple edges). For $A, B \subseteq V$ we define

$$e(A, B) = |E \cap (A \times B)| \quad \text{and} \quad d(A, B) = \frac{e(A, B)}{|A \times B|},$$

and call $d(A, B)$ the *density of edges* between A and B. For $\varepsilon > 0$ the pair (A, B) is said to be *ε-regular* if $|d(X, Y) - d(A, B)| < \varepsilon$ for all those $X \subseteq A$ and $Y \subseteq B$ with $|X| \geqslant \varepsilon|A|$ and $|Y| \geqslant \varepsilon|B|$.

Szemerédi's Regularity Lemma. *Let $0 < \varepsilon < 1$ and $m_0 \in \mathbb{Z}^+$. Then there are positive integers $M = M(\varepsilon, m_0)$ and $N = N(\varepsilon, m_0)$ such that whenever $G = (V, E)$ is an undirected graph with $|V| \geqslant N$ there is a partition $V_0 \cup V_1 \cup \cdots \cup V_m$ of V with*

$$|V_0| \leqslant \varepsilon|V|, \quad |V_1| = \cdots = |V_m|, \quad m_0 \leqslant m \leqslant M \tag{$\star$}$$

and at most εm^2 pairs (V_i, V_j) $(1 \leqslant i < j \leqslant m)$ not ε-regular.

Triangle Removal Lemma (Ruzsa and Szemerédi, 1978)**.** *For each $0 < \delta \leqslant 1$, there exists $0 < c(\delta) < 1$ with the following property: If $G = (V, E)$ is an undirected graph with $|V|$ sufficiently large, and G contains fewer than $c(\delta)|V|^3$ triangles and then it is possible to remove fewer than $\delta|V|^2$ edges from G to create a graph containing no triangles.*

Proof. Let $\varepsilon=\delta/12$ and set $m_0=\lfloor 12/\delta\rfloor+1>1/\varepsilon$. By Szemerédi's regularity lemma, there are positive integers $M=M(\varepsilon,m_0)$ and $N=N(\varepsilon,m_0)$ such that when $n=|V|\geqslant N$ there is a partition $V_0\cup V_1\cup\cdots\cup V_m$ for which $(\star)$ holds and there are at most εm^2 not ε-regular pairs (V_i,V_j) with $1\leqslant i<j\leqslant m$.

Suppose that $|V|\geqslant N$ and let $V_0\cup V_1\cup\cdots\cup V_m$ be a partition of V as described above. Now we delete certain edges by the following rules:

(i) Delete those edges in E with one endpoint in V_0. Since the degree of each vertex is at most $|V|$, the number of edges we delete in this step is at most $|V_0|\cdot|V|\leqslant\varepsilon|V|^2$.

(ii) Delete those edges with two endpoints in the same V_i with $1\leqslant i\leqslant m$. The number of edges we delete in this step is at most

$$\sum_{i=1}^m|V_i|^2\leqslant m\left(\frac{|V|}{m}\right)^2\leqslant\frac{|V|^2}{m_0}<\varepsilon|V|^2.$$

(iii) Delete those edges with one endpoint in V_i and another endpoint in V_j, where $1\leqslant i<j\leqslant m$ and (V_i,V_j) is not ε-regular. As $e(V_i,V_j)\leqslant|V_i|\cdot|V_j|\leqslant(|V|/m)^2$ and there are at most εm^2 not ε-regular pairs, the number of edges we delete in this step is at most $\varepsilon m^2(|V|/m)^2=\varepsilon|V|^2$.

(iv) If (V_i,V_j) is ε-regular with $1\leqslant i<j\leqslant m$ but $d(V_i,V_j)\leqslant 3\varepsilon/2$, then delete those edges with one endpoint in V_i and another endpoint in V_j. The number of edges we delete in this step is at most

$$\binom{m}{2}\frac{3}{2}\varepsilon\frac{|V|}{m}\cdot\frac{|V|}{m}<\frac{3}{4}\varepsilon|V|^2.$$

Let E' denote the set of those edges left after the above four steps. Then $|E\setminus E'|<(3+3/4)\varepsilon|V|^2<4\varepsilon|V|^2$.

