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.

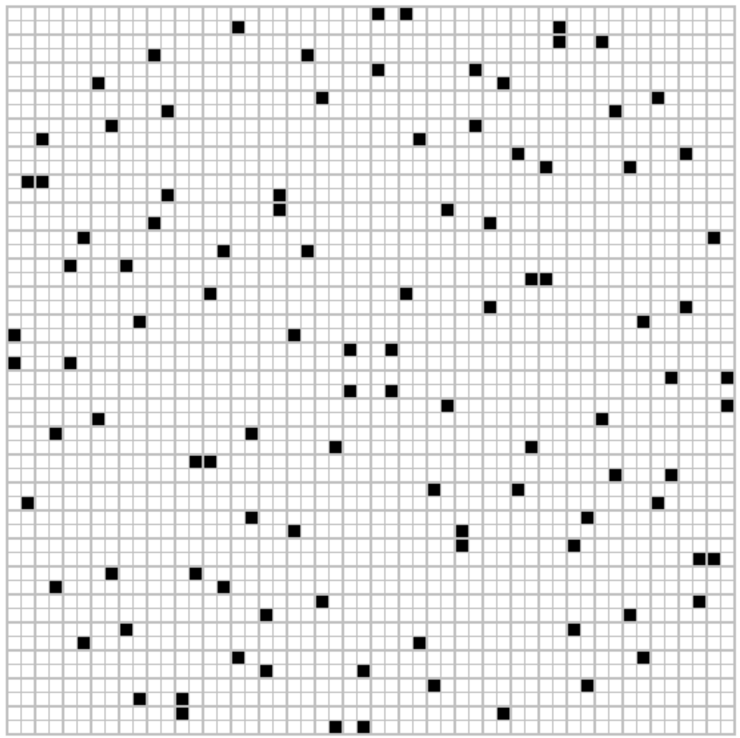
HISTORY

▷ 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 $f(n, \ell)$ such that every set of n points with at most ℓ collinear contains a subset of $f(n, \ell)$ points in general position.

The Problem

- Erdős ('88) asked, determine the largest integer $f(n,\ell)$ such that every set of n points with at most ℓ collinear contains a subset of $f(n,\ell)$ 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 GP(q)

- Gowers noted that $\Omega(q^2) \leq GP(q) \leq O(q^3)$.
  - 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 $\underline{GP(q) \leq O(q^2 \ln q)}$.

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:

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

- Szemeredi-Trotter Theorem ('83):

$$\forall i \qquad \sum_{j \geq i} S_j \leq c\left(\frac{n^2}{i^3} + \frac{n}{i}\right)$$

for some constant c.

Proof:

- Let s_i be the number of lines containing i points.
- Szemeredi-Trotter: $\displaystyle\sum_{j \geq i} s_j \leq c\left(\frac{n^2}{i^3} + \frac{n}{i}\right)$.
- So the number of collinear triples is

$$\sum_{i=2}^{q-1} \binom{i}{3} s_i \leq \sum_{i=2}^{q} i^2 \sum_{j=i}^{q} s_j$$

$$\leq \sum_{i=2}^{q} c i^2 \left(\frac{n^2}{i^3} + \frac{n}{i}\right) \leq c \sum_{i=2}^{q} \left(\frac{n^2}{i} + in\right)$$

$$\leq c\left(n^2 \ln q + q^2 n\right). \qquad \square$$

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 $m < n/3$ then $\alpha(H) > n/2$.
If $m \geq n/3$ then

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

A New Bound

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

<u>Proof:</u> May assume $q^2 < n$, so $m < c\, n^2 \ln q$.

$$q > \alpha(H) > c' \frac{n}{\sqrt{m/n}} \qquad \left(\text{or } q > n/2\right)$$

$$\Rightarrow \qquad q > c'' \frac{n}{\sqrt{n \ln q}}$$

$$\Rightarrow \qquad n \le O(q^2 \ln q).$$

$\square$

Original Problem

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

- Füredi ('91) noted that 'density Hales-Jewett' implies that $f(n,\ell) \leq o(n)$.

- Lefmann (2012) showed that for fixed ℓ

$$f(n,\ell) \geq \Omega\left(\sqrt{n \ln n}\right).$$

(Füredi proved this for $\ell = 3$).

BOUNDS FOR VARIABLE ℓ

- We show that if $\ell \leq O(\sqrt{n})$ then
$$f(n,\ell) \geq \Omega\left(\sqrt{\frac{n}{\ln n}}\right).$$
- Furthermore, if $\ell \leq O(n^{(1-\varepsilon)/2})$ then
$$f(n,\ell) \geq \Omega_{\varepsilon}\left(\sqrt{n \log_{\ell} n}\right).$$
- Method is similar, using Szemeredi-Trotter-based lemma and known results on independent sets in hypergraphs.

CONJECTURES

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

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 ℓ, $f(n, \ell) \geq \Omega(n / \mathrm{polylog}(n))$.

   ▷ True for $[3]^d$.

Subsets with at most k collinear

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

- We show if $k \geqslant 3$ is fixed and $\ell \leqslant O(\sqrt{n})$ then $f(n,\ell,k) \geqslant \Omega\left(\frac{n^{(k-1)/k}}{\ell^{(k-2)/k}}\right)$.

  [This implies $GP_k(q) \leqslant O(q^2)$]

- If $k \geqslant 3$ is fixed and $\ell \leqslant O\left(n^{(1-\varepsilon)/2}\right)$ then $f(n,\ell,k) \geqslant \Omega_\varepsilon\left(\frac{n^{(k-1)/k}}{\ell^{(k-2)/k}}(\ln n)^{1/k}\right)$.