Arithmetic Ramsey Theory

Zdeněk Dvořák

January 4, 2024

Recall:

Theorem 1 (Triangle Removal lemma). *For every $0 < \alpha \leq 1$, there exists $\beta > 0$ and n_0 such that if G is a graph with $n \geq n_0$ vertices, then either*

- *G contains at least βn^3 triangles, or*
- *there exists a set $X \subseteq V(G)$ such that $|X| \leq \alpha n^2$ and $G - X$ contains no trianges.*

1 Arithmetic progressions in dense sets

Removal lemma has many interesting applications, let us give one in arithmetic Ramsey theory. A well-known theorem of Van der Waerden states that in any coloring of integers by finitely many colors, there exist arbitrarily long monochromatic arithmetic sequences. Actually, a stronger claim holds: if we select any subset of integers of positive density, such a subset contains arbitrarily long arithmetic sequences (which implies Van der Waerden's theorem, since if we color integers by k colors, at least one of the color classes has density at least $1/k$).

Theorem 2 (Szemerédi). *For every $\gamma > 0$ and positive integer k, there exists n_2 as follows. If $n \geq n_2$ and $B \subseteq \{1, \ldots, n\}$ has size at least γn, then there exist integers $b, d > 0$ such that $b, b + d, \ldots, b + kd \in B$.*

Here, we show just a restricted subcase of this claim, for arithmetic sequences of length 3. Firstly, we will need a "geometric" statement.

Lemma 3. *For every $\delta > 0$, there exists n_1 as follows. If $n \geq n_1$ and $A \subseteq \{1, \ldots, n\}^2$ has size at least δn^2, then there exist $x, y \in \{1, \ldots, n\}$ and $d \neq 0$ such that $(x, y), (x, y + d), (x + d, y) \in A$.*

Proof. Let $\alpha = \delta/40$. Let $\beta > 0$ and n_0 be the corresponding constants from the Removal lemma. Let $n_1 = \lceil \max(n_0/6, 1/(54\beta)) \rceil$.

Let G be the graph with vertex set $\{r_i, s_i, t_i : 1 \le i \le 2n\}$, and with edges defined as follows:

- $r_x s_y$ is an edge if $(x, y) \in A$,
- $r_x t_z$ is an edge if $(x, z - x) \in A$, and
- $s_y t_z$ is an edge if $(z - y, y) \in A$.

Hence, $r_x s_y t_z$ is a triangle if and only if $(x, y), (x, z-x), (z-y, y) \in A$. Letting $d = z - x - y$, such a triangle gives the required solution unless $d = 0$, i.e., $z = x + y$. Note that there are less than $(2n)^2$ triples $x, y, z \in \{1, \ldots, 2n\}$ such that $x + y = z$. Therefore, if G contains at least $(2n)^2$ triangles, then the lemma holds.

Suppose that G contains less than $(2n)^2 \le \frac{1}{54n_1}(6n)^3 \le \beta |V(G)|^3$ triangles. By the Removal lemma, there exists a set $X \subseteq E(G)$ of size at most $\alpha |V(G)|^2 = 36\alpha n^2$ such that $G - X$ is triangle-free. Let $T = \{r_x s_y t_{x+y} : (x, y) \in A\}$. Note that T is a set of $|A|$ pairwise edge-disjoint triangles in G. Since $G - X$ is triangle-free, X contains an edge in each of the triangles of T, and thus $36\alpha n^2 \ge |X| \ge |T| = |A|$. This is a contradiction, since $|A| \ge \delta n^2 > 36\alpha n^2$. $\square$

Corollary 4 (Roth's theorem)**.** *For every $\gamma > 0$, there exists n_2 as follows. If $n \ge n_2$ and $B \subseteq \{1, \ldots, n\}$ has size at least γn, then there exist integers $b, d > 0$ such that $b, b + d, b + 2d \in B$.*

Proof. Let $\delta = \gamma/4$. Let n_1 be the corresponding constant from Lemma 3. Let $n_2 = \lceil n_1/2 \rceil$.

