Linear Programming, Sparsest Cut, and Metric Embeddings

Lecture 8

Outline

Linear programming

Sparsest cut as an LP relaxation

ℓ_1 metrics are combinations of cuts

Bourgain's embedding theorem

Outline

Linear programming

Sparsest cut as an LP relaxation

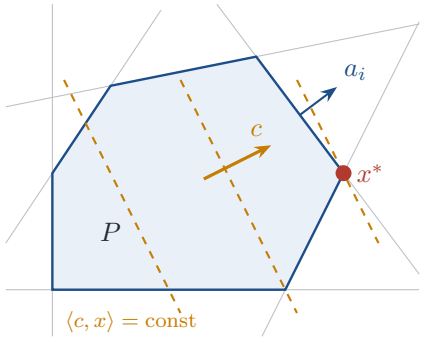
ℓ_1 metrics are combinations of cuts

Bourgain's embedding theorem

A linear program

$$\max \langle c, x \rangle \quad \text{s.t.} \quad Ax \le b$$
$$x \in \mathbb{R}^n, \quad A \in \mathbb{R}^{m \times n}, \quad b \in \mathbb{R}^m$$

- $Ax \le b$: $\langle a_i, x \rangle \le b_i$ for all i
- feasible set P: intersection of half-spaces
- P bounded, nonempty: optimum at a vertex

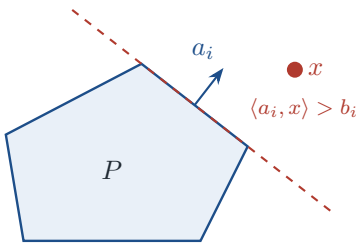


LPs can be solved in polynomial time

Theorem (Khachiyan 1979; Karmarkar 1984)

An optimal x^* (or "infeasible" / "unbounded") in time polynomial in the bit size of (A, b, c).

- the ellipsoid method only needs a separation oracle:
  given $x \notin P$, return a violated constraint
- $\Rightarrow$ exponentially many constraints are fine



Certifying that there is no solution

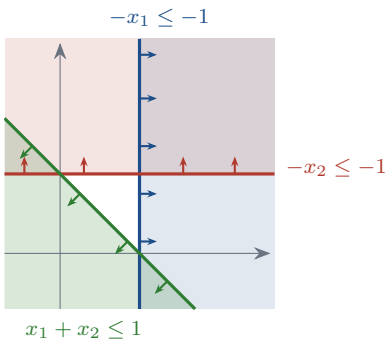
Is there x with $Ax \leq b$?

A certificate for "no":

$$\alpha \geq 0, \qquad \alpha^\top A = 0, \qquad \alpha^\top b < 0$$

Weak duality

a solution would give $0 = \alpha^\top A x \leq \alpha^\top b < 0$



$\alpha = (1, 1, 1)$: the sum is $0 \leq -1$

Farkas' lemma: certificates always exist

Theorem (Farkas)

Exactly one of the following holds:

(i) $Ax \leq b$ for some $x \in \mathbb{R}^n$;

(ii) $\alpha^\top A = 0$ and $\alpha^\top b < 0$ for some $\alpha \geq 0$.

LP duality (when the maximum is finite)

$$\max\{\langle c, x\rangle : Ax \leq b\} \;=\; \min\{\langle b, y\rangle : A^\top y = c,\; y \geq 0\}$$

Three uses of LP in algorithm design

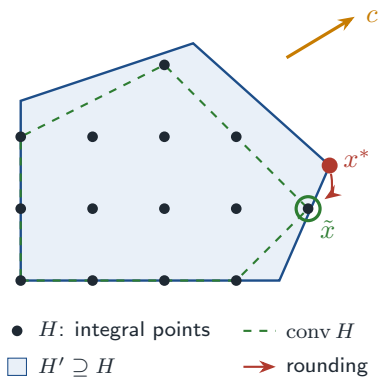
1. **Direct modelling**
   bipartite matching: the LP has integral vertices

2. **Implicit LPs**
   exponentially many constraints + separation oracle

3. **Relaxation & rounding**
   approximation algorithms ← today

LP relaxation and rounding

1. **Problem:** $\max\langle c, x\rangle$, $x \in H$ (hard)
2. **Relax:** polytope $H' \supseteq H$ (easy)
3. **Solve:** LP optimum x^* over H'
4. **Round:** $x^* \mapsto \tilde{x} \in H$,
   $\langle c, \tilde{x}\rangle \geq \gamma \langle c, x^*\rangle$



