On Separating Points by Lines

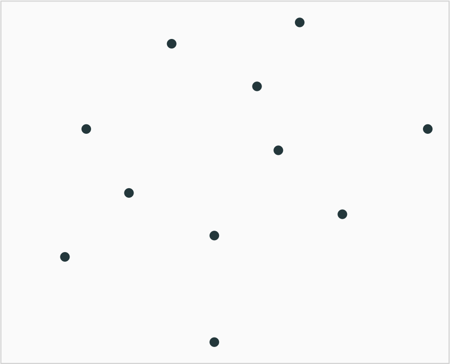
Sariel Har-Peled, Mitchell Jones

January 8, 2018

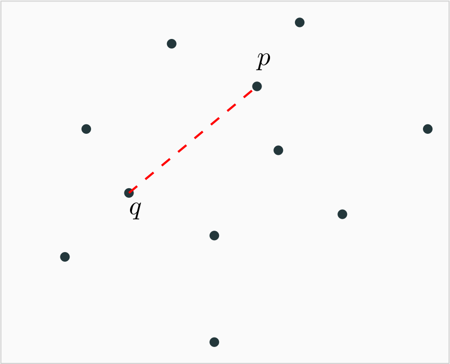
University of Illinois at Urbana-Champaign

Introduction & Motivation

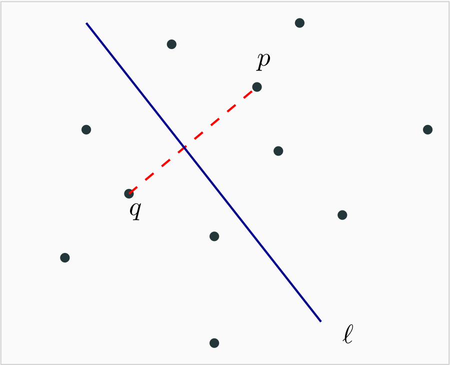
Definition: Separation



Definition: Separation



Definition: Separation



Problem

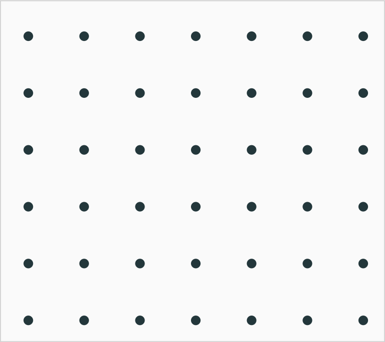
Let P be a set of points in $\mathbb{R}^2$. Find the minimum number of lines needed to separate all pairs of points in P, denoted by $\mathrm{sep}(P)$.

Problem

Let P be a set of points in $\mathbb{R}^2$. Find the minimum number of lines needed to separate all pairs of points in P, denoted by $\text{sep}(P)$.

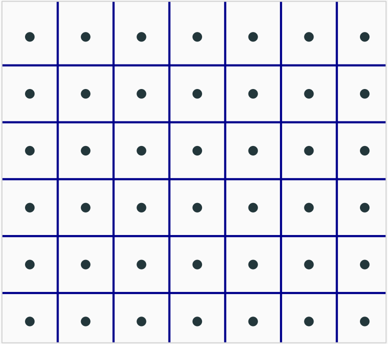
Known: **NP**-complete [Freimer-Mitchell-Piatko '91]

Example: Grid points



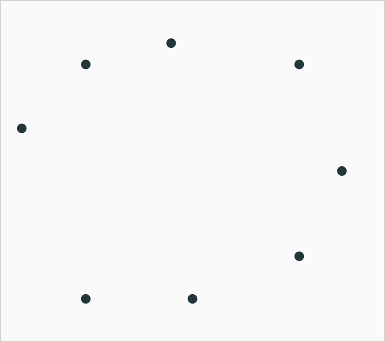
$\sqrt{n} \times \sqrt{n}$ grid

Example: Grid points



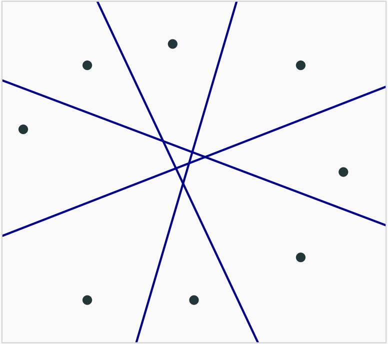
$\sqrt{n} \times \sqrt{n}$ grid $\Rightarrow O(\sqrt{n})$ lines

