Forbidden Configurations in Discrete Geometry

David Eppstein

Southern California Theory Day
University of California, Irvine
November 3, 2018

The big picture

Induced subgraphs and hereditary properties are central to graph theory

> Cliques, coloring, perfect graphs, line graphs, comparability graphs, . . .

Also analogous to permutations and permutation patterns

CC-BY-SA image "Auto-portrait de Van gogh realisé en 2000 polos - Tokyo.JPG" by Arthur Causse from Wikimedia commons

We should build a similar theory in discrete geometry!

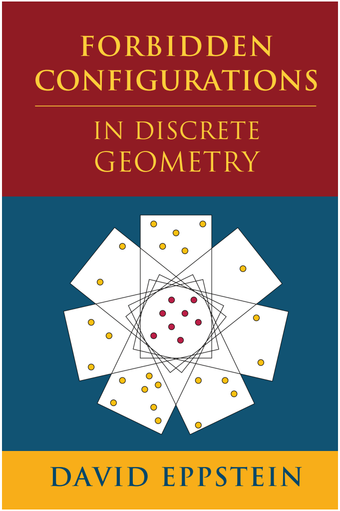
. . . or maybe we already have and we just didn't realize it?

This talk came first

... but it blew up into a book
(Cambridge Univ. Press, 2018)

Algorithmic, but
not about algorithms

(in the same way that most math books include theorems but are not about theorem-proving)



I: A happy ending

Quadrilaterals in five-point sets

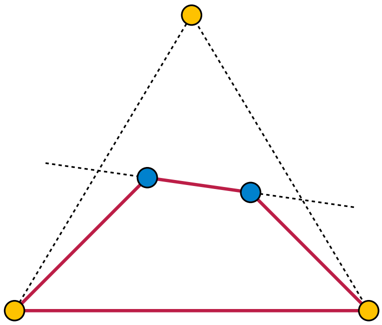
Esther Klein, early 1930s:

Five points in general position
(no three in a line)
contain a convex quadrilateral

Proof:

If convex hull has 4 or 5
vertices, obvious

Otherwise, line through inner
points misses a hull edge; use
that edge + inner points



Erdős and Szekeres, 1935

Generalization to larger convex polygons

Every $\binom{2k-4}{k-2} + 1 < 4^k$ points in general position
contain a convex k-gon

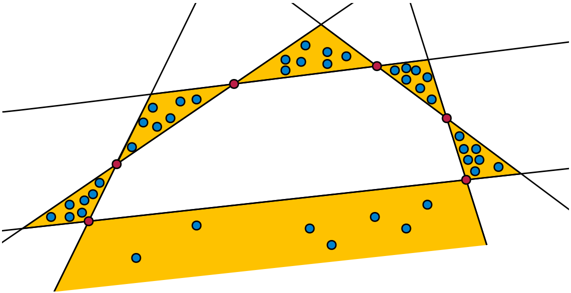
Conjecture: # points needed to ensure a convex k-gon is $2^{k-2} + 1$

Still open, $500 prize for solution

Klein and Szekeres marry and escape the Nazis as refugees in Shanghai and then mathematics professors in Australia commemorated in the theorem name

Suk 2016

Every $2^{k+O(k^{2/3}\log k)}$ points in general position have a convex k-gon



Proof strategy:

Use E–S 1935 to find a big cup or cap (red)
such that each yellow region has many points

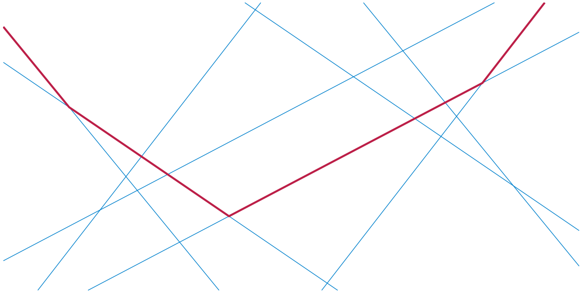
Use Dilworth to find large chains or antichains in each yellow region
for partial ordering by triangle containment

Apply case analysis to glue together yellow regions

The algorithmic version of the problem

How to find the largest convex set in a given input?
Chvátal and Klincsek 1980; Edelsbrunner and Guibas 1989

Biggest polygon with fixed bottom vertex is dual to longest convex chain in line arrangement, found by sweeping the arrangement



Test all choices of bottom vertex: $O(n^3)$ time, $O(n)$ space

(Related algorithms for finding convex subsets with other optimization criteria: E, Overmars, Rote, Woeginger, 1992; E, Erickson, 1994)

Key properties of largest convex polygon

The function that maps point sets to the size of their largest convex polygon has two key properties.