Rounding gives a γ-approximation

Maximization $(\gamma \leq 1)$

$$\gamma \cdot \mathrm{OPT} \underset{H \subseteq H'}{\leq} \gamma \langle c, x^* \rangle \underset{\text{rounding}}{\leq} \langle c, \tilde{x} \rangle \underset{\tilde{x} \in H}{\leq} \mathrm{OPT}$$

Minimization $(\gamma \geq 1)$

$$\mathrm{OPT} \underset{\tilde{x} \in H}{\leq} \langle c, \tilde{x} \rangle \underset{\text{rounding}}{\leq} \gamma \langle c, x^* \rangle \underset{H \subseteq H'}{\leq} \gamma \cdot \mathrm{OPT}$$

Sparsest cut will be a minimization problem.

Outline

Linear programming

Sparsest cut as an LP relaxation

ℓ_1 metrics are combinations of cuts

Bourgain's embedding theorem

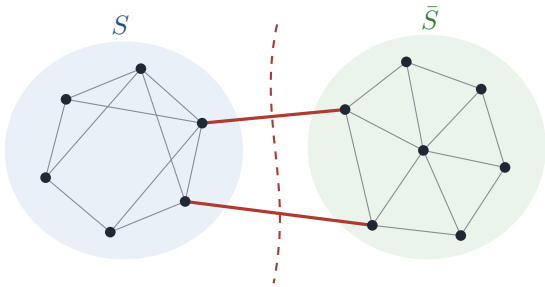
The sparsest cut problem

Graph $G = (V, E)$, $n = |V|$.

$$\phi(S) = \frac{E(S, \bar{S})}{|S| \cdot |\bar{S}|}$$

$$\phi^* = \min_{\emptyset \neq S \subsetneq V} \phi(S)$$

few crossing edges, *and* both sides large

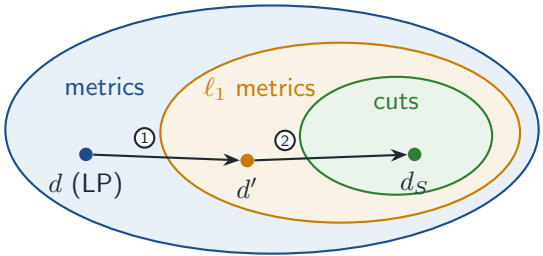


$$\phi(S) = \frac{2}{6 \cdot 7} = \frac{1}{21}$$

Goal: an $O(\log n)$-approximation

Theorem (Leighton–Rao 1988; Linial–London–Rabinovich 1995)

A polynomial-time algorithm finds S with $\phi(S) \le O(\log n) \cdot \phi^*$.



① embed into ℓ_1: lose $O(\log n)$

② split into cuts: no loss

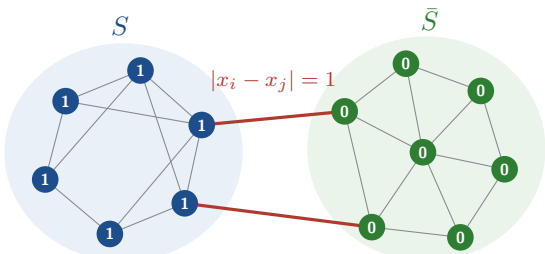
Cuts as $0/1$ vectors

Encode S by $x = \mathbf{1}_S \in \{0,1\}^n$:

$$E(S, \bar{S}) = \sum_{ij \in E} |x_i - x_j|$$

$$|S|\,|\bar{S}| = \sum_{i<j} |x_i - x_j|$$

(sum over unordered pairs)



Linearize with $d_{ij} = |x_i - x_j|$

For a nontrivial cut: $\phi(S) \le \gamma \iff$

$$\sum_{ij \in E} d_{ij} \; - \; \gamma \sum_{i<j} d_{ij} \; \le \; 0, \qquad d \neq 0$$

✓ linear in d

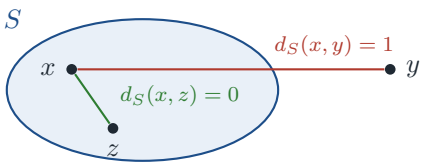
× but d ranges over $\left\{(|x_i - x_j|)_{i<j} : x \in \{0,1\}^n\right\}$

Next: understand this set geometrically, then relax it.