Now assume that any graph formulated by removing fewer than $\delta|V|^2$ edges from G contains triangles. Let l be the maximal number of pairwise edge-disjoint triangles contained in G, and let $T_1, \dots, T_l$ be pairwise disjoint triangles in G. If we remove all edges of $T_1, \dots, T_l$ from G, then the induced graph contains no triangles since each triangle in G shares an edge with some T_i with $1 \leqslant i \leqslant l$. Therefore $3l \geqslant \delta|V|^2$. As

$$|E \setminus E'| < 4\varepsilon|V|^2 = \frac{\delta}{3}|V|^2 \leqslant l,$$

the graph $G' = (V, E')$ must contain a triangle T.

Suppose that the three vertices of the triangle T lie in V_i, V_j and V_k respectively, where $1 \leqslant i < j < k \leqslant m$. By the formulation of E', the pairs $(V_i, V_j), (V_i, V_k), (V_j, V_k)$ are ε-regular and $d(V_i, V_j), d(V_i, V_k), d(V_j, V_k)$ are all greater than $3\varepsilon/2$.

Let

$$V_i^{(j)} = \{v \in V_i : \ d(\{v\}, V_j) \leqslant d(V_i, V_j) - \varepsilon\}.$$

If $|V_i^{(j)}| \geqslant \varepsilon|V_i|$, then $|d(V_i^{(j)}, V_j) - d(V_i, V_j)| < \varepsilon$ since (V_i, V_j) is ε-regular, hence

$$d(V_i, V_j) - \varepsilon < d(V_i^{(j)}, V_j) = \frac{1}{|V_i^{(j)}|} \sum_{v \in V_i^{(j)}} d(\{v\}, V_j) \leqslant d(V_i, V_j) - \varepsilon$$

which leads a contradiction. Thus $|V_i^{(j)}| < \varepsilon|V_i|$. Similarly, $|V_i^{(k)}| < \varepsilon|V_i|$, where

$$V_i^{(k)} = \{v \in V_i : \ d(\{v\}, V_k) \leqslant d(V_i, V_k) - \varepsilon\}.$$

For $U = V_i \setminus (V_i^{(j)} \cup V_i^{(k)})$, we have

$$|U| \geqslant |V_i| - |V_i^{(j)}| - |V_i^{(k)}| > |V_i| - \varepsilon|V_i| - \varepsilon|V_i| = (1 - 2\varepsilon)|V_i|.$$

Let $u \in U$. Then $d(\{u\}, V_j) > d(V_i, V_j) - \varepsilon \geqslant 3\varepsilon/2 - \varepsilon = \varepsilon/2$ and $d(\{u\}, V_k) > d(V_i, V_k) - \varepsilon \geqslant \varepsilon/2$. Thus $|\Gamma_j(u)| > \varepsilon|V_j|$ and $|\Gamma_k(u)| > \varepsilon|V_k|$, where

$$\Gamma_j(u) = \{v \in V_j : (u, v) \in E'\} \text{ and } \Gamma_k(u) = \{v \in V_k : (u, v) \in E'\}.$$

As (V_j, V_k) is ε-regular, we have

$$|d(\Gamma_j(u), \Gamma_k(u)) - d(V_j, V_k)| < \varepsilon$$

and hence

$$e(\Gamma_j(u), \Gamma_k(u)) > (d(V_j, V_k) - \varepsilon)|\Gamma_j(u)| \cdot |\Gamma_k(u)| \geqslant \frac{\varepsilon}{2}\varepsilon|V_j|\varepsilon|V_k|.$$

Note that $e(\Gamma_j(u), \Gamma_k(u))$ is the number of triangles in G with u as a vertex and the other two vertices in V_j and V_k respectively.

By the above, the number of triangles with vertices in V_i, V_j, V_k respectively is at least

$$\sum_{u \in U} e(\Gamma_j(u), \Gamma_k(u)) \geqslant |U|\frac{\varepsilon}{2}\varepsilon|V_j|\varepsilon|V_k| \geqslant \frac{\varepsilon^3}{2}(1 - 2\varepsilon)|V_i||V_j||V_k|.$$

Recall that $|V_i| = |V_j| = |V_k| = (|V| - |V_0|)/m \geqslant (1 - \varepsilon)|V|/M$. So G contains at least $c(\delta)|V|^3$ triangles, where

$$c(\delta) = \frac{\varepsilon^3(1 - \varepsilon)^3(1 - 2\varepsilon)}{2M^3} > 0$$

only depends on δ. This concludes the proof. $\square$

Roth's Theorem. *Let $0 < \delta < 1$. If $n \in \mathbb{Z}^+$ is sufficiently large, then any subset A of $[1,n]$ with $|A| \geqslant \delta n$ contains an AP of length three.*

