Finding Haystacks (and Other Structures) in Geometry

Sariel Har-Peled

[2]UIUC, Illinois, USA

May 25, 2010

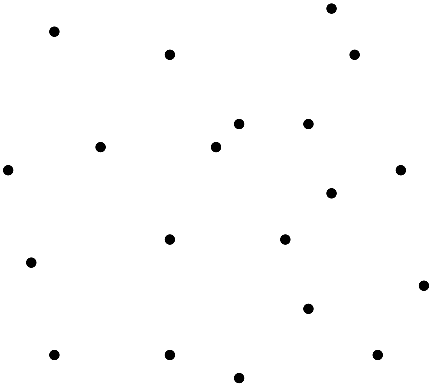
2: Overview

1. Small subsets of the input capture the structure.
2. Example
3. Bedrock for efficient geometric algorithms.
4. In this talk:
    1. Survey of such concepts.
    2. Why such subsets exist?
    3. Some applications.

3: Motivating problem

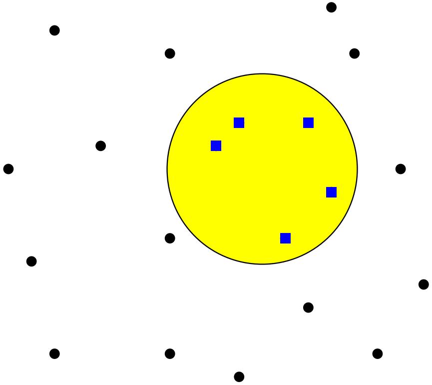
Placing an antenna.

P: Set of **n** points (customers)
Q: Find location of antenna that serves maximum number of customers.

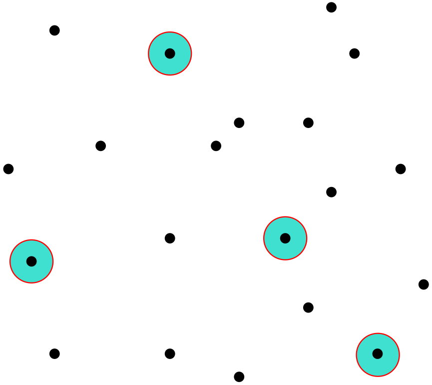


P: Set of **n** points (customers)

Q: Find location of antenna that serves maximum number of customers.



P: Set of **n** points (customers)
Q: Find location of antenna that serves maximum number of customers.



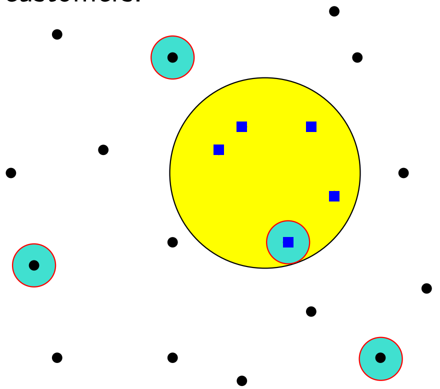
Approach...

- Pick a random sample.
- Count sample points.

3: Motivating problem

Placing an antenna.

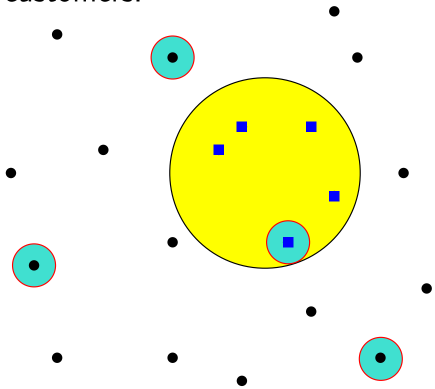
P: Set of **n** points (customers)
Q: Find location of antenna that serves maximum number of customers.



Approach...

- Pick a random sample.
- Count sample points.

P: Set of **n** points (customers)
Q: Find location of antenna that serves maximum number of customers.



Approach...

- Pick a random sample.
- Count sample points.

Challenge...

- Good for all disks.
- ... Infinite number of disks.

$\mathbf{P}$: Set of points in the plane.

Measure

For any disk $\mathbf{D}$ its **measure** is $\overline{\mathbf{P}}(\mathbf{D}) = \dfrac{|\mathbf{D} \cap \mathbf{P}|}{|\mathbf{P}|}$.

ε-Sample

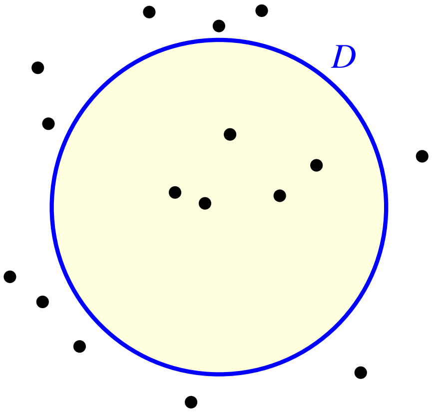
$\varepsilon > \mathbf{0}$: Approx parameter.

$\mathbf{S} \subseteq \mathbf{P}$ is ε-**sample** if

$$\forall \mathbf{D} \quad \left| \overline{\mathbf{P}}(\mathbf{D}) - \overline{\mathbf{S}}(\mathbf{D}) \right| < \varepsilon$$

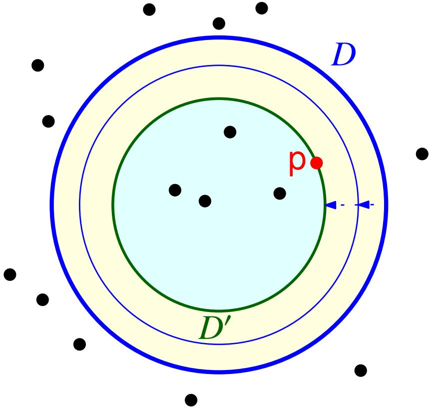
5: Bounding infinity...

...or, it just looks infinite.



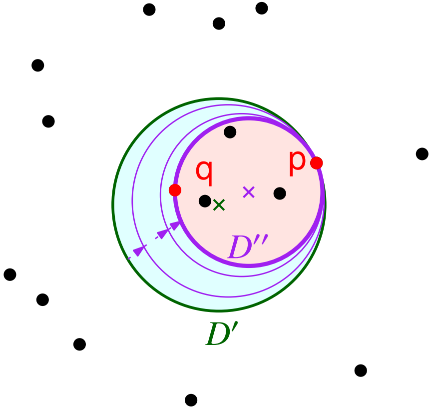
5: Bounding infinity...

...or, it just looks infinite.



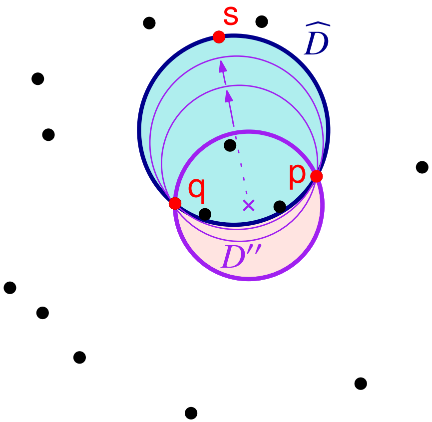
5: Bounding infinity...

...or, it just looks infinite.