Metrics, ℓ_1 metrics, cut metrics

- **metric:** $d(x,x) = 0,\ \ d(x,y) = d(y,x),$
  $d(x,z) \le d(x,y) + d(y,z)$
- ℓ_1 **metric:** $d(x,y) = \|f(x) - f(y)\|_1$
- **cut metric:** $d_S(x,y) = |\mathbf{1}_S(x) - \mathbf{1}_S(y)|$

cuts $\subseteq$ ℓ_1 metrics $\subseteq$ metrics



Sparsest cut optimizes over the cut cone

For a set $\mathcal{D}$ of metrics on V:

$$\lambda(\mathcal{D}) := \min\Big\{ \sum_{uv\in E} d(u,v) \;:\; d\in\mathcal{D},\ \sum_{u<v} d(u,v) = 1\Big\}$$

Fact

$$\lambda(\mathrm{CUT}_n) = \phi^*, \qquad \text{where} \quad \mathrm{CUT}_n := \Big\{ \textstyle\sum_S \alpha_S\, d_S \;:\; \alpha \geq 0\Big\}$$

Why (S nontrivial): if $\sum_{u<v} d = \sum_S \alpha_S |S||\bar{S}| = 1$, then $\sum_{uv\in E} d = \sum_S \alpha_S |S||\bar{S}| \cdot \phi(S) \;\geq\; \phi^*$.

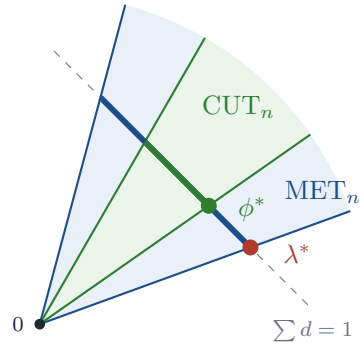
Relaxation: from cuts to all metrics

$$\lambda^* := \lambda(\mathrm{MET}_n) \;\le\; \lambda(\mathrm{CUT}_n) = \phi^*$$

λ^* is an LP:

$$\begin{aligned} \min \quad & \textstyle\sum_{uv \in E} d_{uv} \\ \text{s.t.} \quad & \textstyle\sum_{u<v} d_{uv} = 1 \\ & d_{uv} \ge 0 \\ & d_{uw} \le d_{uv} + d_{vw} \end{aligned}$$

$\binom{n}{2}$ variables, $O(n^3)$ constraints



schematic: minimizing along the slice $\sum d = 1$

Two ingredients

Lemma (A: ℓ_1 metrics are made of cuts)

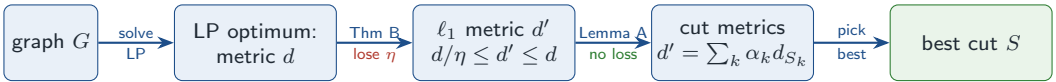
Every ℓ_1 metric is a non-negative combination of cut metrics: $d = \sum_k \alpha_k \, d_{S_k}$.

Theorem (B: Bourgain 1985)

Every metric on n points embeds into ℓ_1 with distortion $\eta = O(\log n)$:

$$\frac{d(x,y)}{\eta} \;\le\; \|f(x) - f(y)\|_1 \;\le\; d(x,y).$$

The rounding algorithm



1. solve the LP $\rightarrow$ metric d
2. embed: $d'(u,v) := \|f(u) - f(v)\|_1$
3. try all threshold cuts $\{u : f(u)_j < t\}$; output the sparsest

Corollary

$$\phi(S) \;\le\; \eta \cdot \lambda^* \;\le\; \eta \cdot \phi^* \;=\; O(\log n) \cdot \phi^*$$