Example: Convex position



n points in
convex position

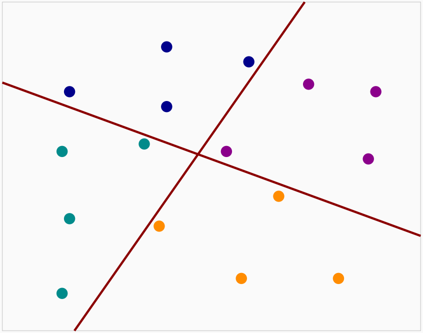
Example: Convex position



n points in convex position $\Rightarrow n/2$ lines

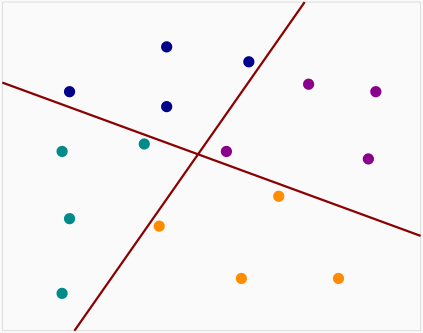
Motivation

- Ham Sandwich Theorem
- Cutting problems
  [Chazelle-Friedman '90]
- Partition problems
  [Matoušek '92]
- Polynomial partition problems
  [Agarwal-Matoušek-Sharir '13]
  Strong properties, less algorithmically convenient



Motivation

- Ham Sandwich Theorem
- Cutting problems
  [Chazelle-Friedman '90]
- Partition problems
  [Matoušek '92]
- Polynomial partition problems
  [Agarwal-Matoušek-Sharir '13]
  Strong properties, less algorithmically convenient
- What else can be can with lines/hyperplanes?



Bounds

Lemma

(1) $\mathrm{sep}(P) \leq \lceil n/2 \rceil$.

Bounds

Lemma

(1) $\mathrm{sep}(P) \leq \lceil n/2 \rceil$.

(2) $\mathrm{sep}(P) = \Omega(\sqrt{n})$.

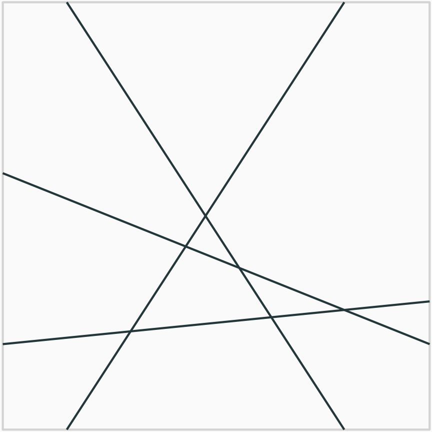
Bounds

Lemma

(1) $\text{sep}(P) \leq \lceil n/2 \rceil$.

(2) $\text{sep}(P) = \Omega(\sqrt{n})$.

- m lines



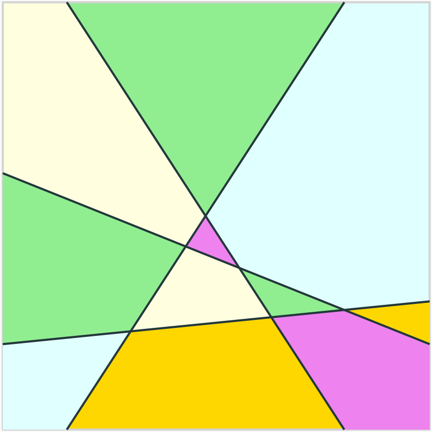
Bounds

Lemma

(1) $\mathrm{sep}(P) \leq \lceil n/2 \rceil$.

(2) $\mathrm{sep}(P) = \Omega(\sqrt{n})$.

- m lines
- $O(m^2)$ faces



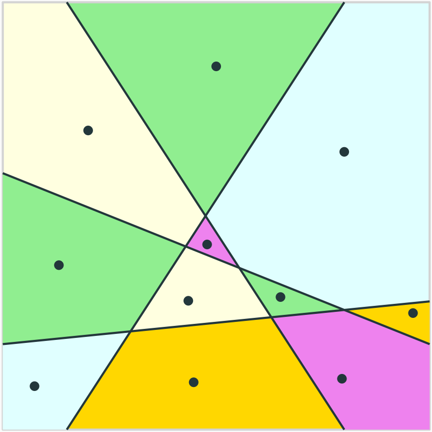
Bounds

Lemma

(1) $\text{sep}(P) \leq \lceil n/2 \rceil$.