Monotone:

Removing points can only reduce largest convex subset

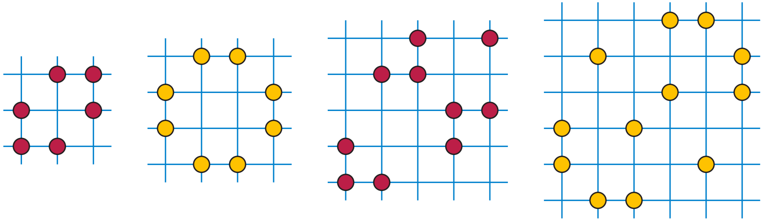
Invariant:

Depends only on relative orientations of triples of points, not on their exact locations

II: A menagerie of monotone invariant problems

The no-three-in-line problem

Dudeney 1917: Largest general-position subset of $n \times n$ grid



Erdős 1951: at least $n(1 - o(1))$

Hall et al. 1975: at least $n(1.5 - o(1))$

Guy 2005: conjectures at most

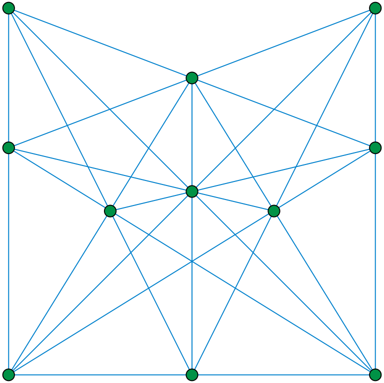
$$\frac{\pi n}{\sqrt{3}} + o(n) \approx 1.814n + o(n).$$

Open: Any upper bound better than $2n$

Orchard planting

Jackson 1821, Lloyd 1914:

How many three-point lines can we form from n points?



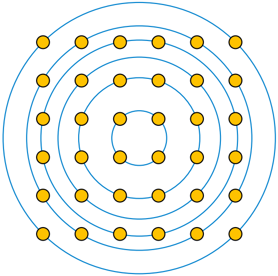
Burr et al. 1974: at least $\left\lfloor \frac{n(n-3)}{6} \right\rfloor + 1$

Green and Tao 2013: Burr et al. is optimal for large n

Onion layers

Repeatedly remove convex hull vertices

How many layers do you get?



Har-Peled and Lidický 2013: $n \times n$ grid has $\Theta(n^{4/3})$ layers

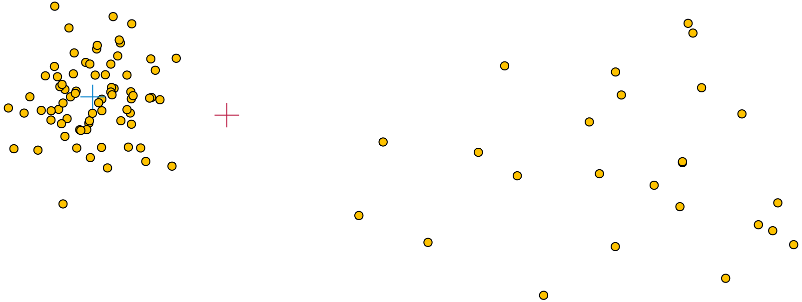
Conjecture (E, Har-Peled, Nivasch, ALENEX 2018): Layers of convex subsets of grid approximate the affine curve-shortening flow

Robust statistics

Depth(q) = minimum # points in a halfplane containing q

Deepest point = estimate of central location

More robust to outliers than centroid (Tukey 1975)