Analysis, step 1: the embedding loses a factor η

Ratio of a nonzero metric: $R(d) := \dfrac{\sum_{uv \in E} d(u,v)}{\sum_{u<v} d(u,v)}$ (so $R(d_S) = \phi(S)$)

$$\sum_{uv \in E} d' \underset{d' \le d}{\le} \sum_{uv \in E} d \underset{\text{LP}}{=} \lambda^* \sum_{u<v} d \underset{d \le \eta d'}{\le} \eta\, \lambda^* \sum_{u<v} d'$$

$$R(d') \;\le\; \eta \cdot \lambda^*$$

Analysis, step 2: some cut is at least as good

By Lemma A: $d' = \sum_k \alpha_k \, d_{S_k}, \quad \alpha_k > 0, \;\; S_k$ nontrivial.

$$\min_k \; \phi(S_k) \;\le\; \frac{\sum_k \alpha_k \, E(S_k, \bar{S}_k)}{\sum_k \alpha_k \, |S_k||\bar{S}_k|} \;=\; R(d') \;\le\; \eta \, \lambda^* \;\le\; \eta \, \phi^*$$

a weighted average of ratios is at least their minimum

The S_k are threshold cuts of the coordinates of f $\Rightarrow$ the algorithm finds one. $\square$

Outline

Linear programming

Sparsest cut as an LP relaxation

ℓ_1 metrics are combinations of cuts

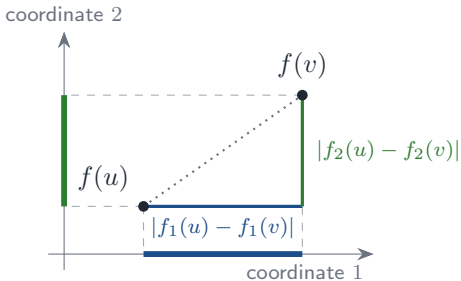
Bourgain's embedding theorem

Lemma A, part 1: ℓ_1 metrics are sums of line metrics

Line metric: $d(u, v) = |g(u) - g(v)|$

$$\|f(u) - f(v)\|_1 = \sum_{k=1}^{M} |f_k(u) - f_k(v)|$$

one line metric per coordinate



Lemma A, part 2: line metrics via random thresholds

Rescale g to $[0,1]$. Pick $\tau \sim U[0,1]$ and $S_\tau := \{u : g(u) < \tau\}$.



Claim

$$d(u,v) \;=\; \mathbb{E}\big[d_{S_\tau}(u,v)\big] \;=\; \textstyle\sum_{A \subseteq X} \Pr[S_\tau = A] \cdot d_A(u,v)$$

Why random thresholds work



For $g(u) \le g(v)$: $\quad S_\tau$ separates u and $v \iff g(u) < \tau \le g(v)$

$$\mathbb{E}\big[d_{S_\tau}(u,v)\big] = \Pr\big[g(u) < \tau \le g(v)\big] = g(v) - g(u) = d(u,v)$$

Only $\le n-1$ different cuts occur: an explicit finite combination.

Outline

Linear programming

Sparsest cut as an LP relaxation

ℓ_1 metrics are combinations of cuts

Bourgain's embedding theorem

Bourgain's theorem

Theorem (Bourgain 1985)

Every n-point metric embeds into ℓ_1 with distortion $O(\log n)$.

Proof plan

1. **Averaging:** random 1-D maps suffice, if they never expand and do not shrink on average.

2. **Bourgain's map:** distance to a random set, at a random scale.

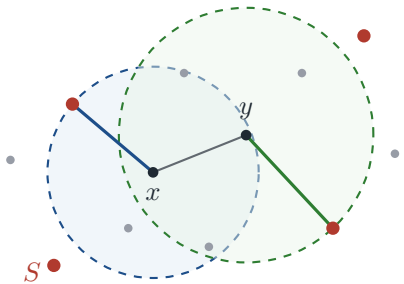
Distance to a set never expands distances

For $\emptyset \neq S \subseteq X$:

$$\pi_S(x) := d(x, S) = \min_{s \in S} d(x, s)$$

$$|d(x, S) - d(y, S)| \;\leq\; d(x, y)$$

Proof: $d(x, S) \leq d(x, y) + d(y, S)$; swap x, y.



red: S; radii $d(x, S)$, $d(y, S)$

Lemma (Averaging)

Let $\pi : X \to \mathbb{R}$ be a random map such that for all x, y:

(a) $|\pi(x) - \pi(y)| \le d(x, y)$ always,

(b) $\mathbb{E}\,|\pi(x) - \pi(y)| \ge d(x, y)/\eta$.

