Geometry and the semi-random method

Péter Hajnal

Bolyai Institute, University of Szeged, Hungary

November 23, 2017

Joint work with Endre Szemerédi

The main lemma in Komlós, Pintz, Szemerédi, 1982

The following result played a crucial role when Heilbron's conjecture ($H_{\triangle}(n) = \Theta(1/n^2)$) was disproved ($H_{\triangle}(n) \gg \frac{\sqrt{\log n}}{n^2}$).

Komlós, Pintz, Szemerédi, 1982

Let $\mathcal{H}$ be a 3-uniform hypergraph on v vertices. Let $\overline{d}$ denote the average degree of $\mathcal{H}$. Assume that $\overline{d} \leq t^2$ and $1 \ll t \ll v^{1/10}$.
If $\mathcal{H}$ doesn't contain simple cycles of length at most 4, then

$$\alpha(\mathcal{H}) = \Omega\left(\frac{v}{t}\sqrt{\log t}\right).$$

The birth of the semi-random method

Methods for constructing combinatorial objects:

- Greedy algorithm.
- Simple random selection (for example Erdős–Rényi graphs).
- The proof of Komlós–Pintz–Szemerédi Lemma (constructing an independent set in an uncrowded 3-uniform hypergraph) was a delicate mixture of the previous methods.

This, with a similar lemma from Ajtai, Komlós, Szemerédi (1981) on Sidon sequences, was the birth of the semi-random method.

Late Rödl developed his famous Rödl's nibble method for solving a diffucult problem of Erdős and Hanani.

In our further applications we might have many 3- and 4-cycles, hence, we need the following strengthening of the basic bound:

Duke, Leffmann, Rödl, 1995

Let $\mathcal{H}$ be a k-uniform hypergraph on v vertices. Let Δ the maximum degree of $\mathcal{H}$. Assume that $\Delta \leq t^{k-1}$ and $1 \ll t$. If $\mathcal{H}$ doesn't contain a 2-cycle (two edges with at least two common vertices), then

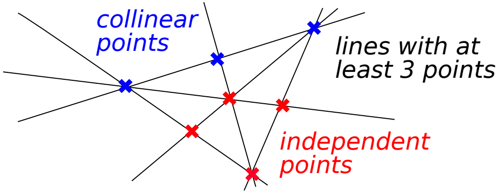
$$\alpha(\mathcal{H}) = \Omega\left(\frac{v}{t}(\log t)^{\frac{1}{k-1}}\right).$$

- A Ramsey problem of Gowers on planar point sets.
- A result in discrepancy theory: Heilbronn's problem for quadrangles.

We will prove the second result.

Gowers' question

Take any planar point set.



Observation (Gowers)

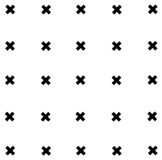
If $\mathcal{P}$ is a large enough point set, then at least n of them collinear or at least n of them are independent.

Definition

$G(n)$ is the threshold function behind "large enough".

II. The grid

$\mathcal{G}_n$: $n \times n$ grid.



(i) $\mathcal{G}_n$ has n^2 points,

(ii) $\mathcal{G}_n$ has no $n+1$ points on a line,

(iii) $\mathcal{G}_n$ has no $2n+1$ independent points.

Observation

$$G(n) \geq \frac{1}{5}n^2.$$

Assume that given n^3 points and no n of them on a line. We want to find a large independent set.

For $k = 0, 1, 2, \ldots$ do the following:
(S) Assume that we are given k of then, that form an independent set. They determine $\binom{k}{2}$ lines. These lines cover at most

$$\binom{k}{2} \cdot n < k^2 n$$

of them. Choose any point that is not covered by these lines.
(GoTo) S

It is obvious that the selection process won't stop before selecting n points. Finding n independent points is guaranteed.

Observation

$$G(n) \leq n^3.$$

II. Probabilistic upper bound

Theorem (Payne, Wood)

$$G(n) \leq 10n^2 \log n.$$

Proof: We are given $10n^2 \log n$ points, with the property that no n of them are on a line. Do the following:

Take a random subset: For each point, independently with probability $\frac{1}{5n \log n}$ keep it. Erase the lines with at least 3 remaining points.

With positive probability we have n leftover points. They are independent.

II. The semi-random method strikes again

With the semi-random method one can obtain an improvement:

Peter Hajnal, Szemerédi

$$G(n) \leq c \cdot \frac{n^2 \log n}{\log \log n},$$

for a suitable constant.

Open problem (Gowers)

Refuse or prove:

$$G(n) \leq c \cdot n^2,$$

for a suitable constant c.