Proof. For each $a \in [1,n]$, call $\{(a,1),\dots,(a,n)\}$ a vertical line, and $\{(1,a),\dots,(n,a)$ a horizontal line. For $m \in [2,2n]$ call $\{(a,b) \in [1,n]^2 : a+b=m\}$ a skew line. Construct a graph $G=(V,E)$ whose vertices are these $4n-1$ lines and whose edges are just those (L_1,L_2) with $L_1 \neq L_2$ and $L_1 \cap L_2 \subseteq X$, where

$$X = \{(a,b) \in [1,n]^2 : \ a+2b \in A\}.$$

Clearly $L_1 \cap L_2 = \emptyset$ if L_1 and L_2 are lines of the same kind. If L_1 and L_2 are lines of different kinds, then there is a unique intersection point of L_1 and L_2. Thus a triangle in G is formed by three lines of three different kinds, and so the intersection points of these three lines can be written as $(a,b),(a+d,b),(a,b+d) \in X$.

For each $v=(a,b) \in X$, we let T_v denote the triangle in G whose vertices are three different lines passing v. Clearly those T_v with $v \in X$ are pairwise edge-disjoint. Observe that

$$\begin{aligned}
|X| &= \sum_{c \in A} |\{(a,b) \in [1,n]^2 : \ a = c-2b\}| \\
&\geqslant \sum_{\substack{c \in A \\ c > \delta n/2}} \left(\left\lfloor \frac{c}{2} \right\rfloor - 1 \right) \geqslant \left(\frac{\delta n}{4} - 2 \right) \left| \left\{ c \in A : \ c > \frac{\delta n}{2} \right\} \right| \\
&\geqslant \left(\frac{\delta n}{4} - 2 \right) \frac{\delta n}{2}.
\end{aligned}$$

If $n \geqslant 10/\delta$, then $2 \leqslant \delta n/5$ and hence

$$|X| \geqslant \left(\frac{\delta n}{4} - \frac{\delta n}{5} \right) \frac{\delta n}{2} > \frac{\delta^2}{40} \left(n - \frac{1}{4} \right)^2 = \delta' |V|^2$$

where $\delta'=\delta^2/640$. If we remove fewer than $\delta'|V|^2<|X|$ edges from G, then there is a triangle T_v with $v\in X$ left in the resulting graph. By the Triangle Removal Lemma, if $|V|=4n-1$ is sufficiently large then the graph G contains at least $c|V|^3>n^2\geqslant |X|$ triangles, where $c>0$ only depends on δ.

Assume that $|V|$ is sufficiently large. By the above, G must contain a triangle T different from those T_v with $v\in X$. Let $(a,b),(a+d,b),(a,b+d)\in X$ be the three intersection points of the three lines used as the vertices of T. Then $d\neq 0$, and also

$$a+2b\in A,\ a+d+2b\in A,\ a+2(b+d)=a+2b+2d\in A.$$

So A contains an AP of length three with common difference d. □

To obtain the general case of Szemerédi's theorem in the above spirit, one should extend Szemerédi's Regularity Lemma and the Triangle Removal Lemma to hypergraphs. We mention that there are several different versions of them for hypergraphs. Contributors in this direction include Erdős, Rödl, Skokan, Gowers, Chung, Graham, Tao, and Ishigami. We will discuss this in future lectures.

The initial form of the following result was first obtained by Balog and Szemerédi in 1994 via Szemerédi's regularity lemma.

Balog-Szemerédi-Gowers Theorem. *Let A and B be finite nonempty subsets of an abelian group. Let $E\subseteq A\times B$ with*

$$|E|\geqslant\frac{|A||B|}{K}\quad\text{and}\quad |A\overset{E}{+}B|\leqslant K'\sqrt{|A||B|},$$

where $K \geqslant 1$, $K' > 0$ *and* $A \overset{E}{+} B = \{a+b : (a,b) \in E\}$. *Then there are* $A' \subseteq A$ *and* $B' \subseteq B$ *such that*

$$|A'| \geqslant \frac{|A|}{4\sqrt{2}K},\ |B'| \geqslant \frac{|B|}{4K},\ \text{and } |A'+B'| \leqslant 2^{12}K^5(K')^3\sqrt{|A||B|}.$$