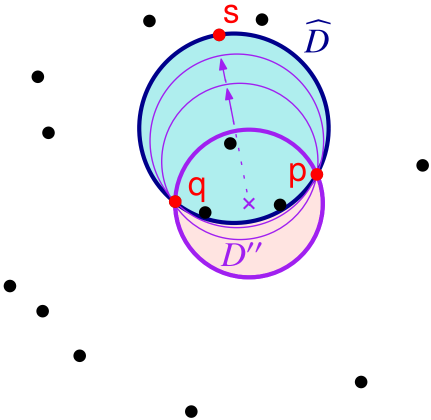
5: Bounding infinity...

...or, it just looks infinite.



5: Bounding infinity...

...or, it just looks infinite.



As such...

n: number of points.
Only $\mathbf{O(n^3)}$ different subsets.

Conclusion: ε-Sample

Std. random sampling...
$\mathbf{O}\left(\frac{\log n}{\varepsilon^2}\right)$.

Known: ε-Sample

[Vapnik and Chervonenkis, 1971] **[Li et al., 2001]** $\mathbf{O}\left(\frac{1}{\varepsilon^2}\right)$.

6: Application

...or why size matters.

Problem

P: **n** points in the plane
Compute: Smallest disk containing half the points.

min_disk$(\mathbf{P}, \mathbf{n/2})$: $\mathbf{O(n^2)}$ time exact algorithm.

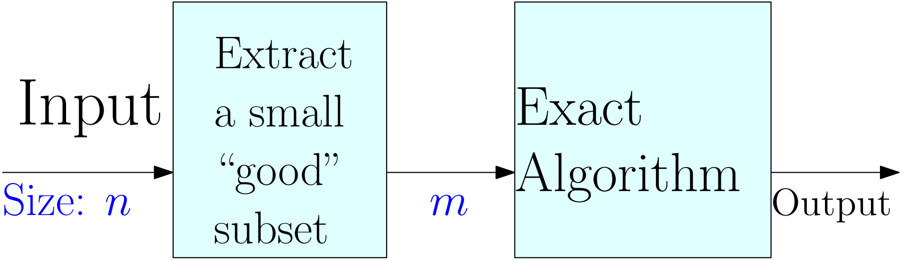
Approximation algorithm

$\varepsilon > \mathbf{0}$: Approx parameter.
$\mathbf{S} \subseteq \mathbf{P}$: ε-Sample of size $\mathbf{m}$.
Return **min_disk**$(\mathbf{S}, (\mathbf{1} - \varepsilon)\mathbf{m/2})$.

Ret disk has $\geq (\mathbf{1/2} - \mathbf{2}\varepsilon)\mathbf{n}$ points and smaller than exact answer.

7: Application

...or why size matters.



Exact Running time: $T_{exact}(n)$

New : $T_{extract}(n) + T_{exact}(m)$

Example: Min disk containing $n/2$ points

$T_{extract}(n) = O(m)$, $T_{exact}(n) = O(n^2)$, $m = 1/\varepsilon^2$.

Running time: $O(1/\varepsilon^4)$.

8: Sketch of a sketch is a sketch

...or how many sketches would a sketch sketch, if a sketch could sketch sketches?

Lemma (Sketch property.)

$\mathbf{X} \subset \mathbf{Y}$: δ-sample.
$\mathbf{Y} \subset \mathbf{Z}$: δ'-sample.
$\implies$ $\mathbf{X}$ is a $(\delta + \delta')$-sample of $\mathbf{Z}$.

$\mathbf{Z} = \mathbf{P}$: Set of $\mathbf{n}$ points
$\exists \mathbf{Y} \subset \mathbf{Z}$: ε-Sample $|\mathbf{Y}| = \mathbf{O}\left(\frac{\log n}{\varepsilon^2}\right)$.
$\exists \mathbf{X} \subset \mathbf{Y}$: ε-Sample $|\mathbf{X}| = \mathbf{O}\left(\frac{\log \log n}{\varepsilon^2}\right)$.
$\implies$
$\mathbf{X}$ is 2ε-sample of $\mathbf{P}$ of size $\mathbf{O}\left(\frac{\log \log n}{\varepsilon^2}\right)$.

Theorem [Vapnik and Chervonenkis, 1971]

Exists ε-sample of size $\mathbf{O}\left(\frac{1}{\varepsilon^2} \log \frac{1}{\varepsilon}\right)$.

8: Sketch of a sketch is a sketch

...or how many sketches would a sketch sketch, if a sketch could sketch sketches?

Lemma (Sketch property.)

$\mathbf{X} \subset \mathbf{Y}$: δ-sample.
$\mathbf{Y} \subset \mathbf{Z}$: δ'-sample.
$\implies$ $\mathbf{X}$ is a $(\delta + \delta')$-sample of $\mathbf{Z}$.

$\mathbf{Z} = \mathbf{P}$: Set of $\mathbf{n}$ points
$\exists \mathbf{Y} \subset \mathbf{Z}$: ε-Sample $|\mathbf{Y}| = \mathbf{O}\left(\frac{\log n}{\varepsilon^2}\right)$.
$\exists \mathbf{X} \subset \mathbf{Y}$: ε-Sample $|\mathbf{X}| = \mathbf{O}\left(\frac{\log \log n}{\varepsilon^2}\right)$.
$\implies$
$\mathbf{X}$ is 2ε-sample of $\mathbf{P}$ of size $\mathbf{O}\left(\frac{\log \log n}{\varepsilon^2}\right)$.

Theorem [Vapnik and Chervonenkis, 1971]

Exists ε-sample of size $\mathbf{O}\left(\frac{1}{\varepsilon^2} \log \frac{1}{\varepsilon}\right)$.

8: Sketch of a sketch is a sketch

...or how many sketches would a sketch sketch, if a sketch could sketch sketches?

Lemma (Sketch property.)

$\mathbf{X} \subset \mathbf{Y}$: δ-sample.
$\mathbf{Y} \subset \mathbf{Z}$: δ'-sample.
$\implies$ $\mathbf{X}$ is a $(\delta + \delta')$-sample of $\mathbf{Z}$.

$\mathbf{Z} = \mathbf{P}$: Set of $\mathbf{n}$ points
$\exists \mathbf{Y} \subset \mathbf{Z}$: ε-Sample $|\mathbf{Y}| = \mathbf{O}\left(\frac{\log n}{\varepsilon^2}\right)$.
$\exists \mathbf{X} \subset \mathbf{Y}$: ε-Sample $|\mathbf{X}| = \mathbf{O}\left(\frac{\log \log n}{\varepsilon^2}\right)$.
$\implies$
$\mathbf{X}$ is 2ε-sample of $\mathbf{P}$ of size $\mathbf{O}\left(\frac{\log \log n}{\varepsilon^2}\right)$.

Theorem [Vapnik and Chervonenkis, 1971]

Exists ε-sample of size $\mathbf{O}\left(\frac{1}{\varepsilon^2} \log \frac{1}{\varepsilon}\right)$.

Lemma (Sketch property.)

$\mathbf{X} \subset \mathbf{Y}$: δ-sample.
$\mathbf{Y} \subset \mathbf{Z}$: δ'-sample.
$\implies$ $\mathbf{X}$ is a $(\delta + \delta')$-sample of $\mathbf{Z}$.