A problem of Gowers

Given a planar point set $\mathcal{P}$, what is the minimal size of $\mathcal{P}$ that guarantees that one can find n points on a line or n independent points (no three on a line) in it?

A problem of Gowers

Given a planar point set $\mathcal{P}$, what is the minimal size of $\mathcal{P}$ that guarantees that one can find n points on a line or n independent points (no three on a line) in it?

Relatively easy:

- the grid shows that $\Omega(n^2)$ many points are necessary,
- in the case of $O(n^3)$ many points without n points on a line, a simple greedy algorithm finds n independent points.

Result of Payne and Wood

They improved the upper bound to $O(n^2 \log n)$,

Result of Payne and Wood

They improved the upper bound to $O(n^2 \log n)$,

They showed that if a planar point set $\mathcal{P}$ of size $cn^2 \log n$ (c is large enogh) without n points on a line must contain n independent points.

Result of Payne and Wood

They improved the upper bound to $O(n^2 \log n)$,

They showed that if a planar point set $\mathcal{P}$ of size $cn^2 \log n$ (c is large enogh) without n points on a line must contain n independent points.

First, Erdős type random sparsification (actually this is an easy theorem of Joel Spencer). Second, a greedy way to get rid of collinear triples. The leftover is an independent set.

Our contribution

Theorem

Let $\mathcal{P}$ be an arbitrary planar point set of size $\Omega\left(\frac{n^2 \log n}{\log\log n}\right)$. Then we can find n points in $\mathcal{P}$ that are either incident to a line or independent.

Our contribution

Theorem

Let $\mathcal{P}$ be an arbitrary planar point set of size $\Omega\left(\frac{n^2 \log n}{\log\log n}\right)$. Then we can find n points in $\mathcal{P}$ that are either incident to a line or independent.

The method: First get rid of 2-cycles, second use the semi-random method.

A basic question (Heilbronn)

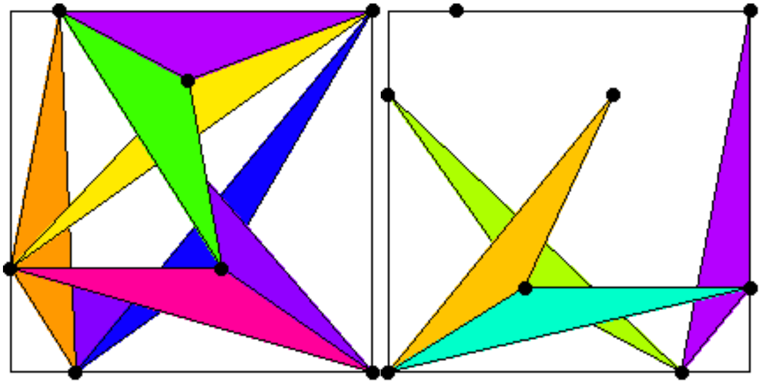
Let $\mathcal{P}$ be n points in a unit square, S. They determine $\binom{n}{3}$ triples/triangles. There is a minimal area determined by these triangles: $\Delta(\mathcal{P})$.

$$\Delta(n) = \max\{\Delta(\mathcal{P}) : \mathcal{P} \subset \mathcal{S}, |\mathcal{P}| = n\}.$$

Determine the order of magnitude of $\Delta(n)$. Give upper and lower bounds on $\Delta(n)$.

I. Extremal configurations, small values

$\Delta(3) = \Delta(4) = 1/2$ is trivial.



Configurations of 7 and 8 points. The minimal area triangles are highlighted by colors. The first configuration is proven to be optimal. The second one is not (source: Erich Friedman packing webpage). This is place for experimental mathematics.

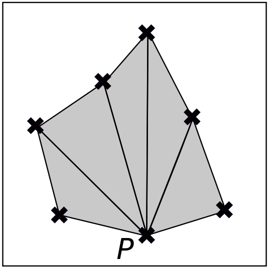
I. Upper bound: The trivial one

Give an upper bound: How small triangle will be guaranteed for sure?

Give an upper bound: How small triangle will be guaranteed for sure?

$$\Delta(n) \leq \frac{1}{n-2}.$$

$n-2$ many disjoint triangle in a unit area:



Minimal area of the disjoint triangles is at most the average area of them. $n-2$ areas sum up to at most 1.

I. Lower bound: Greedy point set

In Roth's survey paper he exhibits a simple greedy argument constructing a set that has no triangle with area smaller that $\frac{1}{10n^2}$.