For convenience, if X is a finite set and $f(x) \in \mathbb{C}$ for all $x \in X$, then we define

$$\mathbb{E}_{x\in X} f(x) := \frac{1}{|X|}\sum_{x\in X} f(x).$$

For a predicate P we let

$$[\![P]\!] = \begin{cases} 1 & \text{if } P \text{ holds,} \\ 0 & \text{otherwise.} \end{cases}$$

A Lemma on Paths of Length Two. *Let* $G = (V,E)$ *be a bipartite graph with vertex classes* A *and* B. *Suppose that* $|E| \geqslant |A|\,|B|/K$ *with* $K \geqslant 1$. *Then for each* $0 < \varepsilon < 1$ *there is* $A' \subseteq A$ *with* $|A'| \geqslant |A|/(\sqrt{2}K)$ *such that*

$$\left|\left\{(a,a') \in A' \times A' :\ |\{b \in B : (a,b),(a',b) \in E\}| < \frac{\varepsilon}{2K^2}|B|\right\}\right| < \varepsilon|A'|^2.$$

Proof. For $b \in B$, the neighborhood of b is $N(b) = \{a \in A : (a,b) \in E\}$. Observe that

$$\mathbb{E}_{b\in B}|N(b)| = \frac{|E|}{|B|} \geqslant \frac{|A|}{K}$$

and hence by the Cauchy-Schwarz inequality we have

$$\mathbb{E}_{b\in B}(|N(b)|^2) \geqslant (\mathbb{E}_{b\in B}|N(b)|)^2 \geqslant \frac{|A|^2}{K^2}.$$

Let

$$\Omega=\left\{(a,a')\in A\times A:\ |N(a)\cap N(a')|<\frac{\varepsilon}{2K^2}|B|\right\}.$$

Clearly,

$$\sum_{b\in B}\sum_{a,a'\in N(b)}[\![(a,a')\in\Omega]\!]=\sum_{(a,a')\in\Omega}|N(a)\cap N(a')|<|\Omega|\frac{\varepsilon}{2K^2}|B|\leqslant\frac{\varepsilon|A|^2B}{2K^2}$$

and so

$$\begin{aligned}&\mathbb{E}_{b\in B}\sum_{a,a'\in N(b)}\left(1-\frac{1}{\varepsilon}[\![(a,a')\in\Omega]\!]\right)\\=&\mathbb{E}_{b\in B}|N(b)|^2-\frac{1}{\varepsilon}\mathbb{E}_{b\in B}\sum_{a,a'\in N(b)}[\![(a,a')\in\Omega]\!]\\\geqslant&\frac{|A|^2}{K^2}-\frac{1}{\varepsilon}\cdot\frac{\varepsilon|A|^2}{2K^2}=\frac{|A|^2}{2K^2}.\end{aligned}$$

Thus, for some $b\in B$ we have

$$\begin{aligned}\frac{|A|^2}{2K^2}&\leqslant\sum_{a,a'\in N(b)}\left(1-\frac{[\![(a,a')\in\Omega]\!]}{\varepsilon}\right)\\&=|N(b)|^2-\frac{1}{\varepsilon}|\{(a,a')\in\Omega:\ a,a'\in N(b)\}|.\end{aligned}$$

Set $A'=N(b)\subseteq A$. Then $|A'|\geqslant|A|/(\sqrt{2}K)$ and

$$|\{(a,a')\in\Omega:\ a,a'\in A'\}|<\varepsilon|A'|^2.$$

This concludes the proof. □

Proof of the Balog-Szemerédi-Gowers Theorem. Let G be a bipartite graph with vertex classes A and B, and with the edge set E. Let

$A_* = \{a \in A : |N(a)| \geqslant |B|/(2K)\}$ and G_* be the induced bipartite graph with classes A_* and B. Clearly $e(A \setminus A_*, B) \leqslant |A||B|/(2K)$ and thus

$$|E(G_*)| \geqslant |E| - \frac{|A||B|}{2K} \geqslant \frac{|A||B|}{2K}.$$

Note that $L = |A|/|A_*| \geqslant 1$ and $|E(G_*)| \geqslant |A_*||B|/(2K/L)$.

Applying the above lemma to the graph G_*, we know that there is $A'_* \subseteq A_*$ such that

$$|A'_*| \geqslant \frac{|A_*|}{\sqrt{2}(2K/L)} = \frac{|A|}{2\sqrt{2}K}$$

and

$$\left| \left\{ (a, a') \in A'_* \times A'_* : |N(a) \cap N(a')| < \frac{1/(16K)}{2(2K/L)^2}|B| \right\} \right| < \frac{|A'_*|^2}{16K}.$$

Thus there are fewer than $|A'_*|^2/(16K)$ pairs $(a, a') \in A'_* \times A'_*$ which are *bad* in the sense that $|N(a) \cap N(a')| < |B|L^2/(128K^3)$.