$\mathbf{Z} = \mathbf{P}$: Set of $\mathbf{n}$ points
$\exists \mathbf{Y} \subset \mathbf{Z}$: ε-Sample $|\mathbf{Y}| = \mathbf{O}\left(\frac{\log n}{\varepsilon^2}\right)$.
$\exists \mathbf{X} \subset \mathbf{Y}$: ε-Sample $|\mathbf{X}| = \mathbf{O}\left(\frac{\log \log n}{\varepsilon^2}\right)$.
$\implies$
$\mathbf{X}$ is 2ε-sample of $\mathbf{P}$ of size $\mathbf{O}\left(\frac{\log \log n}{\varepsilon^2}\right)$.

Theorem [Vapnik and Chervonenkis, 1971]

Exists ε-sample of size $\mathbf{O}\left(\frac{1}{\varepsilon^2} \log \frac{1}{\varepsilon}\right)$.

9: Netting a lot with a little

or, getting less with even less

Smaller ε-samples are better.

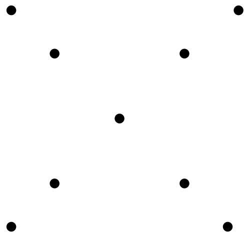
P: **n** points in the plane.

ε-Net

$\varepsilon > 0$: Approx parameter.

$S \subseteq P$ is ε-**net** if

$$\forall D \quad \left|\overline{P}(D)\right| \geq \varepsilon \implies S \cap D \neq \emptyset.$$

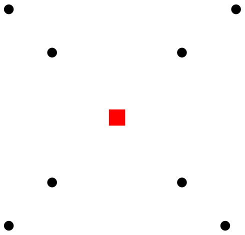


Smaller ε-samples are better.
P: **n** points in the plane.

ε-Net

$\varepsilon > 0$: Approx parameter.
S $\subseteq$ **P** is ε-**net** if

$$\forall D \quad \left|\overline{P}(D)\right| \geq \varepsilon \implies S \cap D \neq \emptyset.$$



9: Netting a lot with a little

or, getting less with even less

Smaller ε-samples are better.

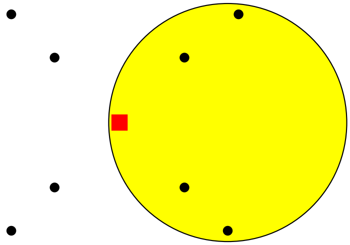
P: **n** points in the plane.

ε-Net

$\varepsilon > \mathbf{0}$: Approx parameter.

$\mathbf{S} \subseteq \mathbf{P}$ is ε-**net** if

$$\forall \mathbf{D} \quad \left|\overline{\mathbf{P}}(\mathbf{D})\right| \geq \varepsilon \implies \mathbf{S} \cap \mathbf{D} \neq \emptyset.$$



10: ε-Sample vs. ε-net

$\mathbf{P}$: $\mathbf{n}$ points in the plane.
$\varepsilon > \mathbf{0}$: Approx parameter.

ε-Net

$\mathbf{S} \subseteq \mathbf{P}$ is ε-**net** if $\quad \forall \mathbf{D} \quad \left|\overline{\mathbf{P}}(\mathbf{D})\right| \geq \varepsilon \implies \mathbf{S} \cap \mathbf{D} \neq \emptyset$.

ε-Sample

$\mathbf{S} \subseteq \mathbf{P}$ is ε-**sample** if $\quad \forall \mathbf{D} \quad \left|\overline{\mathbf{P}}(\mathbf{D}) - \overline{\mathbf{S}}(\mathbf{D})\right| < \varepsilon$.

ε-sample is an ε-net.

Theorem (ε-net theorem)

[Haussler and Welzl, 1987]
Random sample of size $\mathbf{O}\left(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon}\right)$ is an ε-net.

10: ε-Sample vs. ε-net

P: **n** points in the plane.
$\varepsilon > 0$: Approx parameter.

ε-Net

$\mathsf{S} \subseteq \mathsf{P}$ is ε-**net** if $\quad \forall \mathsf{D} \quad \left|\overline{\mathsf{P}}(\mathsf{D})\right| \geq \varepsilon \implies \mathsf{S} \cap \mathsf{D} \neq \emptyset$.

ε-Sample

$\mathsf{S} \subseteq \mathsf{P}$ is ε-**sample** if $\quad \forall \mathsf{D} \quad \left|\overline{\mathsf{P}}(\mathsf{D}) - \overline{\mathsf{S}}(\mathsf{D})\right| < \varepsilon$.

ε-sample is an ε-net.

Theorem (ε-net theorem)

[Haussler and Welzl, 1987]
Random sample of size $\mathsf{O}\left(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}\right)$ *is an* ε*-net.*

10: ε-Sample vs. ε-net

P: **n** points in the plane.
$\varepsilon > 0$: Approx parameter.

ε-Net

$\mathbf{S} \subseteq \mathbf{P}$ is ε-**net** if $\forall \mathbf{D} \quad \left|\overline{\mathbf{P}}(\mathbf{D})\right| \geq \varepsilon \implies \mathbf{S} \cap \mathbf{D} \neq \emptyset$.

ε-Sample

$\mathbf{S} \subseteq \mathbf{P}$ is ε-**sample** if $\forall \mathbf{D} \quad \left|\overline{\mathbf{P}}(\mathbf{D}) - \overline{\mathbf{S}}(\mathbf{D})\right| < \varepsilon$.

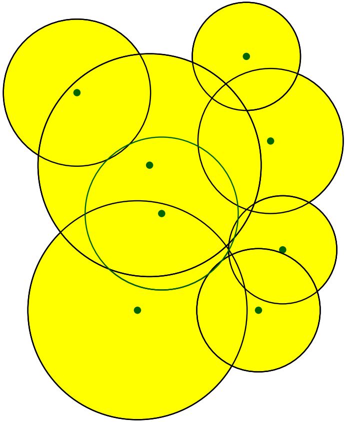
ε-sample is an ε-net.

Theorem (ε-net theorem)

[Haussler and Welzl, 1987]
Random sample of size $\mathbf{O}\left(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon}\right)$ *is an* ε*-net.*

11: Geometric Hitting Set

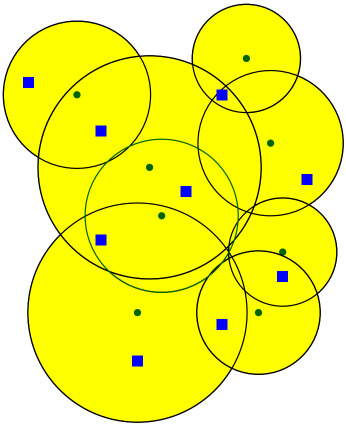
=Set cover



- $\mathcal{F}$: **m** disks in the plane. (Reception areas)
- **P**: possible loc for antennas.
- **Q**: Find min # of antennas serving all $\implies$ Hitting set.
- **Q**: Min size $\mathbf{X} \subseteq \mathbf{P}$ stabbing/hitting all disk in $\mathcal{F}$.

Observation

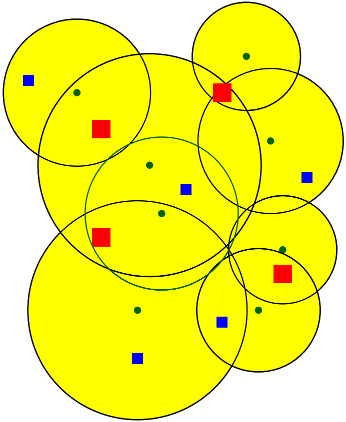
An instance of hitting set = set cover.