"... proof was included in Schmidt's paper. But the argument may well have been discovered independently by most of those mathematicians who have made unsuccessful attempts to disprove Heilbronn's conjecture ..."

If I have time I will sketch the argument later.

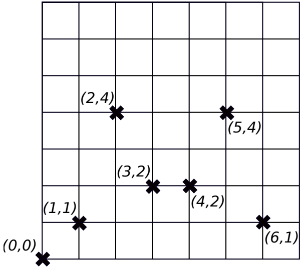
I. Lower bound: Erdős' point set

Erdős gave an elegant construction when n is a prime ($n = p$). We assume that our square is $[0, p] \times [0, p]$, and relative areas (to the base square) are considered.

Let

$$P_i = \left(i, i^2 \pmod{p}\right).$$

for $i = 0, 1, 2, \ldots, p - 1$.



The case p=7.

I. Erdős' point set: The argument

A point set, that is a part of a grid and not contains three points on a line:

$$\det\begin{pmatrix} 1 & i & i^2 \pmod p \\ 1 & j & j^2 \pmod p \\ 1 & k & k^2 \pmod p \end{pmatrix} = \det\begin{pmatrix} 1 & i & i^2 \\ 1 & j & j^2 \\ 1 & k & k^2 \end{pmatrix} \pmod p \neq 0.$$

Any triangle has area at least the area of an empty grid-triangle, $1/2$.

The relative area of a triangle (relative to the whole square) is at least

$$\frac{1/2}{p^2} = \frac{1}{2n^2}.$$

I. Lower bound: The semi-random method

An improvement of the greedy point selection.

The semi-random method has several applications in various part of mathematics: Ramsey theory, combinatorial number theory, designs, ...

Theorem (Komlós, Pintz, Szemerédi)

$$\Delta(n) \geq \frac{c \log n}{n^2},$$

for suitable constant c.

Disproof of Heilbronn's conjecture: Heilbronn conjectured (more then 30 years before the theorem, above) that for any point set of size n the smallest area among the triangles is $\mathcal{O}(1/n^2)$.

After small steps Roth made a breakthrough:

Theorems

For arbitrary $\varepsilon > 0$ and large enough n

(i) Roth 1972
$$\Delta(n) \leq n^{-1.105\ldots+\varepsilon}$$

(ii) Roth 1972
$$\Delta(n) \leq n^{-1.117\ldots+\varepsilon}$$

(iii) Komlós, Pintz, Szemerédi 1981
$$\Delta(n) \leq n^{-\frac{8}{7}+\varepsilon}$$

I. Further questions

(1) Can you prove that

$$\Delta(n) \leq \frac{c_\varepsilon}{n^{2-\varepsilon}}$$

for any $\varepsilon > 0$ and suitable constant?

(2) What happens if you consider the smallest area quadrangle determined by our points. Introduce $\square(n)$. Is it true, that

$$\square(n) = o\left(\frac{1}{n}\right)?$$

Heilbron's problem

Take a "nice" unit area domain D (for example a square). Place n points into D and find the smallest area among the triangles determined by the chosen points. Let $H_\triangle(n)$ denote the maximum of this parameter over all possible n points.

Heilbron's problem

Take a "nice" unit area domain D (for example a square). Place n points into D and find the smallest area among the triangles determined by the chosen points. Let $H_\triangle(n)$ denote the maximum of this parameter over all possible n points.

The easy bounds:

$$\frac{1}{n^2} \ll H_\triangle(n) \ll \frac{1}{n}.$$

Heilbron's problem

Take a "nice" unit area domain D (for example a square). Place n points into D and find the smallest area among the triangles determined by the chosen points. Let $H_\triangle(n)$ denote the maximum of this parameter over all possible n points.

The easy bounds:

$$\frac{1}{n^2} \ll H_\triangle(n) \ll \frac{1}{n}.$$

Roth and Schmidt proved fundamental theorems on this problem.

A breakthrough result is:

Komlós, Pintz, Szemerédi, 1982

$$H_{\triangle}(n) = \Omega\left(\frac{\sqrt{\log n}}{n^2}\right).$$

One can take k-tuples of our point set and consider the area of the convex hull of the k chosen points. We denote the corresponding parameter by $H_k(n)$ (so $H_3(n) = H_\triangle(n)$).

Trivial observations:

$$H_3(n) \ll H_4(n) \ll H_5(n) \ll \ldots \ll \mathcal{O}\left(\frac{1}{n}\right).$$

Two major open problems: $H_\triangle(n) = O(1/n^{2-o(1)})$ and $H_4(n) = o(1/n)$.