Then there is $f : X \to \mathbb{R}^T$ with

$$\frac{d(x,y)}{2\eta} \;\le\; \|f(x) - f(y)\|_1 \;\le\; d(x,y).$$

Warm-up: if π has finite support

$\pi = \pi^{(k)}$ with probability p_k, $k = 1, \dots, K$:

$$f(x) := \big(p_1\,\pi^{(1)}(x),\ \dots,\ p_K\,\pi^{(K)}(x)\big)$$

$$\|f(x) - f(y)\|_1 = \sum_{k=1}^{K} p_k \left|\pi^{(k)}(x) - \pi^{(k)}(y)\right| = \mathbb{E}\,|\pi(x) - \pi(y)|$$

✓ distortion exactly η

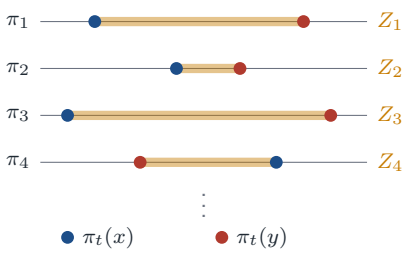
× K can be $\approx 2^n$ $\Rightarrow$ sample instead

Sampling T independent copies

$$f(x) := \tfrac{1}{T}\big(\pi_1(x), \ldots, \pi_T(x)\big)$$

- by (a): $\|f(x) - f(y)\|_1 \le d(x,y)$
- $\|f(x) - f(y)\|_1 = \bar{Z}$ is an average of i.i.d. $Z_t \in [0, D]$ with mean $\mu \ge D/\eta$

$D := d(x,y), \quad Z_t := |\pi_t(x) - \pi_t(y)|$



Concentration: Chebyshev and a union bound

$$\operatorname{Var} \bar{Z} = \frac{\operatorname{Var} Z_1}{T} \le \frac{\mathbb{E} Z_1^2}{T} \underset{Z_1 \le D}{\le} \frac{D\mu}{T}$$

$$\Pr\left[\bar{Z} < \frac{\mu}{2}\right] \underset{\text{Chebyshev}}{\le} \frac{4 \operatorname{Var} \bar{Z}}{\mu^2} \le \frac{4D}{T\mu} \underset{\mu \ge D/\eta}{\le} \frac{4\eta}{T}$$

- $T := \lceil 4\eta n^2 \rceil$: each pair fails with probability $\le 1/n^2$
- union bound over $< n^2/2$ pairs: with probability $> \frac{1}{2}$, for all pairs,

$$\|f(x) - f(y)\|_1 \ge \frac{\mu}{2} \ge \frac{d(x,y)}{2\eta}. \qquad \square$$

Chernoff instead of Chebyshev: $T = O(\eta \log n) = O(\log^2 n)$ suffices.

Step 2: Bourgain's random map

Lemma (Random scale)

Let $m := \lfloor \log_2 n \rfloor + 2$. Pick $k \in \{1, \dots, m\}$ uniformly, then $S \subseteq X$ with each point independently with probability 2^{-k}, and set $\pi(x) := d(x, S)$. Then

(a) $|\pi(x) - \pi(y)| \le d(x, y)$;

(b) $\mathbb{E}\,|\pi(x) - \pi(y)| \;\ge\; \dfrac{c_0}{2m}\, d(x, y), \qquad c_0 := \frac{1}{4}\big(1 - e^{-1/4}\big) > \frac{1}{20}.$

Averaging lemma $\Rightarrow$ distortion $4m/c_0 = O(\log n)$.

Random sets at different scales



$$\mathbb{E}\,|\pi(x) - \pi(y)| = \frac{1}{m}\sum_{i=1}^{m} \mathbb{E}\,\big|d(x, S_i) - d(y, S_i)\big| \qquad (S_i\text{: density } 2^{-i})$$

It suffices to prove, for all x, y:

$$(\star) \qquad \sum_{i=1}^{m} \mathbb{E}\,\big|d(x, S_i) - d(y, S_i)\big| \;\geq\; \frac{c_0}{2}\, d(x, y)$$