- $\mathcal{F}$: **m** disks in the plane. (Reception areas)
- **P**: possible loc for antennas.
- **Q**: Find min # of antennas serving all
  $\implies$ Hitting set.
- **Q**: Min size $\mathbf{X} \subseteq \mathbf{P}$ stabbing/hitting all disk in $\mathcal{F}$.

Observation

An instance of hitting set = set cover.



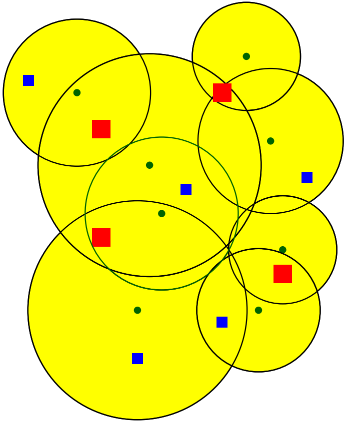
- $\mathcal{F}$: **m** disks in the plane. (Reception areas)
- **P**: possible loc for antennas.
- **Q**: Find min # of antennas serving all
  $\implies$ Hitting set.
- **Q**: Min size $\mathbf{X} \subseteq \mathbf{P}$ stabbing/hitting all disk in $\mathcal{F}$.

Observation

An instance of hitting set = set cover.

11: Geometric Hitting Set

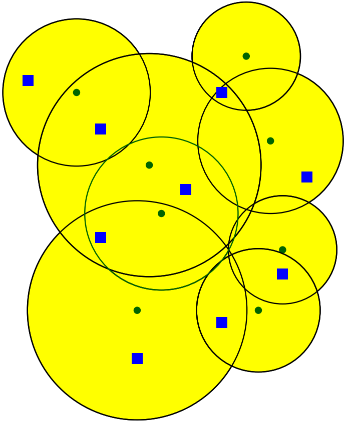
=Set cover



- $\mathcal{F}$: **m** disks in the plane. (Reception areas)
- **P**: possible loc for antennas.
- **Q**: Find min # of antennas serving all
  $\implies$ Hitting set.
- **Q**: Min size $\mathbf{X} \subseteq \mathbf{P}$ stabbing/hitting all disk in $\mathcal{F}$.

Observation

An instance of hitting set = set cover.



- $\mathcal{F}$: **m** of disks in the plane.
- **P**: set of points
- Every disk $\mathbf{D} \in \mathcal{F}$ corresponds to the subset $\mathbf{P} \cap \mathbf{D}$.
- Induced **set system**:

$$\left(\mathbf{P}, \left\{ \mathbf{P} \cap \mathbf{D} \;\middle|\; \mathbf{D} \in \mathcal{F} \right\}\right).$$

13: Hitting set via LP relaxation

$\mathsf{V} = (\mathsf{P}, \mathcal{F})$: geometric set system
$\mathsf{P} = \{\mathsf{p}_1, \ldots, \mathsf{p}_n\}$
$\mathcal{F} = \{\mathsf{D}_1, \ldots, \mathsf{D}_m\}$

IP for hitting set.

$$\min \quad \tau = \sum_{i=1}^{n} x_i$$

$$\text{s.t.} \sum_{p_i \in D} x_i \geq 1 \qquad \forall D \in \mathcal{F}$$

$$x_i \in \{0, 1\} \qquad \forall p_i \in P$$

LP relax' of hitting set.

$$\min \quad \mathrm{Opt} = \sum_{i=1}^{n} x_i$$

$$\text{s.t.} \sum_{p_i \in D} x_i \geq 1 \quad \forall D \in \mathcal{F}$$

$$x_i \geq 0 \quad \forall p_i \in P$$

13: Hitting set via LP relaxation

$\mathbf{V} = (\mathbf{P}, \mathcal{F})$: geometric set system
$\mathbf{P} = \{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$
$\mathcal{F} = \{\mathbf{D}_1, \ldots, \mathbf{D}_m\}$

IP for hitting set.

$$\min \quad \tau = \sum_{i=1}^{n} x_i$$
$$\text{s.t.} \sum_{p_i \in D} x_i \geq 1 \qquad \forall D \in \mathcal{F}$$
$$x_i \in \{0, 1\} \qquad \forall p_i \in P$$

LP relax' of hitting set.

$$\min \quad \mathrm{Opt} = \sum_{i=1}^{n} x_i$$
$$\text{s.t.} \sum_{p_i \in D} x_i \geq 1 \quad \forall D \in \mathcal{F}$$
$$x_i \geq 0 \quad \forall p_i \in P$$

$\mathbf{V} = (\mathbf{P}, \mathcal{F})$: set sys' VC dim $\mathbf{d}$ $\quad$ $\mathbf{P} = \{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$
$\mathcal{F} = \{\mathbf{D}_1, \ldots, \mathbf{D}_m\}$

Rewritten: $\varepsilon = 1/\mathrm{Opt}$

$$\begin{aligned} \max \quad & \varepsilon \\ & \sum_{i=1}^{n} z_i = 1 \\ \text{s.t.} \quad & \sum_{p_i \in D} z_i \geq \varepsilon \qquad \forall D \in \mathcal{F} \\ & z_i \geq 0 \qquad \forall i \end{aligned}$$

By ε-net theorem...

Exists a set $\mathbf{X} \subseteq \mathbf{P}$ of size

$$O\left(\frac{d}{\varepsilon} \log \frac{d}{\varepsilon}\right) = O(\mathrm{Opt} \log(\mathrm{Opt}))$$

hitting all ranges.
Algorithm from **[Long, 2001]**

LP relax' of hitting set.

$$\min \quad \mathrm{Opt} = \sum_{i=1}^{n} x_i$$

$$\text{s.t.} \sum_{p_i \in D_j} x_i \geq 1 \qquad \forall j$$

$$x_i \geq 0 \qquad \forall i$$

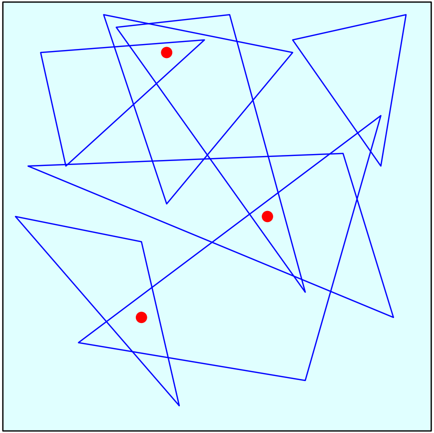
Lemma [Long, 2001]

In geometric settings there is an ε-net of size $O(K/\varepsilon)$ iff the integrality gap is K.

16: On small ε-nets

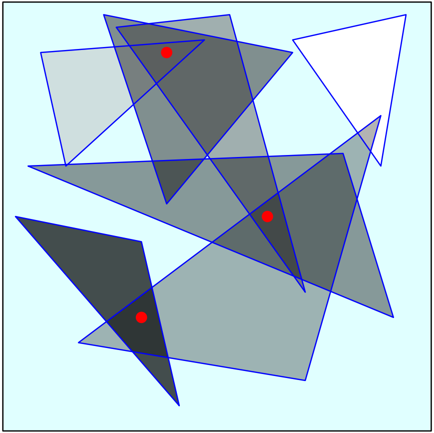
- $\mathbf{O}\left(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}\right)$: ε-net theorem.
- $\mathbf{O}(1/\varepsilon)$: halfplanes, halfspaces, pseudo-disks.
- $\mathbf{\Omega}\left(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}\right)$: Lower bound **[Komlós et al., 1992]**.
- $\mathbf{\Omega}\left(\frac{1}{\varepsilon}\mathbf{w}\left(\frac{1}{\varepsilon}\right)\right)$: points and lines **[Alon, 2010]**.
- $\mathbf{O}\left(\frac{1}{\varepsilon}\log\frac{\mathbf{U}(1/\varepsilon)}{1/\varepsilon}\right)$
  [Aronov, Ezra and Sharir, 2009]
  $\mathbf{U}(\mathbf{n})$: Union complexity of $\mathbf{n}$ shapes.