The lower bound on $H_4(n)$

Schmidt proved that

$$H_4(n) = \Omega(n^{-3/2}).$$

The proof is a construction of a point set by a simple greedy algorithm.

Our contribution

With the help of the semirandom method we are able to improve this bound.

Theorem (H_4)

There exists a point set of size n in the unit square that doesn't contain four points with convex hull of area at most $\mathcal{O}(n^{-3/2}(\log n)^{1/2})$.

Let

$$S := \{(x, y) \in \mathbb{R}^2 : |x|, |y| \leq 1/2\}$$

be a unit square on the plane. Choose N (a parameter that will be chosen later) random points (independently with uniform distribution) from

$$(1/2)S = \{(x/2, y/2); (x, y) \in S\}.$$

Let $\mathcal{P}$ be the random point set $\{P_1, P_2, \ldots, P_N\}$ we obtain this way.

Observation: any connecting line of two points from $\mathcal{P}$ has an intersection with S of length $\Theta(1)$.

4-uniform hypergraph $\mathcal{Q}$ on the vertex set $\mathcal{P}$

A four element point set $\{P, Q, R, S\}$ forms an edge iff

$$Area(PQRS) < \tau,$$

here τ is a threshold to be determined later.

$\mathcal{Q}$ is a random 4-uniform hypergraph.

The major part of the proof is bounding the expected values of combinatorial parameters of $\mathcal{Q}$.

Let $A, B \in \mathcal{P}$ two different points and

$$deg(A, B) = |\{\{C, D\} : \{A, B, C, D\} \in \mathcal{Q}\}|$$

Clearly,

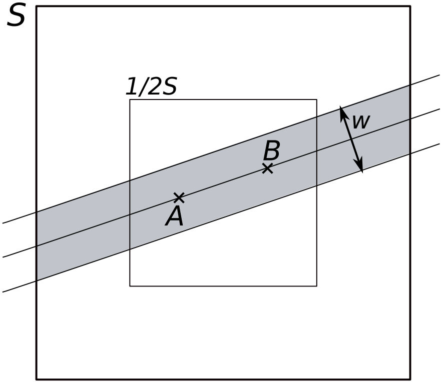
$$deg(A, B) = \frac{1}{2}|\{(C, D) : \{A, B, C, D\} \in \mathcal{Q}\}|$$

That is, $deg(A, B)$ denotes the number of edges of $\mathcal{Q}$ that contains both A and B.

Our goal is to give an upper bound for this parameter. We will count how many ordered pair of points C, D are considered when $deg(A, B)$ is determined.

Strips

Let $strip(AB, w)$ denote the set of points from S that are in the strip of width w with midline AB (see the Figure).
I.e. $strip(AB, w)$ contains those points of S that have distance at most $w/2$ from line AB.

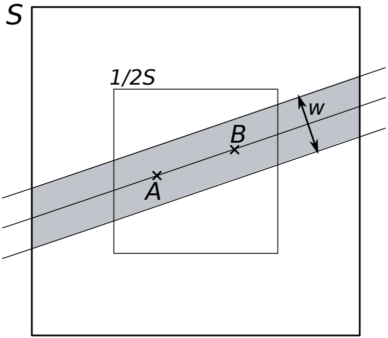


Note that $strip(AB, w)$ has area $\Theta(w)$.

Bounding $d(A, B)$

Strategy: Fix A and B, let $d = dist(A, B)(< 1)$. $deg(A, B)$ counts certain C, D pairs of points.

We distinguish cases according to the position of C, an arbitrary point from $\mathcal{P} - \{A, B\}$ and we bound the possible positions of the D's that contribute to $deg(A, B)$ with the current C.



3 cases

Case 1: $C \notin strip(AB, 4\tau/d)$.
In this case the area of $ABC\triangle$ is at least τ, hence this C doesn't contribute to $deg(A, B)$.

Case 2: $C \in strip(AB, 4\tau/\sqrt{d})$. Note that $strip(AB, 4\tau/\sqrt{d})$ has area $\Theta(\tau/\sqrt{d})$, assuming $\tau \ll \sqrt{d} < d$.
Outside of $strip(AB, 4\tau/d)$ no D contributes to $deg(A, B)$ (see Case 1).
So we can bound the possible positions of contributing D's to $strip(AB, 4\tau/d)$. This strip has area $\Theta(\tau/d)$.