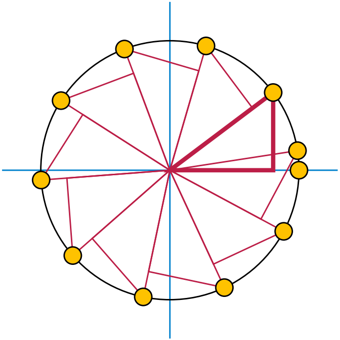


Deepest point in the plane: not invariant

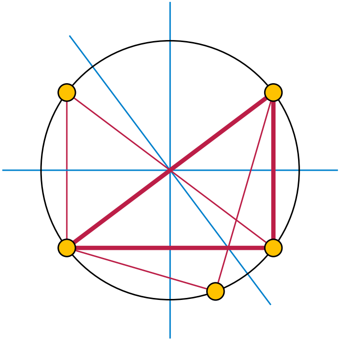
Deepest from a given point set: invariant and monotone

Realizability with rational distances

Euler (1862) proved: For every k there exists a convex k-gon with all pairwise distances rational



Rotate unit vector by the angle of a Pythagorean triangle



Reflect integer-sided triangle across perpendicular bisectors

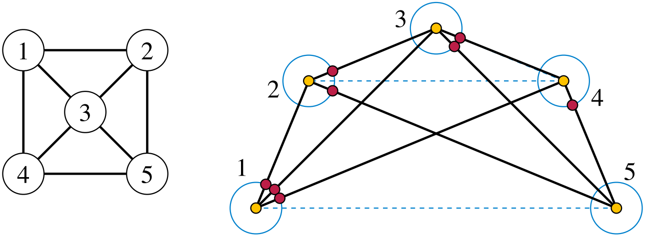
Open: are all rational-distance sets near-convex or near-linear?

III: Algorithmic perspectives

Finding a matching subconfiguration

Testing whether n points include a given k-point subconfiguration is NP-hard and W[1]-hard, and requires time n^{ck} for some $c > 0$ under standard assumptions (the exponential time hypothesis)

Reduction from clique-finding in graphs



Parameterized deletion to a hereditary property

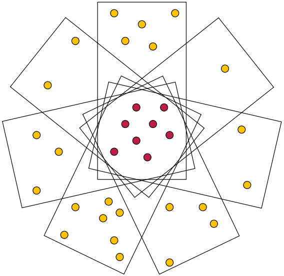
Removing k points so remaining subset has a property defined by $O(1)$ obstacles takes time $f(\mathrm{k}) \times$ polynomial(input size)

("fixed-parameter tractable")

Erdős–Rado sunflower lemma: many obstacles
$\Rightarrow$ big subfamily has equal pairwise intersects

Can safely ignore non-intersection points of all but $k+1$ of the obstacles in the family

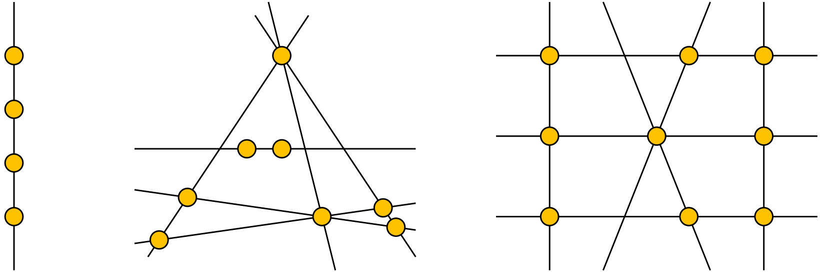
Repeat until few obstacles remain
$\Rightarrow$ small equivalent subproblem



Parameterized subsets with a hereditary property

Finding k points that have a property defined by given obstacles is Σ_2^{P}-complete for variable obstacles and k

Even when the obstacle set is fixed and k is a parameter it is not FPT – it requires time $n^{\Omega(k/\log k)}$ (under ETH)



Three obstacles that are hard to avoid

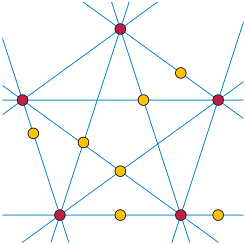
E & Lokshtanov [IPEC 2018] reduce from subgraph isomorphism

Special case of parameterized subsets: No-three-in-line

NP-hard and APX-hard
(reduction from independent set in bounded-degree graphs)

Fixed-parameter tractable in size of general-position subset
and approximable to within $O(\sqrt{n})$

Both based on principle: if G is a maximal subset in general position, whole set can be covered by $\binom{|G|}{2}$ lines