(2) $\text{sep}(P) = \Omega(\sqrt{n})$.

- m lines
- $O(m^2)$ faces
- $n = O(m^2)$ points



Bounds

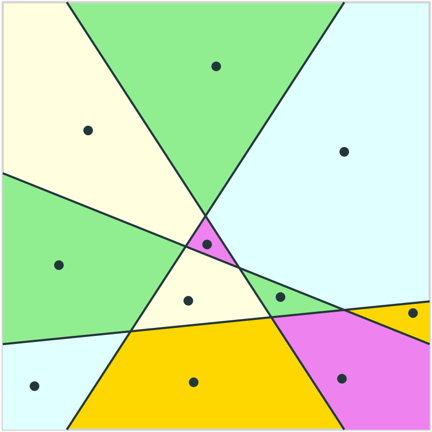
Lemma

(1) $\mathrm{sep}(P) \le \lceil n/2 \rceil$.

(2) $\mathrm{sep}(P) = \Omega(\sqrt{n})$.

- m lines
- $O(m^2)$ faces
- $n = O(m^2)$ points

$\implies m = \Omega(\sqrt{n})$



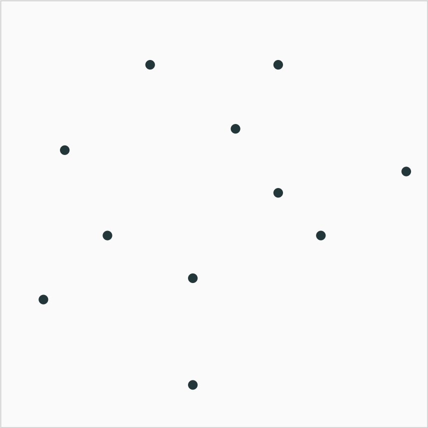
Separating random points

Results

Theorem

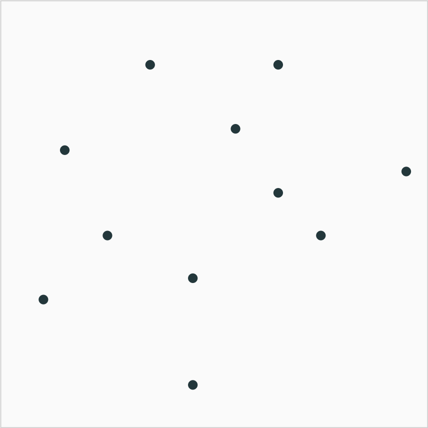
Let P be a set of n points chosen UAR from $[0, 1]^2$. With high probability, $\text{sep}(P) = \Omega(n^{2/3} \log \log n / \log n)$.