Case 3: $C \in strip(AB, 4\tau/d) - strip(AB, 4\tau/\sqrt{d})$ (note that $d < \sqrt{d} < 1$). $strip(AB, 4\tau/d) - strip(AB, 4\tau/\sqrt{d})$ has area $\mathcal{O}(\tau/d)$.
The contributing D's must come from $strip(AB, 4\tau/d) \cap strip(AC, 4\tau/dist(A, C))$.

Bounding the positions of contributing D's in case 3

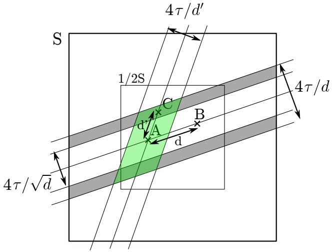


Figure: The shaded region is the space for those C's where Case 3 applies. The green region contains those D's that can form an edge of $\mathcal{Q}$ with A, B and C.

Case 3(cont'd): By elementary geometry the green parallelogram on the above figure, bounding the possible positions of contributing D's, has area $\Theta(\tau^2/Area(ABC\triangle))$. As $Area(ABC\triangle) = \Omega(d \cdot \tau/\sqrt{d})$, the parallelogram has area $\mathcal{O}(\tau/\sqrt{d})$.

Bounding $\mathbb{E}\,(deg(A,B))$

The expected numbers of C's come from the area bounds

$$\text{Case 1:}\quad 0,\qquad \text{Case 2:}\quad \mathcal{O}(N\tau/\sqrt{d}),\qquad \text{Case 3:}\quad \mathcal{O}(N\tau/d).$$

We can bound the corresponding numbers of D's

$$\text{Case 1:}\quad 0,\qquad \text{Case 2:}\quad \mathcal{O}(N\tau/d),\qquad \text{Case 3:}\quad \mathcal{O}(N\tau/\sqrt{d}).$$

Since the choice of C and D are independent, the number of contributing (C,D)'s in expectation is a product of two expectations.
In each of the three cases this product is $\mathcal{O}(\tau^2 d^{-3/2} N^2)$. Hence

$$\mathbb{E}\,(deg(A,B)) = \mathcal{O}(\tau^2 dist^{-3/2}(A,B) N^2).$$

Bounding the number of 2-cycles through $A, B \in \mathcal{P}$

$\mathcal{Q}$ may have two types of 2-cycles (now different symbols denote different points):

(I): $\{A, B, C, D\}$, $\{A, B, C', D'\}$,

(II): $\{A, B, C, D\}$, $\{A, B, C, D'\}$

Let $\mathcal{C}_I(A, B)$, resp. $\mathcal{C}_{II}(A, B)$ denote the number of 2-cyles of type (I), resp. type (II) through given A, B points.

Bounding the expected value of $\mathcal{C}_I(A, B)$ is easy, based on the previous calculation

$$\mathbb{E}\left(\mathcal{C}_I(A, B)\right) = \mathcal{O}(\tau^4 dist^{-3}(A, B) N^4).$$

Type (II) cycles

Bounding the expected value of $\mathcal{C}_{II}(A, B)$ is a little bit more technical. We distinguish the contribution of C's that satisfy Case 2 and those that satisfy Case 3:

$$\mathbb{E}\left(\mathcal{C}_{II}(A, B)\right) = \mathcal{O}\left((N\tau/\sqrt{d})(N\tau/d)^2 + (N\tau/d)(N\tau/\sqrt{d})^2\right),$$

hence,

$$\mathbb{E}\left(\mathcal{C}_{II}(A, B)\right) = \mathcal{O}(\tau^3 N^3 d^{-2.5}).$$

Sparsification to lower bound the minimal distance

We sparsify our point set a little bit in order to have a lower bound on the minimal distance determined by our points.

Let $\delta = \frac{1}{100} N^{-1/2}$. Let $C(\mathcal{P})$ be the set of pairs of $\mathcal{P}$ that are closer than δ (this is a random set).
Let $C_A(\mathcal{P})$ be the set of points that contains points $B \in \mathcal{P}$ closer than δ to A (this is a random set too).
Let $Disc(A;\delta)$ denote the disc of radius δ centered at A.

Obviously,

$$C_A(\mathcal{P}) = \mathcal{P} \cap Disc(A;\delta),$$

$$|C(\mathcal{P})| = 1/2 \sum_A |C_A(\mathcal{P})|,$$

and

$$\mathbb{E}\ |C_A(\mathcal{P})| \leq (N-1) Area(Disc(A;\delta)) = 1/2\pi \cdot \delta^{-2} N < 1/1000.$$

We get that $\mathbb{E}\left(|C(\mathcal{P})|\right) \leq N/1000$, hence with sufficiently large probability

$$|C(\mathcal{P})| \leq N/4.$$

After deleting these pairs we will obtain $\mathcal{P}_0$, our new point set. $|\mathcal{P}_0| \geq N/2$ with high probability, and any two points of it are at least distance δ apart.