17: Covering points by triangles



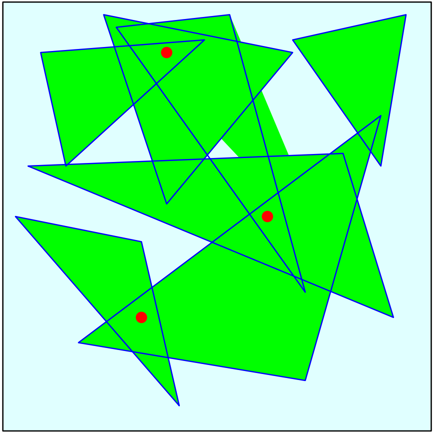
- Solve LP
  $\mathbf{k}$ = fract. # of triangles
- ε-net theorem:
  $\mathbf{O(k \log k)}$ sample $\Longrightarrow$
  covers all points.
- Do better...
  Sample only $\mathbf{k}$.

17: Covering points by triangles



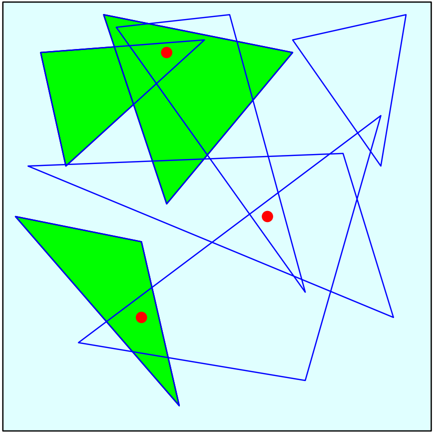
- Solve LP
  $\mathbf{k}$ = fract. # of triangles
- ε-net theorem:
  $\mathbf{O(k \log k)}$ sample $\implies$ covers all points.
- Do better...
  Sample only $\mathbf{k}$.

17: Covering points by triangles



- Solve LP
  $\mathbf{k}$ = fract. # of triangles
- ε-net theorem:
  $\mathbf{O(k \log k)}$ sample $\Longrightarrow$ covers all points.
- Do better...
  Sample only $\mathbf{k}$.

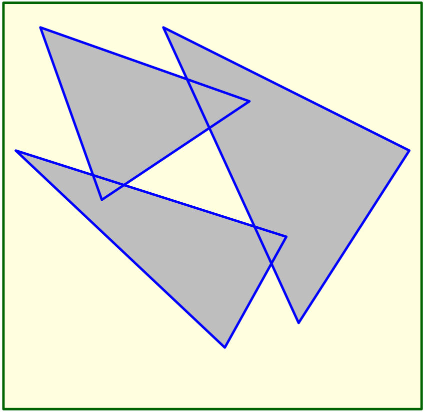
17: Covering points by triangles



- Solve LP
  k = fract. # of triangles
- ε-net theorem:
  O(k log k) sample $\Longrightarrow$ covers all points.
- Do better...
  Sample only **k**.

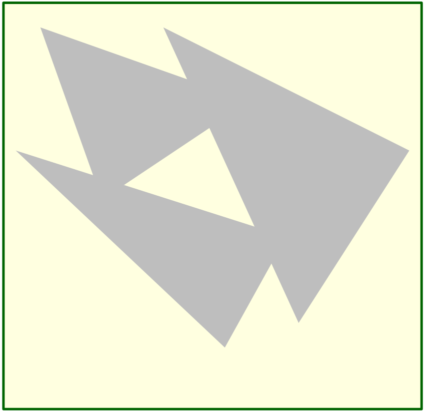
18: Union complexity

Length of description of the union.



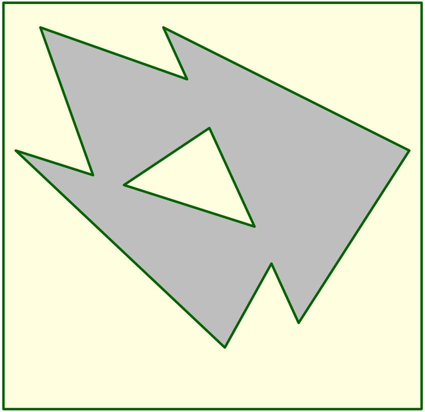
18: Union complexity

Length of description of the union.



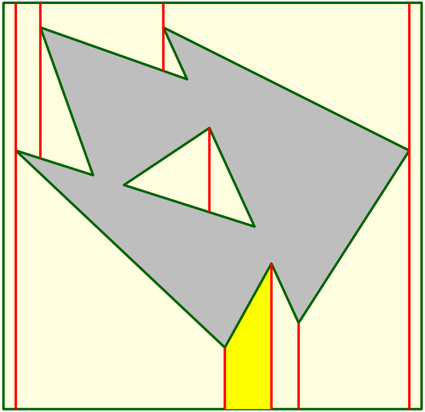
18: Union complexity

Length of description of the union.

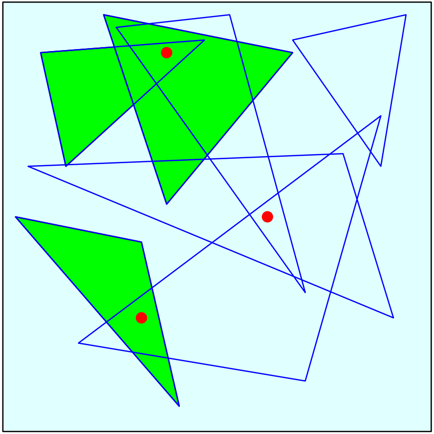


18: Union complexity

Length of description of the union.

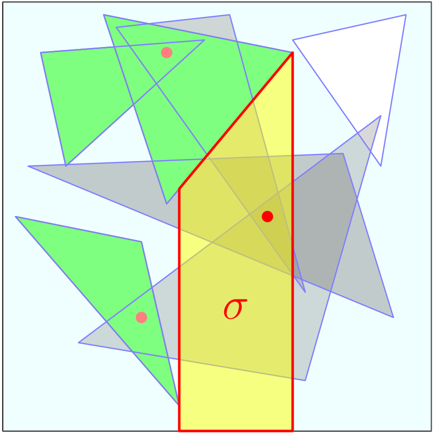


19: Back to Covering points by triangles



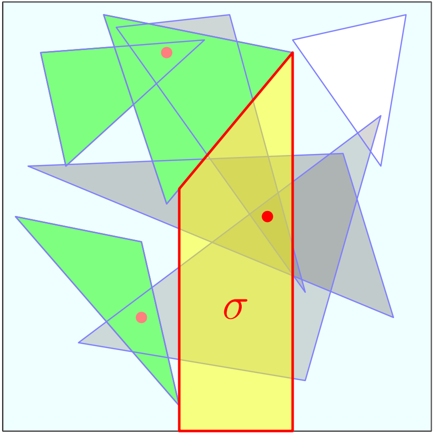
- Vertical decomposition.
- $\mathbf{k}$ = fract. # of triangles
- α = weight $\triangle$s inter. σ
- $\mathbf{E}[\alpha] = \mathbf{O}(1)$.
- $\mathbf{Pr}[\alpha > \mathbf{t}] = \mathbf{O}(2^{-\mathbf{t}})$.
- Fix: $\mathbf{O}(\alpha \log \alpha) = \mathbf{O}(1)$.
- Cover of size $\mathbf{O}(\mathbf{U}(\mathbf{k}))$. $\mathbf{U}(\mathbf{k})$ = union complexity of $\mathbf{k}$ triangles.
- $\implies$ ε-net for $(\mathbf{triangles}, \mathbf{points})$ of size $\mathbf{O}(\mathbf{U}(1/\varepsilon))$.