Expected number of lines: $O(n^{2/3})$



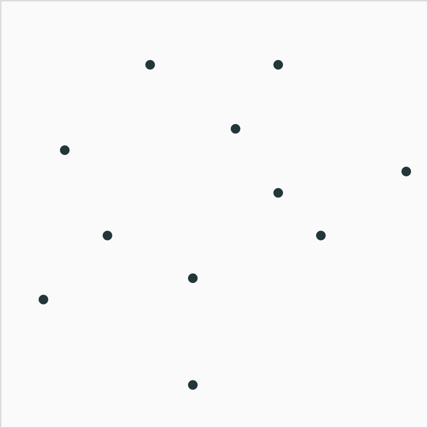
Expected number of lines: $O(n^{2/3})$

- Form a $n^{2/3} \times n^{2/3}$ grid



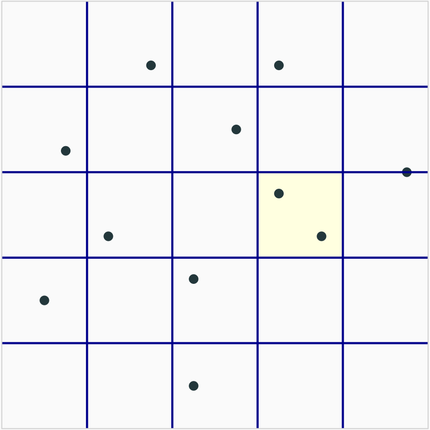
Expected number of lines: $O(n^{2/3})$

- Form a $n^{2/3} \times n^{2/3}$ grid
- Area of grid cell $= 1/n^{4/3}$



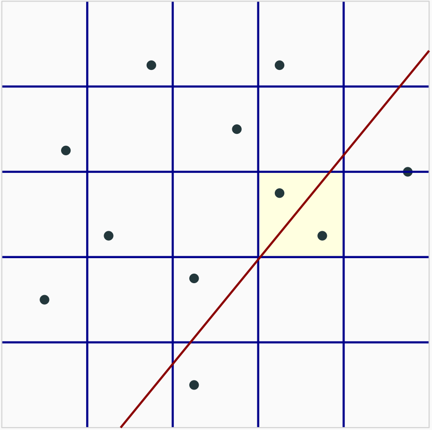
Expected number of lines: $O(n^{2/3})$

- Form a $n^{2/3} \times n^{2/3}$ grid
- Area of grid cell $= 1/n^{4/3}$
- Separate points in different grid cells using $O(n^{2/3})$ lines



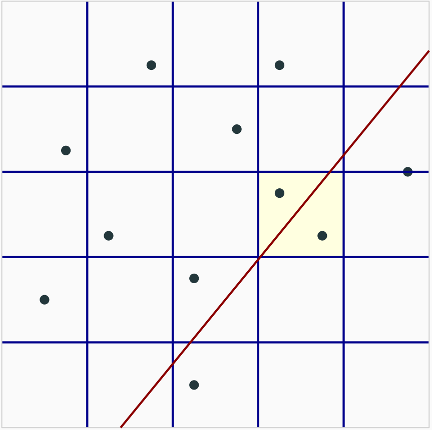
Expected number of lines: $O(n^{2/3})$

- Form a $n^{2/3} \times n^{2/3}$ grid
- Area of grid cell $= 1/n^{4/3}$
- Separate points in different grid cells using $O(n^{2/3})$ lines



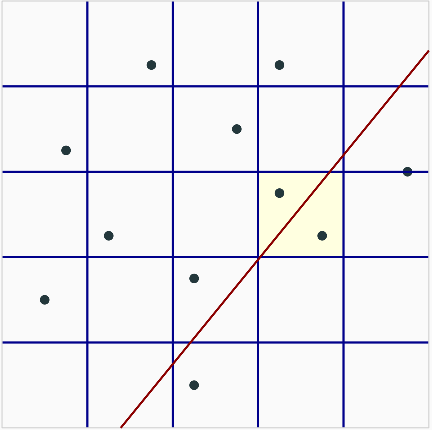
Expected number of lines: $O(n^{2/3})$

- Form a $n^{2/3} \times n^{2/3}$ grid
- Area of grid cell $= 1/n^{4/3}$
- Separate points in different grid cells using $O(n^{2/3})$ lines
- Expected number of colliding pairs is $\binom{n}{2}\frac{1}{n^{4/3}} = O(n^{2/3})$



Expected number of lines: $O(n^{2/3})$

- Form a $n^{2/3} \times n^{2/3}$ grid
- Area of grid cell $= 1/n^{4/3}$
- Separate points in different grid cells using $O(n^{2/3})$ lines
- Expected number of colliding pairs is $\binom{n}{2}\frac{1}{n^{4/3}} = O(n^{2/3})$
  $\implies O(n^{2/3})$ lines needed



Lower bound (sketch): $\Omega(n^{2/3} \log\log n / \log n)$

Interpret it as a balls (points) and bins (cells) problem on $T \times T$ grid, $T = n^{2/3}$.

1. How many **heavy** cells?
2. How many cells can a line intersect?
3. Of these cells, how many **heavy** cells?

Lower bound (sketch): $\Omega(n^{2/3} \log \log n / \log n)$

Interpret it as a balls (points) and bins (cells) problem on $T \times T$ grid, $T = n^{2/3}$.

1. How many **heavy** cells?
2. How many cells can a line intersect?
3. Of these cells, how many **heavy** cells?

Lower bound (sketch): $\Omega(n^{2/3} \log \log n / \log n)$

Interpret it as a balls (points) and bins (cells) problem on $T \times T$ grid, $T = n^{2/3}$.

1. How many **heavy** cells? $\Theta(n^{2/3})$
2. How many cells can a line intersect?
3. Of these cells, how many **heavy** cells?

Lower bound (sketch): $\Omega(n^{2/3} \log\log n / \log n)$

Interpret it as a balls (points) and bins (cells) problem on $T \times T$ grid, $T = n^{2/3}$.

1. How many **heavy** cells? $\Theta(n^{2/3})$
2. How many cells can a line intersect?
3. Of these cells, how many **heavy** cells?

Lower bound (sketch): $\Omega(n^{2/3} \log\log n / \log n)$

Interpret it as a balls (points) and bins (cells) problem on $T \times T$ grid, $T = n^{2/3}$.