Let

$$A' = \left\{ a \in A'_* : |\{a' \in A'_* : (a, a') \text{ is bad}\}| < \frac{|A'_*|}{8K} \right\}.$$

If $|A'_* \setminus A'| \geqslant |A'_*|/2$, then

$$\frac{|A'_*|^2}{16K} > |\{(a, a') \in (A'_* \setminus A') \times A'_* : \ (a, a') \text{ is bad}\}| \geqslant \frac{|A'_*|}{2} \cdot \frac{|A'_*|}{8K}$$

which leads a contradiction. So $|A'| \geqslant |A'_*|/2 \geqslant |A|/(4\sqrt{2}K)$. Set

$$B' = \left\{ b \in B : |\{a \in A'_* : (a, b) \in E\}| \geqslant \frac{|A'_*|}{4K} \right\}.$$

Then

$$\begin{aligned}|A'_*||B'|&\geqslant\sum_{b\in B'}|\{a\in A'_*:(a,b)\in E\}|\\&=\sum_{b\in B}|\{a\in A'_*:(a,b)\in E\}|-\sum_{b\in B\setminus B'}|\{a\in A'_*:(a,b)\in E\}|\\&>|A'_*|\frac{|B|}{2K}-\sum_{b\in B\setminus B'}\frac{|A'_*|}{4K}\geqslant\frac{|A'_*||B|}{4K}\end{aligned}$$

and hence $|B'|\geqslant|B|/(4K)$.

Let $a\in A'$ and $b\in B'$. Then

$$|\{a'\in A'_*:(a',b)\in E\}|\geqslant\frac{|A'_*|}{4K}\text{ and }|\{a'\in A'_*:(a,a')\text{ is bad}\}|<\frac{|A'_*|}{8K}.$$

Therefore

$$\begin{aligned}&\left|\left\{a\in A'_*:(a',n)\in E\text{ and }|N(a)\cap N(a')|\geqslant\frac{L^2|B|}{128K^3}\right\}\right|\\=&|\{a\in A'_*:(a',n)\in E\text{ and }(a,a')\text{ is not bad}\}|\\\geqslant&\frac{|A'_*|}{4K}-\frac{|A'_*|}{8K}=\frac{|A'_*|}{8K}\geqslant\frac{|A|}{16\sqrt{2}K^2}\end{aligned}$$

and hence

$$\begin{aligned}&|\{(a',b')\in A\times B:(a,b'),(a',b'),(a',b)\in E\}|\\&\geqslant\frac{|A|}{16\sqrt{2}K^2}\cdot\frac{L^2|B|}{2^7K^3}\geqslant\frac{|A||B|}{2^{12}K^5}.\end{aligned}$$

Note that $(a+b')-(a'+b')+(a'+b)=a+b$. So

$$|\{(x,y,z):x,y,z\in A\overset{E}{+}B,\ x-y+z=a+b\}|\geqslant\frac{|A||B|}{2^{12}K^5}.$$

By the above,

$$\begin{aligned}|A'+B'|\frac{|A||B|}{2^{12}K^5}\leqslant&|\{(x,y,z):x,y,z\in A\overset{E}{+}B,\ x-y+z\in A'+B'\}|\\\leqslant&|A\overset{E}{+}B|^3\leqslant(K'\sqrt{|A||B|})^3\end{aligned}$$

and it follows that $|A'+|B'|\leqslant 2^{12}K^5(K')^3\sqrt{|A||B|}$. We are done. □

References

[B] B. J. Green, *Structure theory of set addition*, Lecture notes given at Edinburgh in 2002.

[F] G. A. Freiman, *Foundations of a Structural Theory of Set Addition*, Translations of Math. Monographs, Vol. 37, Amer. Math. Soc., Providence, R.I., 1973.

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

[N] M. B. Nathanson, *Additive Number Theory: Inverse Problems and the Geometry of Sumsets* (Graduate texts in math.; 165), Springer, New York, 1996.

[S] Z. W. Sun, *Hall's theorem revisited*, Proc. Amer. Math. Soc. **129** (2001), 3129–3131.

[TV] T. Tao and V. H. Vu, *Additive Combinatorics*, Cambridge Univ. Press, Cambridge, 2006.