Let $A \subseteq \{1, ..., 2n\}^2$ such that $(x, y) \in A$ if and only if $y - x \in B$. Note that if $b \in B$, then $(1, b+1), (2, b+2), \ldots, (n, b+n) \in A$, hence $|A| \ge |B|n \ge \gamma n^2 = (\gamma/4)(2n)^2 = \delta(2n)^2$. By Lemma 3, there exist x, y and $d \ne 0$ such that $(x, y), (x, y+d), (x+d, y) \in A$. Hence $(y-x)-d, y-x, (y-x)+d \in B$, giving the arithmetic progression as required. $\square$

How to modify the argument to prove the full version of Szemerédi's theorem? It suffices to generalize Lemma 3 to higher dimension, i.e., show that a dense subset A of $\{1, \ldots, n\}^k$ contains a point $(x_1, x_2, \ldots, x_k)$ such that $(x_1 + d, x_2, \ldots, x_k), (x_1, x_2 + d, \ldots, x_k), \ldots, (x_1, x_2, \ldots, x_k + d) \in A$. Proving this requires a variant of Removal lemma for hypergraphs: If H is a k-uniform hypergraph, then either H contains $\Omega(|V(H)|^{k+1})$ copies of the complete k-uniform hypergraph on $k + 1$ vertices, or all the copies can be destroyed by removing $o(|V(H)|^k)$ hyperedges.

Unfortunately, a straightforward generalization of the Regularity lemma to the hypergraph setting is too weak to prove the Removal lemma. The problem is that it is not sufficient to control the density of hyperedges: let G be a random graph with edge density $1/2$, and let H be the 3-uniform hypergraph where $uvw \in E(H)$ if and only if $uv, uw, vw \in E(G)$. It is easy to check H is "random-like" in the sense of Regularity lemma, and it has hyperedge density $1/8$. So, we would expect the density of the complete 3-uniform hypergraph on 4 vertices (which has 4 hyperedges) in H to be $(1/8)^4 = 1/4096$. However, actually the complete 3-uniform hypergraphs on 4 vertices in H correspond exactly to the subgraphs of K_4 in G, which have density $(1/2)^6 = 1/64$. Hence, a much more involved version of hypergraph Regularity lemma (which is beyond the scope of this lecture) is needed to prove hypergraph Removal lemma.

Szemerédi's Theorem probably is not the best possible. Erdös conjectured the following much stronger statement.

Conjecture 1. *Any infinite set B of integers such that*

$$\sum_{n \in B} \frac{1}{n} = \infty$$

contains arbitrarily long arithmetic progressions.

Recently, Green and Tao proved an interesting special case of this conjecture: primes contain arbitrarily long arithmetic progressions. From the other side, we know that there are quite dense subsets without arithmetic progressions.

Theorem 5. *For every N_0, there exists $N \geq N_0$ and a subset $B \subseteq \{0, \ldots, N-1\}$ such that*

$$|B| \geq \frac{N}{16^{\sqrt{\log_2 N}}}$$

and B does not contain a 3-term arithmetic progression.

Proof. For a positive integer n, let $d = 2^{n-1}$ and $N = (2d)^n$. Note that $\log_2 N = n(1 + \log_2 d) = n^2$.

For $0 \leq k \leq n(d-1)^2$, let

$$S_k = \{\vec{x} \in \{0, \ldots, d-1\}^n : \|\vec{x}\| = \sqrt{k}\}.$$

Since S_k is a subset of the sphere of radius k, S_k does not contain a 3-term arithmetic progression, i.e., for all distinct $\vec{x}, \vec{z} \in S_k$ we have $\frac{\vec{x}+\vec{z}}{2} \notin S_k$.

Moreover, we have $\{0, \ldots, d-1\}^n = \bigcup_{k=0}^{n(d-1)^2} S_k$, and thus there exists k such that $|S_k| \geq \frac{d^n}{n(d-1)^2+1} > d^{n-2}/n$. Let us fix this value of k.

For $\vec{x} = (x_1, \ldots, x_n) \in \{0, \ldots, 2d-1\}^n$, let us define $f(\vec{x}) = \sum_{i=1}^{n} (2d)^{i-1} x_i$; i.e., we view $\vec{x}$ as a number in the base $2d$. Clearly $f(\vec{x}) \in \{0, \ldots, N-1\}$. Furthermore, f is a bijection and for $\vec{x}, \vec{y}, \vec{z} \in \{0, \ldots, d-1\}^n$, we have $f(\vec{x}) + f(\vec{z}) = f(\vec{x} + \vec{z})$ and $2f(\vec{y}) = f(2\vec{y})$. Since S_k does not contain 3-term arithmetic progression, it follows that $f(S_k)$ does not contain a 3-term arithmetic progression. For $B = f(S_k)$, we have

$$\frac{|B|}{N} = \frac{|S_k|}{N} > \frac{d^{n-2}/n}{(2d)^n} = \frac{1}{n 2^n d^2} \geq \frac{1}{4^{n+\log_2 d}} > \frac{1}{4^{2n}} = \frac{1}{16^{\sqrt{\log_2 N}}}.$$

□

2 Hales-Jewett Theorem

Another interesting generalization of Van der Waerden's theorem essentially considers higher-dimensional arithmetic progressions.