1. How many **heavy** cells? $\Theta(n^{2/3})$
2. How many cells can a line intersect? $\leq 2T$
3. Of these cells, how many **heavy** cells?

Lower bound (sketch): $\Omega(n^{2/3} \log\log n / \log n)$

Interpret it as a balls (points) and bins (cells) problem on $T \times T$ grid, $T = n^{2/3}$.

1. How many **heavy** cells? $\Theta(n^{2/3})$
2. How many cells can a line intersect? $\leq 2T$
3. Of these cells, how many **heavy** cells?

Lower bound (sketch): $\Omega(n^{2/3} \log\log n / \log n)$

Interpret it as a balls (points) and bins (cells) problem on $T \times T$ grid, $T = n^{2/3}$.

1. How many **heavy** cells? $\Theta(n^{2/3})$
2. How many cells can a line intersect? $\leq 2T$
3. Of these cells, how many **heavy** cells? $O(\log n / \log\log n)$

Lower bound (sketch): $\Omega(n^{2/3} \log\log n / \log n)$

Interpret it as a balls (points) and bins (cells) problem on $T \times T$ grid, $T = n^{2/3}$.

1. How many **heavy** cells? $\Theta(n^{2/3})$
2. How many cells can a line intersect? $\leq 2T$
3. Of these cells, how many **heavy** cells? $O(\log n / \log\log n)$

Lower bound (sketch): $\Omega(n^{2/3} \log\log n / \log n)$

Interpret it as a balls (points) and bins (cells) problem on $T \times T$ grid, $T = n^{2/3}$.

1. How many **heavy** cells? $\Theta(n^{2/3})$
2. How many cells can a line intersect? $\leq 2T$
3. Of these cells, how many **heavy** cells? $O(\log n / \log\log n)$

   $\Longrightarrow \Omega(n^{2/3} \log\log n / \log n)$ lines needed

Higher dimensions

Theorem

Let P be a set of n points chosen UAR from $[0,1]^d$. With high probability, the minimum number of hyperplanes separating P is $\Omega(n^{2/(d+1)} \log\log n / \log n)$.

In expectation, one needs $O(dn^{2/(d+1)})$ separating hyperplanes.

Approximating the minimum separating set

A slightly weaker result

Lemma

Let P be a set of points in $\mathbb{R}^2$ and $\mathrm{OPT} := \mathrm{sep}(P)$.

There is an algorithm that finds set of separating set of lines of size $O(\mathrm{OPT} \log \mathrm{OPT})$, expected running time $O(n^2 \mathrm{OPT} \log \mathrm{OPT})$.

A slightly weaker result

Lemma

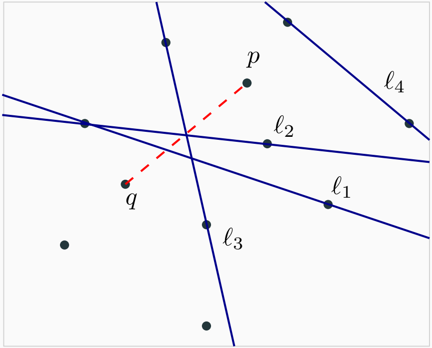
Let P be a set of points in $\mathbb{R}^2$ and $\mathsf{OPT} := \mathsf{sep}(P)$.

There is an algorithm that finds set of separating set of lines of size $O(\mathsf{OPT} \log \mathsf{OPT})$, expected running time $O(n^2 \mathsf{OPT} \log \mathsf{OPT})$.

Known: 2-approximation when separating lines are axis-parallel [Calinescu-Dumitrescu-Karloff-Wan '05]

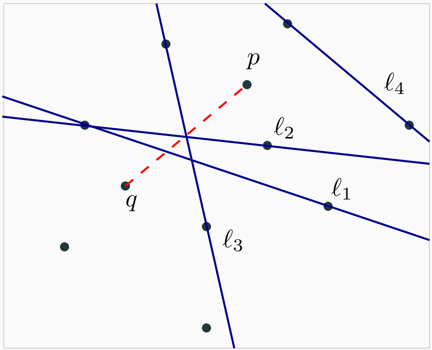
First algorithm with Hitting sets

- Suffices to consider lines passing through pairs of points



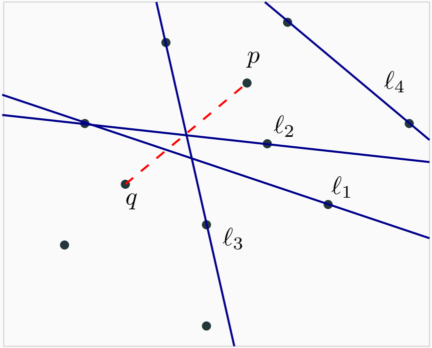
First algorithm with Hitting sets

- Suffices to consider lines passing through pairs of points
- Generate $O(n^2)$ lines ($\mathcal{C}$)



