ON THE GENERAL POSITION SUBSET SELECTION PROBLEM

Michael S. Payne
University of Melbourne

David R. Wood
Monash University

MCW XIX
Prague - 1/8/13
BASIC IDEAS

△ We consider finite point sets $P$ in the Euclidean plane.

△ $P$ is in general position if no three points of $P$ are collinear.

△ Given a point set $P$ not in general position, we are interested in selecting the largest possible subset in general position.
Dudeney (1917) **No-three-in-line problem:**

“What is the size of the largest subset of the $n \times n$ grid in general position?”

A best possible example for $n = 52$:

(Flammenkamp '98)
The Problem

Erdős ('88) asked, determine the largest integer $f(n,l)$ such that every set of $n$ points with at most $l$ collinear contains a subset of $f(n,l)$ points in general position.
The Problem

Erdős (’88) asked, determine the largest integer $f(n,l)$ such that every set of $n$ points with at most $l$ collinear contains a subset of $f(n,l)$ points in general position.

Gowers (MathOverflow 2011) asked:

What is the minimum integer $GP(q)$ such that every set of $GP(q)$ points in the plane contains $q$ collinear points or $q$ points in general position?
Bounds on $\text{GP}(q)$

- Gowers noted that $\Omega(q^2) \leq \text{GP}(q) \leq O(q^3)$.
  - Lower bound: $\frac{q-1}{2} \times \frac{q-1}{2}$ grid.

- Upper bound:
  - Suppose we choose $q-1$ points in $g,p$.
  - Others lie on $\binom{q-1}{2}$ lines.
  - At most $q-1$ per line $\Rightarrow O(q^3)$.

- We will show that $\text{GP}(q) \leq O(q^2 \log q)$. 
**A GEOMETRIC LEMMA**

**Lemma:** Let $P$ be a set of $n$ points with no $q$ collinear. Then the number of collinear triples in $P$ is at most $c(n^2 \ln q + q^2 n)$ for some constant $c$.

**Proof:**

1. Let $S_i$ be the number of lines containing $i$ points.
2. Szemeredi-Trotter Theorem (’83):

   \[ orall i \quad \sum_{j \geq i} S_j \leq c \left( \frac{n^2}{i^2} + \frac{n}{i} \right) \]

   for some constant $c$. 
Proof:

Let $S_i$ be the number of lines containing $i$ points.

Szemerédi-Trotter: \[ \sum_{j \geq i} S_j \leq c \left( \frac{n^2}{i^3} + \frac{n}{i} \right). \]

So the number of collinear triples is

\[
\sum_{i=2}^{q-1} \binom{i}{3} S_i = \sum_{i=2}^{q} i^2 \sum_{j=i}^{q} S_j \leq \sum_{i=2}^{q} i^2 \left( \frac{n^2}{i^3} + \frac{n}{i} \right) \leq c \sum_{i=2}^{q} \left( \frac{n^2}{i} + \ln \frac{q}{i} \right) \leq c \left( n^2 \ln q + q^2 n \right). \]

\[ \square \]
A Hypergraph Lemma

We consider the 3-uniform hypergraph $H(P)$ of collinear triples in $P$.

A subset in general position is an independent set in $H(P)$.

Lemma (Spencer '72): Let $H$ be a 3-uniform hypergraph with $n$ vertices and $m$ edges. If $m < \frac{n^2}{3}$ then $\alpha(H) > \frac{n}{2}$. If $m \geq \frac{n^2}{3}$ then

$$\alpha(H) > c' \sqrt{\frac{n}{m/n}}.$$
A NEW Bound

**Theorem:** Let $P$ be a set of $n$ points with no $q$ collinear and no $q$ in general position. Then $n \leq O(q^2 \ln q)$.

**Proof:** May assume $q^2 < n$, so $m < cn^2 \ln q$.

$$q > \alpha(H) > c' \frac{n}{\sqrt{m/n}} \quad \text{(or } q > n^{1/2})$$

$$\Rightarrow \quad q > c'' \frac{n}{\sqrt{m \ln nq}}$$

$$\Rightarrow \quad n \leq O(q^2 \ln q) \quad \Box$$
Original Problem

- Erdős ('88) asked, determine the largest integer $f(n,t)$ such that every set of $n$ points with at most $t$ collinear contains a subset of $f(n,t)$ points in general position.

- Füredi ('91) noted that ‘density Hales-Jewett’ implies that $f(n,t) \leq o(n)$.

- Lehmann (2012) showed that for fixed $t$

  $$f(n,t) \geq \Omega (\sqrt{n \ln n}).$$

  (Füredi proved this for $t=3$).
Bounds for variable $\ell$

- We show that if $\ell \leq O(\sqrt{n})$ then
  \[ f(n, \ell) \geq \Omega(\sqrt{n}). \]

- Furthermore, if $\ell \leq O(n^{(1-\varepsilon)/2})$ then
  \[ f(n, \ell) \geq \Omega_{\varepsilon}(\sqrt{n \log \ell n}). \]

- Method is similar, using Szemeredi-Trotter-based lemma and known results on independent sets in hypergraphs.
**Conjectures**

1) $f(n, \sqrt{n}) \geq \Omega(\sqrt{n})$ (or $GP(q) \leq O(q^2)$).

2) Every set of $n$ points with at most $\sqrt{n}$ collinear can be coloured with $O(\sqrt{n})$ colours such that each colour is in general position.

   ▷ (1) and (2) are true for the grid.

3) For fixed $t$, $f(n, t) \geq \Omega(n / \text{polylog}(n))$.

Subsets with at most \( k \) Collinear

Determine the largest integer \( f(n, l, k) \) such that every set of \( n \) points with at most \( l \) collinear contains a subset of \( f(n, l, k) \) points with at most \( k \) collinear, where \( k < l \).

\[ \text{We show if } k \geq 3 \text{ is fixed and } l \leq O(\sqrt{n}) \]

\[ \text{then } f(n, l, k) \geq \Omega \left( \frac{n^{(k-1)/k}}{\ell^{(k-2)/k}} \right). \]

[This implies \( G_{P_k}(q) \leq O(q^2) \)]

\[ \text{If } k \geq 3 \text{ is fixed and } l \leq O \left( n^{(1-\varepsilon)/2} \right) \]

\[ \text{then } f(n, l, k) \geq \Omega \left( \frac{n^{(k-1)/k}}{\ell^{(k-2)/k}} \left( \ln n \right)^{1/k} \right). \]