We will consider subsets of the d-dimensional cube $\{1, \ldots, n\}^d$. Any sequence S of stars and integers $\{1, \ldots, n\}$ of length d that contains at least one star is called a *root*. For $i = 1, \ldots, n$, let $S(i)$ denote the sequence obtained from S by replacing the stars by i.

The *combinatorial line* $L(S)$ described by S is the set $\{S(1), S(2), \ldots, S(n)\}$. Example for $d = 2$ and $n = 4$:

- $L(\star, 2) = \{(1, 2), (2, 2), (3, 2), (4, 2)\}$ is the second row of the 4×4 matrix.
- $L(3, \star) = \{(3, 1), (3, 2), (3, 3), (3, 4)\}$ is the third column of the 4×4 matrix.
- $L(\star, \star) = \{(1, 1), (2, 2), (3, 3), (4, 4)\}$ is the main diagonal of the 4×4 matrix.

Theorem 6 (Hales-Jewett)**.** . *For every n, k, there exists D such that every k-coloring of the D-dimensional cube $\{1, \ldots, n\}^D$ contains a monochromatic combinatorial line.*

Informally, if we play n-in-a-row in sufficiently large dimension, then it is impossible to draw. Firstly, let us show that Hales-Jewett theorem implies Van der Waerden's theorem.

Proof of Van der Waerden's theorem. Consider any k-coloring of the integers, and suppose that we are looking for an arithmetic progression of length n. Let D be the dimension from Theorem 6. For $X = (x_1, \ldots, x_D) \in \{1, \ldots, n\}^D$, let $f(X) = x_1 + \ldots + x_D$ and assign to X the color of $f(X)$. By Theorem 6, there exists a monochromatic combinatorial line; let S be its root. Then $f(S(1)), f(S(2)), \ldots, f(S(n))$ is a monochromatic arithmetic progression (with step equal to the number of stars in S). □

Note that a density version of Hales-Jewett theorem (where we do not give a k-coloring, but just a set containing at least one k-th of the elements of the cube) is also true, and implies Szemeredi's theorem in the same way.

Hales-Jewett theorem is much stronger than Van der Waerden's theorem. For example, it implies the following generalization showing that the arithmetic sequences are not really that special.

Theorem 7 (Gallai-Witt)**.** *For any finite set $T \subseteq \mathbf{N}^t$ and for any coloring of $\mathbf{N}^t$ by k colors, there exist $a \in \mathbf{N}^t$ and a positive integer d such that the set $\{a + dt : t \in T\}$ is monochromatic.*

Proof. Let $T = \{t_1, \ldots, t_n\}$ and let D be the dimension from Theorem 6. For $X = (x_1, \ldots, x_D) \in \{1, \ldots, n\}^D$, let $f(X) = t_{x_1} + \ldots + t_{x_D}$, and assign to X the color of $f(X)$. By Theorem 6, there exists a monochromatic combinatorial line; let $S = (s_1, \ldots, s_D)$ be its root. Let $I \subseteq \{1, \ldots, D\}$ be the set of indices on that S has stars, and $J = \{1, \ldots, D\} \setminus I$. Let

$$a = \sum_{j \in J} t_{s_j}$$
$$d = |I|$$

Then $\{f(S(1)), f(S(2)), \ldots, f(S(n))\} = \{a + dt : t \in T\}$ is monochromatic. □

Note that with $T = \{1, \ldots, n\}$, Gallai-Witt theorem becomes the Van der Waerden's theorem. Let us now prove Hales-Jewett Theorem. Let us start with a few definitions.

Suppose that $a = (a_1, \ldots, a_d)$ is an element of $\{1, \ldots, n\}^d$ and for $i = 1, \ldots, d$, let S_i be a root of length D_i. Let S be the concatenation of $S_1, \ldots, S_d$. By $S(a)$, we mean the concatenation of $S_1(a_1)$, $S_2(a_2)$, …, $S_d(a_d)$.

Two elements $(a_1, \ldots, a_d), (b_1, \ldots, b_d) \in \{1, \ldots, n\}^d$ are *adjacent* if there exists $r \in \{1, \ldots, d\}$ such that $a_i = b_i$ for $i \neq r$ and $\{a_r, b_r\} \subseteq \{n-1, n\}$.