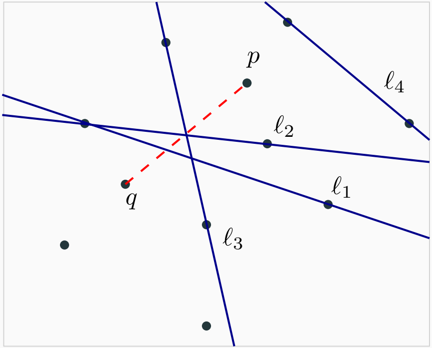
First algorithm with Hitting sets

- Suffices to consider lines passing through pairs of points
- Generate $O(n^2)$ lines ($\mathcal{C}$)
- For each segment pq, determine all lines intersecting pq (L_{pq})



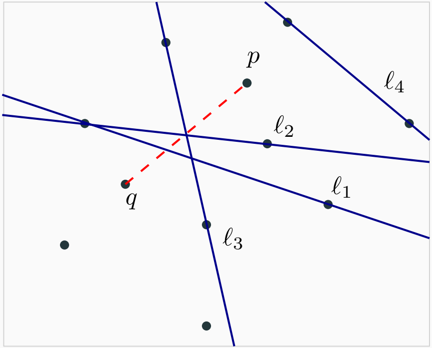
First algorithm with Hitting sets

- Suffices to consider lines passing through pairs of points
- Generate $O(n^2)$ lines ($\mathcal{C}$)
- For each segment pq, determine all lines intersecting pq (L_{pq})
- Repeat for all segments, $\mathcal{L} = \{L_{pq}\}_{p,q \in P}$

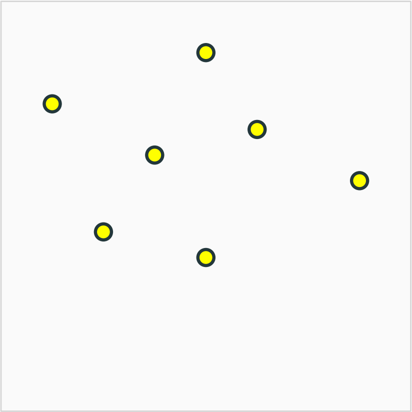


First algorithm with Hitting sets

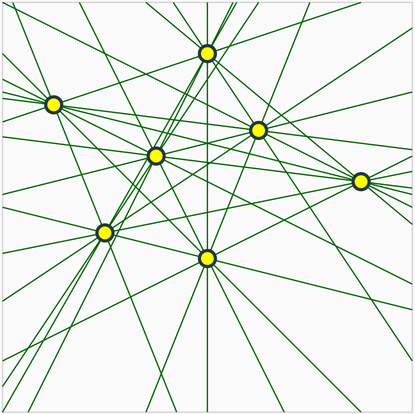
- Suffices to consider lines passing through pairs of points
- Generate $O(n^2)$ lines ($\mathcal{C}$)
- For each segment pq, determine all lines intersecting pq (L_{pq})
- Repeat for all segments, $\mathcal{L} = \{L_{pq}\}_{p,q \in P}$ $\implies$ $(\mathcal{C}, \mathcal{L})$ hitting set instance (finite VC dimension)