Let $\mathcal{Q}_0$ be the restriction of $\mathcal{Q}$ to $\mathcal{P}_0$. From now on we will work with $\mathcal{Q}_0$.

Lemma on sets with bounded minimal distance

Lemma

Take M points from S so that the minimal distance among them is at least δ. Let $P \in S$. Let $Ann_i(P, \delta)$ be the annulus

$$Ann_i(P; \delta) = \{X \in \mathbb{R}^2 : (i-1)\delta < dist(P, X) \leq i\delta\}.$$

$Ann_1(P; \delta)$, $Ann_2(P; \delta)$, ..., $Ann_{\mathcal{O}(\delta^{-1})}(P; \delta)$ are disjoint and cover S (hence they cover our point set). Furthermore at most $\mathcal{O}(i)$ of our M points can be covered by $Ann_i(P, \delta)$.

Proof by picture

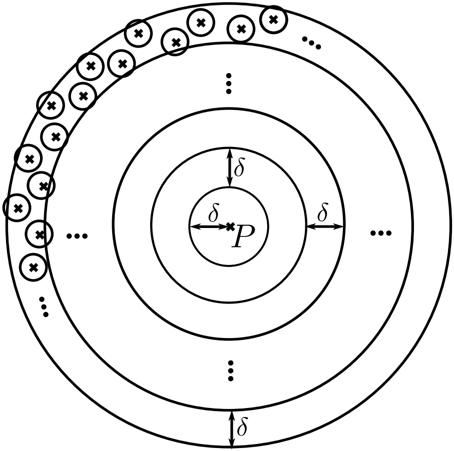


Figure: The annuluses around P, and the elementary volume argument in the proof.

Proof by words

The covering property is obvious.

The bound on the number of points in the annulus is a simple volume argument: Draw $Disc(A, \delta/3)$ for all points A of us from $Ann_i(P; \delta)$. These discs are disjoint within $\{X \in \mathbb{R}^2 : (i - 4/3)\delta < dist(P, X) \leq (i + 1/3)\delta\}$.

The lemma follows immediately.

Easy calculation: Bounding $\mathbb{E}\ |\mathcal{Q}_0|$

$$\begin{aligned}
\mathbb{E}\ deg_{\mathcal{Q}_0}(A) \leq \sum_B \mathbb{E}\ deg(A,B) = &\sum_{i=1}^{\mathcal{O}(N^{1/2})} \sum_{B\in Ann_i(A,\delta)} deg(A,B) \\
\leq &\sum_{i=1}^{\mathcal{O}(N^{1/2})} \sum_{B\in Ann_i(A,\delta)} \mathcal{O}(\tau^2 N^2 (i/\sqrt{N})^{-3/2}) \\
\leq &\sum_{i=1}^{\mathcal{O}(N^{1/2})} i\cdot\mathcal{O}(\tau^2 N^{2.75} i^{-3/2}) = \mathcal{O}(\tau^2 N^{2.75}) \sum_{i=1}^{\mathcal{O}(N^{1/2})} i\cdot i^{-3/2} \\
= &\mathcal{O}(\tau^2 N^{2.75})\mathcal{O}(N^{0.25}) = \mathcal{O}(\tau^2 N^3).
\end{aligned}$$

Hence

$$\mathbb{E}\ |\mathcal{Q}_0| = \mathcal{O}(\tau^2 N^4).$$

Easy calculation: Bounding $\mathbb{E}\left(\mathcal{C}_I\right)$

$$
\begin{aligned}
\mathbb{E}\,\mathcal{C}_I &\le \sum_{A,B} \mathbb{E}\,\mathcal{C}_I(A,B) = \sum_{A} \sum_{i=1}^{\mathcal{O}(N^{1/2})} \sum_{B\in Ann_i(A,\delta)} \mathbb{E}\,\mathcal{C}_I(A,B) \\
&= \sum_{A} \sum_{i=1}^{\mathcal{O}(N^{1/2})} \sum_{B\in Ann_i(A,\delta)} \mathcal{O}(\tau^4 \cdot i^{-3} N^{1.5} \cdot N^4) \\
&= \sum_{A} \sum_{i=1}^{\mathcal{O}(N^{1/2})} \mathcal{O}(\tau^4 \cdot i^{-2} \cdot N^{5.5}) = \sum_{A} \mathcal{O}(\tau^4 N^{5.5}) \sum_{i=1}^{\mathcal{O}(N^{1/2})} i^{-2} = \\
&= \color{red}{\mathcal{O}(\tau^4 N^{6.5}).}
\end{aligned}
$$