Proof of Theorem 6. We proceed by induction on n. For $n = 1$, the claim holds trivially. Hence, assume that $n > 1$ and that the claim holds for $n - 1$

(with the same number of colors), and let d be the corresponding bound on the dimension. Let

$$D_m = k^{n^{d+D_1+\ldots+D_{m-1}}}$$

for $m = 1, \ldots, d$, and let $D = D_1 + \ldots + D_d$.

Let φ be any k-coloring of $\{1, \ldots, n\}^D$. We need to prove that it contains a monochromatic combinatorial line.

Claim 1. *There exist roots $S_1, \ldots, S_d$ of lengths $D_1, \ldots, D_d$ with concatenation S such that for any two adjacent $a, b \in \{1, \ldots, n\}^d$,*

$$\varphi(S(a)) = \varphi(S(b)).$$

Proof. We define $S_d, S_{d-1}, \ldots, S_1$ in order. Suppose that $S_d, S_{d-1}, \ldots, S_{m+1}$ is already defined. Let $E_m = D_1 + \ldots + D_{m-1}$.

For $t = 0, \ldots, D_m$, let

$$W_t = (\underbrace{n-1, \ldots, n-1}_{t \text{ times}}, \underbrace{n, \ldots, n}_{D_m - t \text{ times}})$$

and define a coloring φ_t of $\{1, \ldots, n\}^{E_m+d-m}$ by

$$\varphi_t(x_1, \ldots, x_{E_m}, y_{m+1}, \ldots, y_d) = \varphi(x_1, \ldots, x_{E_m}, W_t, S_{m+1}(y_{m+1}), \ldots, S_d(y_d)).$$

The number of all k-colorings of $\{1, \ldots, n\}^{E_m+d-m}$ is

$$k^{n^{E_m+d-m}} \leq k^{n^{E_m+d}} = D_m.$$

However, we defined $D_m + 1$ colorings $\varphi_0, \ldots, \varphi_{D_m}$, and thus two of them are the same, say $\varphi_r = \varphi_s$ for some $0 \leq r < s \leq D_m$. We define

$$S_m = (\underbrace{n-1, \ldots, n-1}_{r \text{ times}}, \underbrace{\star, \ldots, \star}_{s-r \text{ times}}, \underbrace{n, \ldots, n}_{D_m - s \text{ times}}).$$

After defining $S_1, \ldots, S_d$ this way, we need to verify that the conclusion of the claim holds. Suppose that a and b are adjacent and differ only in the m-th coordinate, $a = (a_1, \ldots, a_d)$ with $a_m = n - 1$. Then

$$S(a) = S_1(a_1) \ldots S_{m-1}(a_{m-1}) W_s S_{m+1}(a_{m+1}) \ldots S_d(a_d)$$
$$S(b) = S_1(a_1) \ldots S_{m-1}(a_{m-1}) W_r S_{m+1}(a_{m+1}) \ldots S_d(a_d)$$

And thus

$$\begin{aligned}\varphi(S(a)) &= \varphi_s(S_1(a_1) \ldots S_{m-1}(a_{m-1}) a_{m-1} \ldots a_d) \\ &= \varphi_r(S_1(a_1) \ldots S_{m-1}(a_{m-1}) a_{m-1} \ldots a_d) \\ &= \varphi(S(b))\end{aligned}$$

□

Let S be the root from Claim 1. Note that if $a, b \in \{1, \ldots, n\}^d$ differ only in coordinates where they are both equal to n or $n-1$, then there exist $a = c_0, c_1, \ldots, c_m = b$ such that c_i is adjacent to c_{i-1} for $i = 1, \ldots, m$, and thus $\varphi(S(a)) = \varphi(S(b))$,

Now, consider the cube $\{1, \ldots, n-1\}^d$, and define its k-coloring ψ by setting $\psi(a) = \varphi(S(a))$ for every $a \in \{1, \ldots, n-1\}^d$. By the induction hypothesis, there exists a root T of length d such that $\psi(T(1)) = \ldots = \psi(T(n-1))$. Hence, $\varphi(S(T(1))) = \ldots = \varphi(S(T(n-1)))$. However, $S(T(n-1))$ and $S(T(n))$ only differ in coordinates where they are both equal to n or $n-1$, and thus $\varphi(S(T(n-1))) = \varphi(S(T(n)))$. Therefore, $\{S(T(1)), \ldots, S(T(n))\}$ is a monochromatic combinatorial line in φ. $\square$