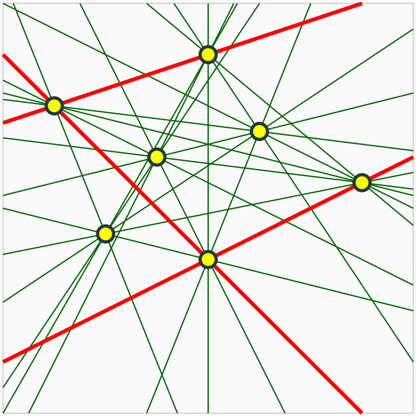
First algorithm



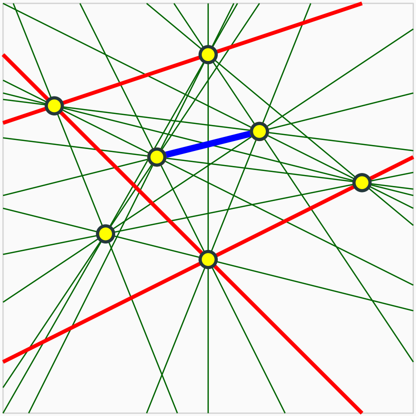
First algorithm



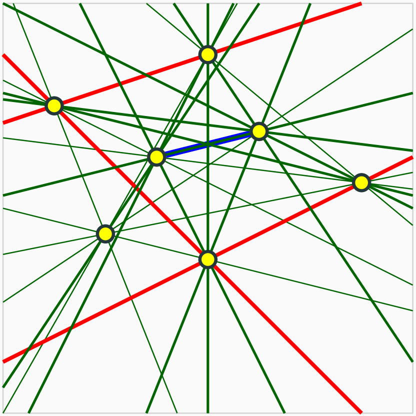
First algorithm



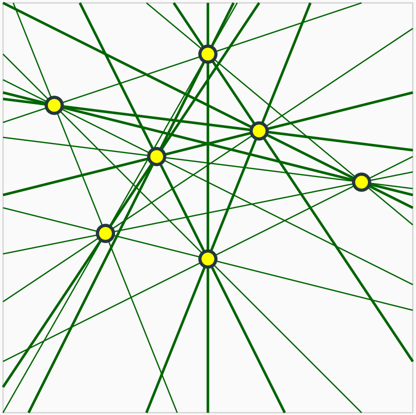
First algorithm



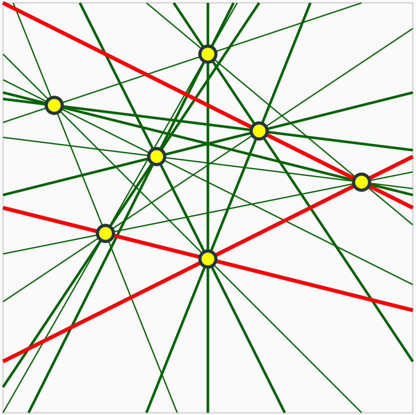
First algorithm



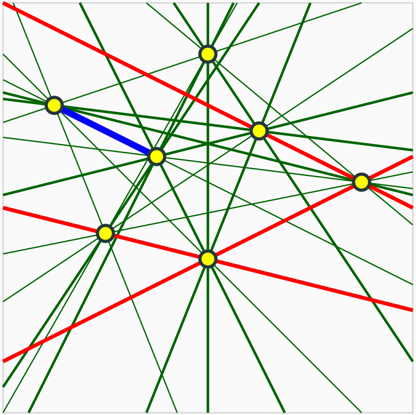
First algorithm



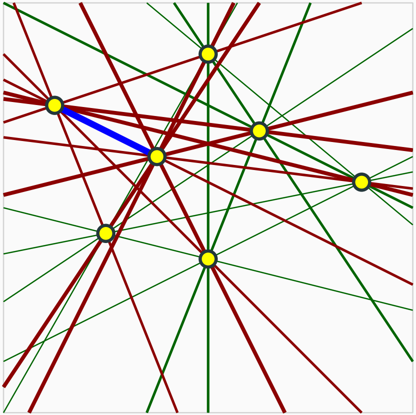
First algorithm



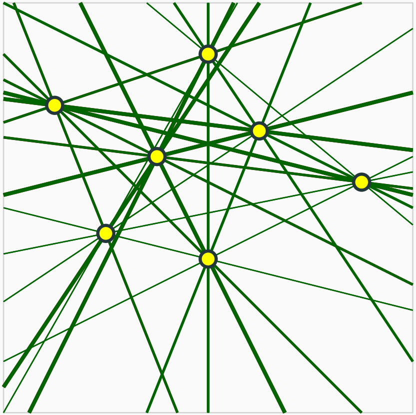
First algorithm



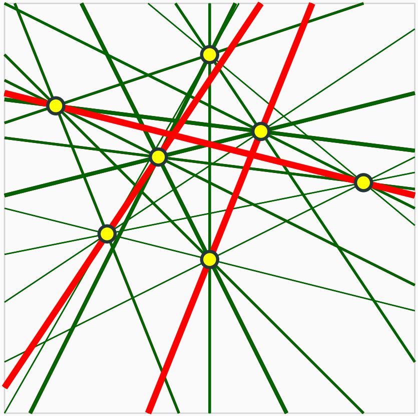
First algorithm



First algorithm



First algorithm



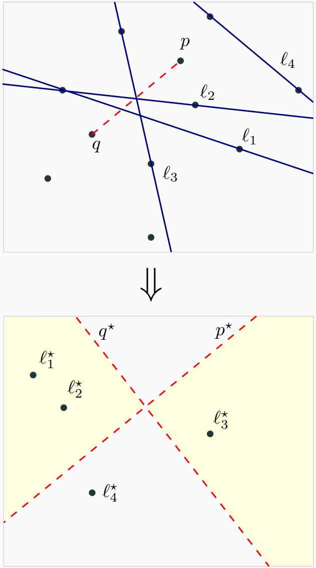
Analysis

Lemma

Algorithm returns a set of separating lines of size $O(\mathrm{OPT} \log \mathrm{OPT})$, expected running time $O(n^2 \mathrm{OPT} \log \mathrm{OPT})$.

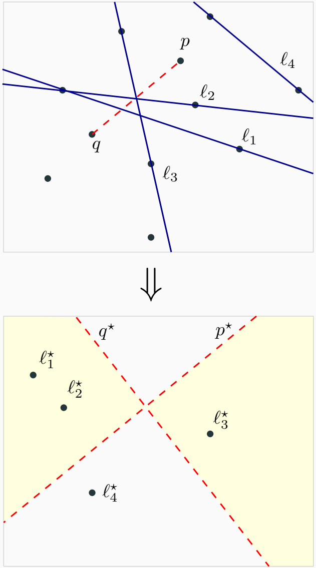
Improvement

- Bottleneck is maintaining weights of $O(n^2)$ lines



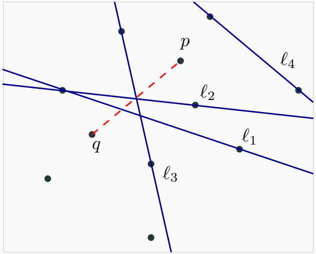
Improvement

- Bottleneck is maintaining weights of $O(n^2)$ lines
- Use duality (lines $\rightarrow$ points, segments $\rightarrow$ wedges)

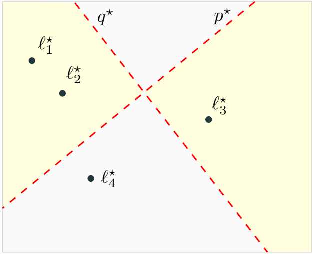


Improvement

- Bottleneck is maintaining weights of $O(n^2)$ lines
- Use duality (lines $\rightarrow$ points, segments $\rightarrow$ wedges)
- Maintain weights as they are updated



$\Downarrow$