19: Back to Covering points by triangles



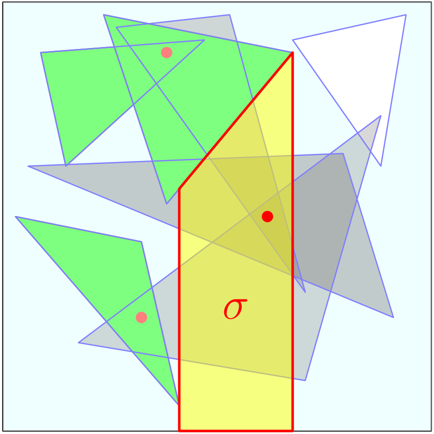
- Vertical decomposition.
- $\mathbf{k}$ = fract. # of triangles
- α = weight $\triangle$s inter. σ
- $\mathbf{E}[\alpha] = \mathbf{O}(1)$.
- $\mathbf{Pr}[\alpha > t] = \mathbf{O}(2^{-t})$.
- Fix: $\mathbf{O}(\alpha \log \alpha) = \mathbf{O}(1)$.
- Cover of size $\mathbf{O}(\mathbf{U}(\mathbf{k}))$. $\mathbf{U}(\mathbf{k})$ = union complexity of $\mathbf{k}$ triangles.
- $\implies$ ε-net for **(triangles, points)** of size $\mathbf{O}(\mathbf{U}(1/\varepsilon))$.

19: Back to Covering points by triangles



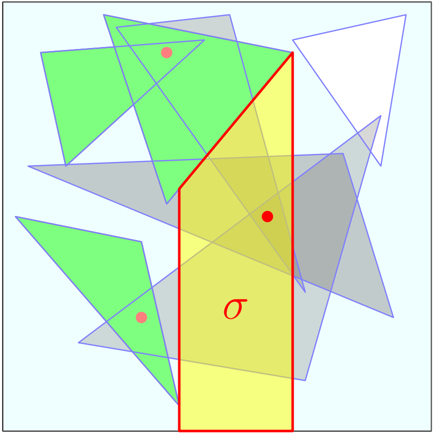
- Vertical decomposition.
- $\mathbf{k}$ = fract. # of triangles
- α = weight $\triangle$s inter. $\boldsymbol{\sigma}$
- $\mathbf{E}[\alpha] = \mathbf{O}(1)$.
- $\mathbf{Pr}[\alpha > t] = \mathbf{O}(2^{-t})$.
- Fix: $\mathbf{O}(\alpha \log \alpha) = \mathbf{O}(1)$.
- Cover of size $\mathbf{O}(\mathbf{U}(\mathbf{k}))$. $\mathbf{U}(\mathbf{k})$ = union complexity of $\mathbf{k}$ triangles.
- $\Longrightarrow$ ε-net for $\mathbf{(triangles, points)}$ of size $\mathbf{O}(\mathbf{U}(1/\varepsilon))$.

19: Back to Covering points by triangles



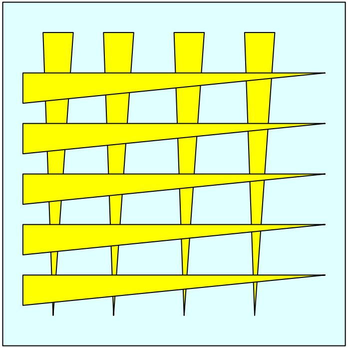
- Vertical decomposition.
- $\mathbf{k}$ = fract. # of triangles
- α = weight $\triangle$s inter. $\boldsymbol{\sigma}$
- $\mathbf{E}[\alpha] = \mathbf{O}(1)$.
- $\mathbf{Pr}[\alpha > \mathbf{t}] = \mathbf{O}(2^{-\mathbf{t}})$.
- Fix: $\mathbf{O}(\alpha \log \alpha) = \mathbf{O}(1)$.
- Cover of size $\mathbf{O}(\mathbf{U}(\mathbf{k}))$.
  $\mathbf{U}(\mathbf{k})$ = union complexity of $\mathbf{k}$ triangles.
- $\implies$ ε-net for $(\mathbf{triangles}, \mathbf{points})$ of size $\mathbf{O}(\mathbf{U}(1/\varepsilon))$.

19: Back to Covering points by triangles



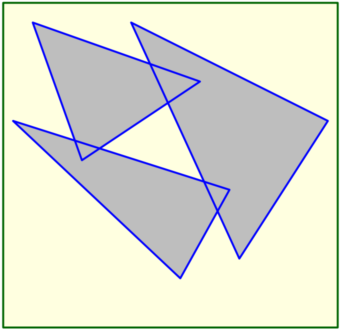
- Vertical decomposition.
- $\mathbf{k}$ = fract. # of triangles
- $\boldsymbol{\alpha}$ = weight $\triangle$s inter. $\boldsymbol{\sigma}$
- $\mathbf{E}[\alpha] = \mathbf{O}(1)$.
- $\mathbf{Pr}[\alpha > \mathbf{t}] = \mathbf{O}(2^{-\mathbf{t}})$.
- Fix: $\mathbf{O}(\alpha \log \alpha) = \mathbf{O}(1)$.
- Cover of size $\mathbf{O}(\mathbf{U}(\mathbf{k}))$. $\mathbf{U}(\mathbf{k})$ = union complexity of $\mathbf{k}$ triangles.
- $\implies$ ε-net for $(\mathbf{triangles}, \mathbf{points})$ of size $\mathbf{O}(\mathbf{U}(1/\varepsilon))$.

20: Union complexity of triangles



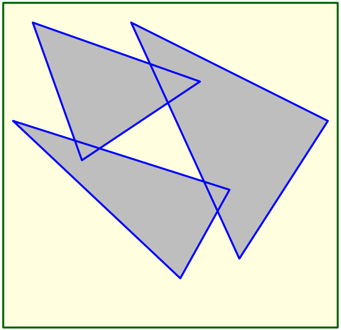
- Worst case complexity $\Omega(n^2)$.
- Fat triangles... $O(n \log \log n)$. [Matoušek, Pach, Sharir, Sifrony, and Welzl, 1994]
- For (fat triangles, points) $\implies$ ε-net of size $O\left(\frac{1}{\varepsilon} \log \log \frac{1}{\varepsilon}\right)$.
- For (fat triangles, points) $\implies$ ε-net of size $O\left(\frac{1}{\varepsilon} \log \log \log \frac{1}{\varepsilon}\right)$. [Aronov, Ezra and Sharir, 2009]