Easy calculation: Bounding $\mathbb{E}\ (\mathcal{C}_{II})$

$$\mathbb{E}\ \mathcal{C}_{II} \leq \sum_{A,B} \mathbb{E}\ \mathcal{C}_{II}(A,B) = \sum_{A} \sum_{i=1}^{\mathcal{O}(N^{1/2})} \sum_{B \in Ann_i(A,\delta)} \mathbb{E}\ \mathcal{C}_{II}(A,B)$$

$$= \sum_{A} \sum_{i=1}^{\mathcal{O}(N^{1/2})} \sum_{B \in Ann_i(A,\delta)} \mathcal{O}(\tau^3 \cdot i^{-2.5} N^{1.25} \cdot N^3)$$

$$= \sum_{A} \sum_{i=1}^{\mathcal{O}(N^{1/2})} \mathcal{O}(\tau^3 \cdot i^{-1.5} \cdot N^{4.25}) = \sum_{A} \mathcal{O}(\tau^3 N^{4.25}) \sum_{i=1}^{\mathcal{O}(N^{1/2})} i^{-1.5} =$$

$$= \mathcal{O}(\tau^3 N^{5.25}).$$

Hence,

$$\mathbb{E}\ \mathcal{C} = \mathbb{E}\ (\mathcal{C}_I + \mathcal{C}_{II}) = \mathcal{O}(\tau^4 N^{6.5}) + \mathcal{O}(\tau^3 N^{5.25}).$$

The fruits of our bounds

Fix the random bits of the probabilistic process producing $\mathcal{P}_0$ and $\mathcal{Q}_0$ so that

- the number of points is $N/2$,
- the number of quadruples is $\mathcal{O}(\tau^2 N^4)$, and
- the number of 2-cycles is $\mathcal{O}(\tau^4 N^{6.5}) + \mathcal{O}(\tau^3 N^{5.25})$.

Let $\mathcal{Q}_1$ be the 4-uniform hypergraph we obtained this way.

Sparsification to get rid of 2-cycles

Random sparsification for getting rid of the 2-cycles: with probability p keep a point and with probability $1-p$ throw it away. Do this independently for all points.

Let $\mathcal{Q}_1$ be the random 4-uniform hypergraph we obtain this way.

Its parameters can be easily bounded:

$$\mathbb{E}\ |V(\mathcal{Q}_1)| = \Theta(pN), \quad \mathbb{E}\ |\mathcal{Q}_1| = \mathcal{O}(p^4\tau^2N^4),$$

$$\mathbb{E}\ (\mathcal{C}) = \mathcal{O}(p^6\tau^4N^{6.5}) + \mathcal{O}(p^5\tau^3N^{5.25}).$$

Sparsification to get rid of 2-cycles

Choose p so that

$$\mathcal{C} \ll |V(\mathcal{Q}_1)|.$$

Using Markov's inequality fix the random choices so that we obtain a hypergraph with the property that after deleting the points of the 2-cycles we obtain a leftover hypergraph on $\Theta(pN)$ points, $\mathcal{O}(p^4\tau^2N^4)$ edges, and no 2-cycles.

Let $\overline{d}$ denote the average degree. Throw away the points with degree at least $10\overline{d}$.

The leftover hypergraph (without 2-cycles) is denoted by $\mathcal{L}$ and its parameters are:

$$|V(\mathcal{L})| = \Theta(pN), \qquad |\mathcal{L}| = \mathcal{O}(p^4\tau^2N^4), \qquad \Delta(\mathcal{L}) = \mathcal{O}(p^3\tau^2N^3).$$

The right choices for parameters

Choose N, τ in such a way that $\alpha(\mathcal{L}) \geq n$ will hold. The n points forming an independent set will prove our main result.
Do some arithmetics:

$$p := \mathcal{O}(n^{-0.001}), \quad N := \mathcal{O}(n^{1.01}), \quad \tau = \mathcal{O}(n^{-3/2}\sqrt{\log n}).$$

The relevant computation is as follows:

$$\mathbb{E}\,(\mathcal{C}) = \mathcal{O}(p^6\tau^4N^{6.5}) + \mathcal{O}(p^5\tau^3N^{5.25}) = o(n)$$

with the above choice of p, N and τ, and at the same time

$$\mathbb{E}\,|V(\mathcal{Q}_1)| = \Theta(pN) = \Theta(n^{1.009}).$$

Hence, getting rid of 2-cycles is easy.