Improvement

Theorem

Improved algorithm returns a set of separating lines of size $O(\mathrm{OPT} \log \mathrm{OPT})$, expected running time $O(n^{2/3}\mathrm{OPT}^{5/3} \log^{O(1)} n)$.

Improvement

Theorem

Improved algorithm returns a set of separating lines of size $O(\text{OPT} \log \text{OPT})$, expected running time $O(n^{2/3}\text{OPT}^{5/3} \log^{O(1)} n)$.

- $\text{sep}(P) = O(\sqrt{n}) \implies O(n^{3/2} \log^{O(1)} n)$

Improvement

Theorem

Improved algorithm returns a set of separating lines of size $O(\mathrm{OPT} \log \mathrm{OPT})$, expected running time $O(n^{2/3}\mathrm{OPT}^{5/3} \log^{O(1)} n)$.

- $\mathrm{sep}(P) = O(\sqrt{n}) \implies O(n^{3/2} \log^{O(1)} n)$
- $\mathrm{sep}(P) = O(n) \implies O(n^{7/3} \log^{O(1)} n)$

Summary

- For n random points in $[0,1]^2$, expected $O(n^{2/3})$ lines needed
- With high probability, $\Omega(n^{2/3} \log\log n / \log n)$ lines needed
- Can compute a separating set of size $O(\mathrm{OPT} \log \mathrm{OPT})$ in time $\widetilde{O}(n^{2/3}\mathrm{OPT}^{5/3})$

Summary

- For n random points in $[0,1]^2$, expected $O(n^{2/3})$ lines needed
- With high probability, $\Omega(n^{2/3} \log\log n / \log n)$ lines needed
- Can compute a separating set of size $O(\mathrm{OPT} \log \mathrm{OPT})$ in time $\widetilde{O}(n^{2/3}\mathrm{OPT}^{5/3})$

Thank you!