20: Union complexity of triangles



- Worst case complexity $\Omega(n^2)$.
- Fat triangles... $O(n \log\log n)$. **[Matoušek, Pach, Sharir, Sifrony, and Welzl, 1994]**
- For (fat triangles, points) $\implies$ ε-net of size $O\left(\frac{1}{\varepsilon}\log\log\frac{1}{\varepsilon}\right)$.
- For (fat triangles, points) $\implies$ ε-net of size $O\left(\frac{1}{\varepsilon}\log\log\log\frac{1}{\varepsilon}\right)$. **[Aronov, Ezra and Sharir, 2009]**

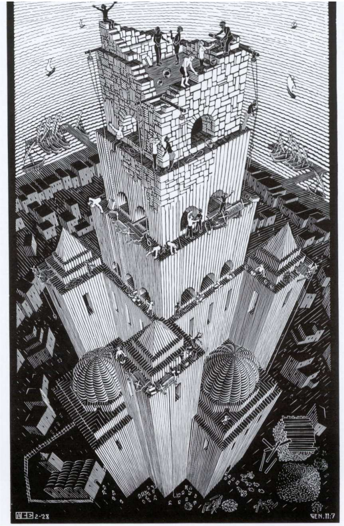
20: Union complexity of triangles



- Worst case complexity $\mathbf{\Omega(n^2)}$.
- Fat triangles...
  $\mathbf{O(n \log \log n)}$.
  [Matoušek, Pach, Sharir, Sifrony, and Welzl, 1994]
- For (fat triangles, points) $\implies$ ε-net of size $\mathbf{O}\left(\frac{1}{\varepsilon} \log \log \frac{1}{\varepsilon}\right)$.
- For (fat triangles, points) $\implies$ ε-net of size $\mathbf{O}\left(\frac{1}{\varepsilon} \log \log \log \frac{1}{\varepsilon}\right)$.
  [Aronov, Ezra and Sharir, 2009]

Name	Property $\forall \mathbf{D} \in \mathcal{F}$	Sample size
ε-net	$\overline{\mathbf{m}} = \overline{\mathbf{m}}(\mathbf{r})$, $\bar{\mathbf{s}} = \bar{\mathbf{s}}(\mathbf{r})$ $\overline{\mathbf{m}} \geq \varepsilon \quad \Rightarrow \quad \bar{\mathbf{s}} > 0$	$\mathbf{O}\left(\frac{\delta}{\varepsilon} \log \frac{1}{\varepsilon}\right)$
ε-sample	$\lvert \overline{\mathbf{m}} - \bar{\mathbf{s}} \rvert \leq \varepsilon$	$\mathbf{O}\left(\frac{\delta}{\varepsilon^2}\right)$
Sensitive ε-approx.	$\lvert \overline{\mathbf{m}} - \bar{\mathbf{s}} \rvert \leq \frac{\varepsilon}{2}\left(\sqrt{\overline{\mathbf{m}}} + \varepsilon\right)$	$\mathbf{O}\left(\frac{\delta}{\varepsilon^2} \log \frac{1}{\varepsilon}\right)$
Relative $(\varepsilon, \mathbf{p})$-approx.	$\overline{\mathbf{m}} \geq \mathbf{p} \quad \Rightarrow$ $(1 - \varepsilon)\overline{\mathbf{m}} \leq \bar{\mathbf{s}} \leq (1 + \varepsilon)\overline{\mathbf{m}}$	$\mathbf{O}\left(\frac{\delta}{\varepsilon^2 \mathbf{p}} \log \frac{1}{\mathbf{p}}\right)$

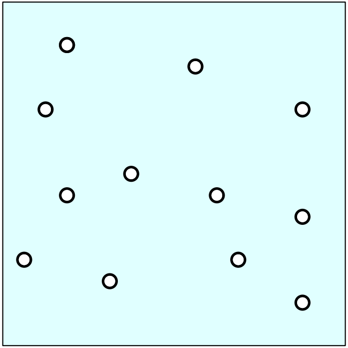
22: Also...



Q: What is best approximation to **P** using half the points?

Discrepancy

Color the points by red/blue and take the red subset as the approximation.



E: # of edges of matching crossing **h**
Result: Discrepancy $\mathbf{O}\left(\sqrt{\mathbf{E} \log \mathbf{n}}\right)$

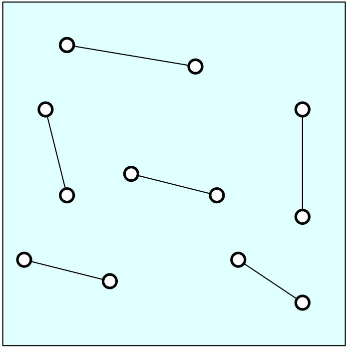
Lemma

Red points form $\widetilde{\mathbf{O}}(\mathbf{1}/\sqrt{\mathbf{n}})$-sample.

Q: What is best approximation to **P** using half the points?

Discrepancy

Color the points by red/blue and take the red subset as the approximation.



E: # of edges of matching crossing **h**
Result: Discrepancy $\mathbf{O}\left(\sqrt{\mathbf{E} \log \mathbf{n}}\right)$

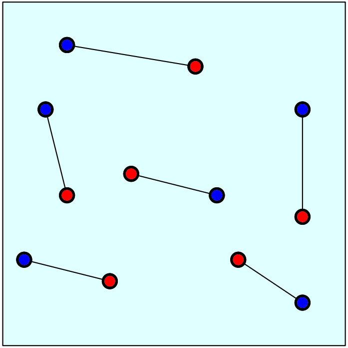
Lemma

Red points form $\widetilde{\mathbf{O}}(\mathbf{1}/\sqrt{\mathbf{n}})$-sample.

Q: What is best approximation to **P** using half the points?

Discrepancy

Color the points by red/blue and take the red subset as the approximation.



E: # of edges of matching crossing **h**
Result: Discrepancy $O\left(\sqrt{E \log n}\right)$

Lemma

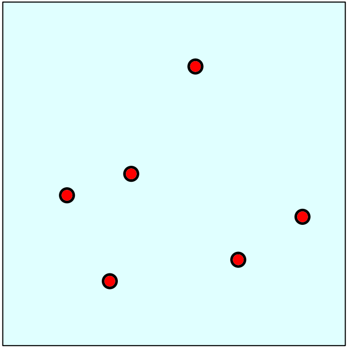
Red points form $\tilde{O}(1/\sqrt{n})$-sample.

Best sample using half the points?

Q: What is best approximation to **P** using half the points?

Discrepancy

Color the points by red/blue and take the red subset as the approximation.



E: # of edges of matching crossing **h**

Result: Discrepancy $O\left(\sqrt{E \log n}\right)$

Lemma

Red points form $\widetilde{O}(1/\sqrt{n})$-sample.

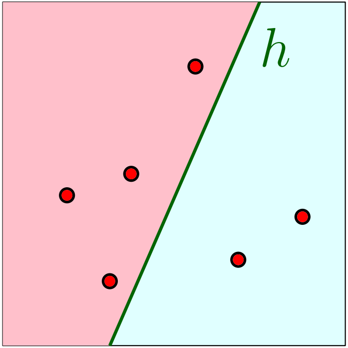
23: Getting a good sketch via discrepancy

Best sample using half the points?

Q: What is best approximation to **P** using half the points?

Discrepancy

Color the points by red/blue and take the red subset as the approximation.