Recalling Theorem [DLR 1995]

Theorem (Duke, Leffmann, Rödl, 1995)

Let $\mathcal{H}$ be a k-uniform hypergraph on v vertices. Let Δ the maximum degree of $\mathcal{H}$. Assume that $\Delta \le t^{k-1}$ and $1 \ll t$. If $\mathcal{H}$ doesn't contain a 2-cycle (two edges with at least two common vertices), then

$$\alpha(\mathcal{H}) = \Omega\left(\frac{v}{t}(\log t)^{\frac{1}{k-1}}\right).$$

Application of Theorem [DLR 1995]

In order to apply Theorem [DLR 1995] we introduce a parameter t, such that $\Delta(\mathcal{L}) \leq t^3$.

Since $\Delta(\mathcal{L}) = \mathcal{O}(p^3\tau^2N^3)$ the right choice for t is

$$t = \Theta(p\tau^{2/3}N) = \Theta(n^{0.001}(n^{-1}\log^{1/3} n)n^{1.01}) = \Theta(n^{0.009}\log^{1/3} n).$$

Hence Theorem [DLR 1995] is applicable and it gives the following bound:

$$\alpha(\mathcal{L}) \geq \frac{\Omega(pN)}{t}\log^{1/3} t = \tau^{-2/3}\cdot\Omega(\log^{1/3} n) = \Omega(n).$$

After scaling, we obtain $\alpha(\mathcal{L}) \geq n$.

The theorem is proved by an independent set of size n in $\mathcal{L}$.

Appendix: Schmidt's greedy bound on Heilbronn's $\Delta(n)$

Assume that we have chosen $P_1, P_2, \ldots, P_k$ points from $\mathcal{S}$, the unit square.

We have two assumptions:

- The distance of any two different chosen points is at least

$$\frac{1}{10} \cdot \frac{1}{\sqrt{n}}.$$

- The area of any triangle determined by three different chosen points is at least

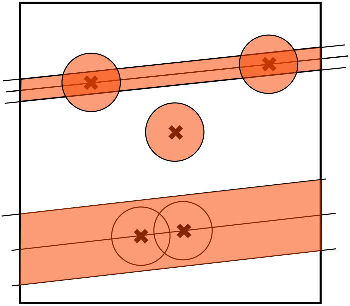
$$\frac{1}{10} \cdot \frac{1}{n^2}.$$

We choose the next point, and meanwhile we maintain these two properties (assuming that fewer than n points are chosen so far)..

Appendix: Forbidden regions

(I) Draw circles centered at the chosen points of radius $\frac{1}{10} \cdot \frac{1}{\sqrt{n}}$.

(II) Draw strips for any P_i, P_j pair of points: its mid-line is the line P_iP_j, its width $\frac{4}{10n^2 d(P_i,P_j)}$.



These regions (only two strips are drawn) are forbidden if we maintain the two properties.

How big is the area of the forbidden region?

(I) The circles cover at most

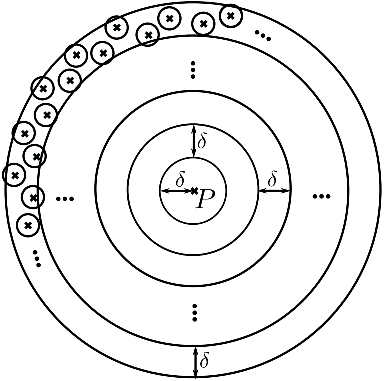
$$k \cdot \left(\frac{1}{10\sqrt{n}}\right)^2 \cdot \pi$$

area. That is small.

(II) The area, covered by the strips is more complicated to bound. The area of a strip depends on its width. The width depends on the distance of the corresponding two points. In order to bound the area covered by the strip we need some knowledge on the distribution of distances determined by our points.

Appendix: Distance distribution

Take any of our points, P. We cover the square with annuluses centered at P of width $\delta = \frac{1}{10\sqrt{n}}$:



Easy to see that our first condition ensures that the i^{th} annulus contains at most $\mathcal{O}(i)$ many points.

Appendix: The end

The rest is arithmetic:

The area of the union of the strip is small: We know, that among the PP_i segments how many contribute a strip of approximate width of $\frac{4}{10n^2 \cdot i\delta}$.

Easy to see that the forbidden regions cannot cover our square (assuming we have chosen fewer than n points). The selection of the next point is possible.

Since the chosen point is outside the strips, we won't have a triangle with "small" area.

Thank you for your attention!