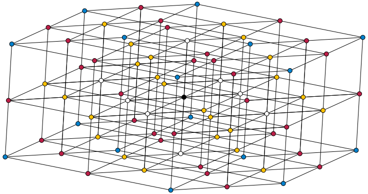


Open: Tighten the approximation gap

Tradeoff collinearity vs general position

Payne & Wood 2013: Every set of n points includes a subset of $\Omega(\sqrt{n/\log n})$ that is either collinear or in general position

Algorithm: eliminate points in unusually many collinear triples, then apply entropy compression (algorithmic LLL) to partition remaining points into few general-position subsets



Balogh & Solymosi 2018: $\exists$ sets with no four in line, largest general position subset $O(n^{5/6})$

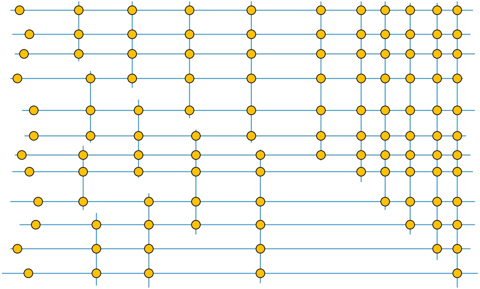
Open: What is the optimal tradeoff?

Approximate projective clustering

How many lines are needed to cover all points of a point set?

NP-complete [Megiddo and Tamir 1982]
and FPT [Langerman and Morin 2005]

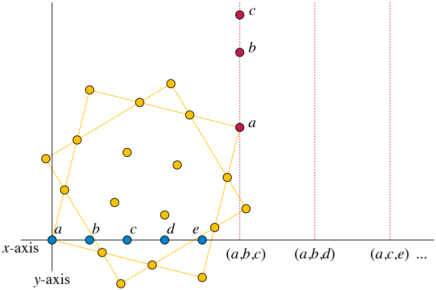
Greedy cover gives only logarithmic approximation



Open: Can we approximate it more accurately?

Partition into few general-position subsets

NP-complete even for two subsets



But # subsets $\geq \left\lceil \frac{\text{\# points on longest line}}{2} \right\rceil$

with equality if all points belong to few lines

$\Rightarrow$ Finding an optimal partition is FPT

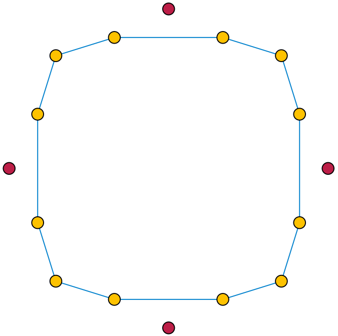
when parameterized by projective clustering

Property testing

Theorem: For obstacles of size $\leq s$ sampling $n^{1-1/s}$ points distinguishes obstacle-free sets from far-from-free sets w.h.p.

Sometimes $c > 0$ is necessary: $\Theta(n^{2/3})$ sample size for convexity [Csumaj, Sohler, & Ziegler, ESA 2000]

Sometimes $O(1)$ points suffice (e.g. projective clustering)

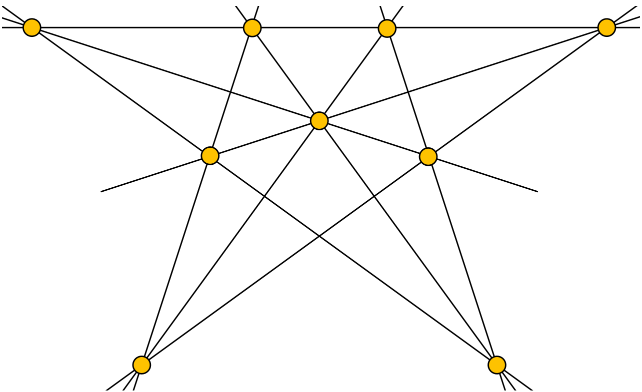


This set is $n/4$-far from convex but samples of $o(n^{2/3})$ points are w.h.p. convex

Realizability with integer coordinates

The Perles configuration (Perles, 1960s)

No combinatorially-equivalent set of points has integer coordinates



Open: Is integer realization decidable?

Conclusions

Monotonicity provides a unifying framework for many famous problems in algorithmic discrete geometry

Characterization by obstacles leads to algorithms of many types (exact, parameterized, approximation, property testing)

Much more remains to be done!