E: # of edges of matching crossing **h**

Result: Discrepancy $\mathbf{O}\left(\sqrt{\mathbf{E} \log \mathbf{n}}\right)$

Lemma

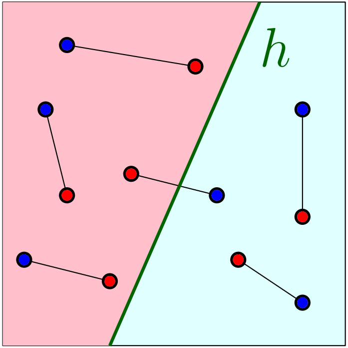
Red points form $\widetilde{\mathbf{O}}(\mathbf{1}/\sqrt{\mathbf{n}})$-sample.

Best sample using half the points?

Q: What is best approximation to **P** using half the points?

Discrepancy

Color the points by red/blue and take the red subset as the approximation.



E: # of edges of matching crossing **h**

Result: Discrepancy $\mathbf{O}\left(\sqrt{\mathbf{E} \log \mathbf{n}}\right)$

Lemma

Red points form $\widetilde{\mathbf{O}}(\mathbf{1}/\sqrt{\mathbf{n}})$-sample.

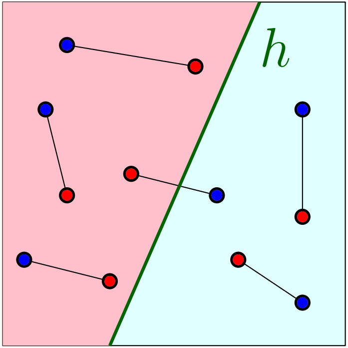
23: Getting a good sketch via discrepancy

Best sample using half the points?

Q: What is best approximation to **P** using half the points?

Discrepancy

Color the points by red/blue and take the red subset as the approximation.



E: # of edges of matching crossing **h**
Result: Discrepancy $\mathbf{O}\left(\sqrt{\mathbf{E} \log \mathbf{n}}\right)$

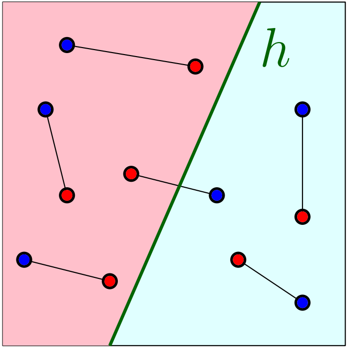
Lemma

Red points form $\widetilde{\mathbf{O}}(1/\sqrt{\mathbf{n}})$-sample.

Q: What is best approximation to **P** using half the points?

Discrepancy

Color the points by red/blue and take the red subset as the approximation.



E: # of edges of matching crossing **h**
Result: Discrepancy $\mathbf{O}\left(\sqrt{\mathbf{E}\log \mathbf{n}}\right)$

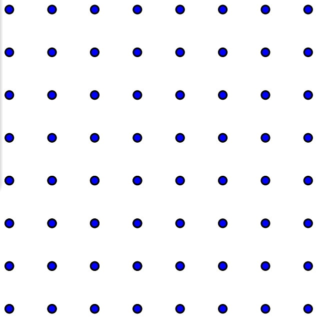
Lemma

Red points form $\widetilde{\mathbf{O}}(1/\sqrt{\mathbf{n}})$-sample.

24: Spanning tree with low crossing number

[Welzl, 1992]

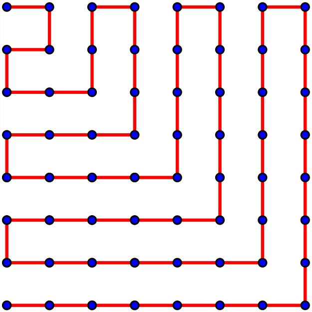
P: **n** points
T: spanning tree for **P**
Such that any line ℓ crosses
$O(\sqrt{n})$ edges of **T**
Proof: Uses reweighting.



Any point set behaves like a grid.
$\Longrightarrow$ tour of **P** with
$\mathrm{cr}(P) = O(\sqrt{n})$.
$\Longrightarrow$ matching of **P** with
$\mathrm{cr}(P) = O(\sqrt{n})$.

[Welzl, 1992]

P: **n** points
T: spanning tree for **P**
Such that any line ℓ crosses
$O(\sqrt{n})$ edges of **T**
Proof: Uses reweighting.



Any point set behaves like a grid.
$\Longrightarrow$ tour of **P** with
$\mathrm{cr}(P) = O(\sqrt{n})$.
$\Longrightarrow$ matching of **P** with
$\mathrm{cr}(P) = O(\sqrt{n})$.

$\mathbf{n} = |\mathbf{P}|$

Plugging matching into disc. construction:

Theorem

$\mathbf{P}$: *Set of* $\mathbf{n}$ *points in the plane.*
One can compute coloring with discrepancy $\mathbf{O}\left(\sqrt{\mathbf{E} \log \mathbf{n}}\right) = \widetilde{\mathbf{O}}\left(\mathbf{n}^{1/4}\right)$.

$\implies \widetilde{\mathbf{O}}(1/\mathbf{n}^{3/4})$-sample using $\mathbf{n}/2$ points.

Theorem

For $(\textit{points}, \textit{halfplanes})$ *there is* ε*-sample of size* $\widetilde{\mathbf{O}}(1/\varepsilon^{4/3})$.

Spanning tree with relative crossing number

[Har-Peled and Sharir, 2010]:

- $\mathbf{P}$: set of points
  $\mathbf{n} = |\mathbf{P}|$
  $\exists$ $\mathbf{T}$ spanning tree of $\mathbf{P}$
- $\mathbf{h}^+$: halfplane
  $\mathbf{k} = |\mathbf{h}^+ \cap \mathbf{P}|$.
  $\implies$ $\mathbf{h}$ crosses $\mathbf{O}\left(\sqrt{\mathbf{k}} \log(\mathbf{n}/\mathbf{k})\right)$ of $\mathbf{T}$.

$\implies$ smaller Relative $(\varepsilon, \mathbf{p})$-approx.

27: Conclusions

- Most of bounds not tight.
- Open problem:

Approximate ε-sample

Given that $\exists$ ε-sample of size $\mathbf{t}$, compute a sample of size $\mathbf{O(t \log(1/\varepsilon))}$.

Alon, N. (2010).
A non-linear lower bound for planar ε-nets.
Manuscript.

Har-Peled, S. and Sharir, M. (2010).
Relative $(\mathbf{p}, \varepsilon)$-approximations in geometry.
Discrete Comput. Geom.
To appear.

Haussler, D. and Welzl, E. (1987).
ε-nets and simplex range queries.
Discrete Comput. Geom., 2:127–151.

Komlós, J., Pach, J., and Woeginger, G. (1992).
Almost tight bounds for ε-nets.
Discrete Comput. Geom., 7:163–173.

Li, Y., Long, P. M., and Srinivasan, A. (2001).
Improved bounds on the sample complexity of learning.

J. Comput. Syst. Sci., 62(3):516–527.

Long, P. M. (2001).
Using the pseudo-dimension to analyze approximation algorithms for integer programming.
In *Proc. 7th Workshop Algorithms Data Struct.*, volume 2125 of *Lecture Notes Comput. Sci.*, pages 26–37.

Vapnik, V. N. and Chervonenkis, A. Y. (1971).
On the uniform convergence of relative frequencies of events to their probabilities.
Theory Probab. Appl., 16:264–280.

Welzl, E. (1992).
On spanning trees with low crossing numbers.
In *Data Structures and Efficient Algorithms, Final Report on the DFG Special Joint Initiative*, volume 594 of *Lect. Notes in Comp. Sci.*, pages 233–249. Springer-Verlag.