Growing balls around x and y

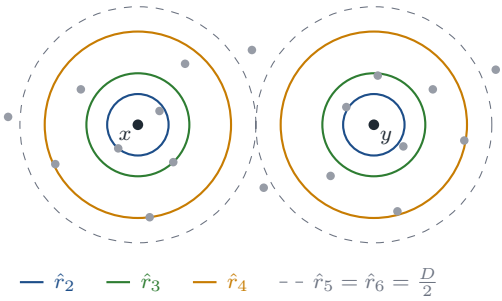
Fix x, y with $D := d(x,y) > 0$.

$$N(r) := |B(x,r)| + |B(y,r)|$$

$$r_i := \min\{r : \ N(r) \geq 2^i\}$$

$$\hat{r}_i := \min(r_i, \ D/2)$$

closed balls; $r_i := \infty$ if $2^i > 2n$



$n = 20$, $m = 6$: $\hat{r} = (0, \ 0, \ 0.78, \ 1.31, \ 2.37, \ 3, \ 3)$

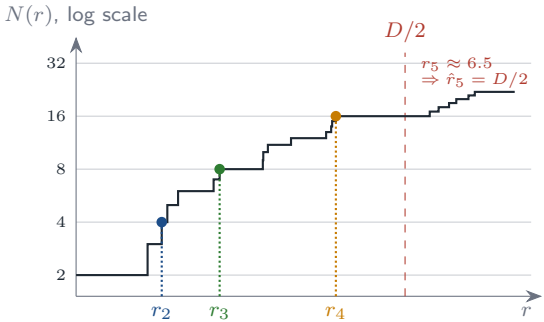
Three facts about the radii

(F0) $\hat{r}_0 = 0, \quad \hat{r}_m = D/2$

(F1) $|B^\circ(x, \hat{r}_i)| + |B^\circ(y, \hat{r}_i)| < 2^i$

(F2) $|B(x, r_{i-1})| + |B(y, r_{i-1})| \geq 2^{i-1}$

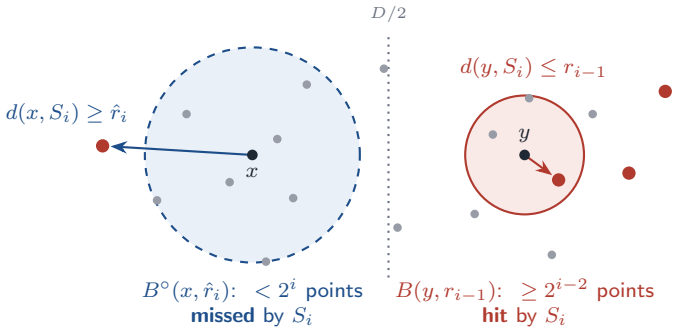
B°: open ball. (F2) needs $r_{i-1} < \infty$.



Key claim: every scale contributes

Claim

$$\mathbb{E}\left|d(x,S_i) - d(y,S_i)\right| \;\geq\; c_0\,(\hat{r}_i - \hat{r}_{i-1}) \qquad \text{for every } i \in \{1,\dots,m\}$$

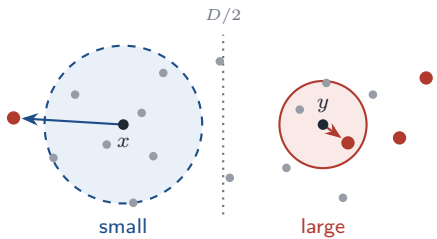


example point set, $i = 4$; red: a sample S_4 in the good event

The two balls

Assume $\hat{r}_i > \hat{r}_{i-1}$ (else there is nothing to show); then $r_{i-1} < D/2$.

- **small:** $|B^\circ(x, \hat{r}_i)| < 2^i$ (F1)
- **large:** $|B(y, r_{i-1})| \geq 2^{i-2}$ (F2)
- **disjoint:** $\hat{r}_i + r_{i-1} < D$ ($\triangle$-ineq.)



The good event

Good event

$$\mathcal{A} := \left\{S_i \cap B^\circ(x, \hat{r}_i) = \emptyset\right\} \;\cap\; \left\{S_i \cap B(y, r_{i-1}) \neq \emptyset\right\}$$

On $\mathcal{A}$: $d(x, S_i) \geq \hat{r}_i$ and $d(y, S_i) \leq r_{i-1} = \hat{r}_{i-1}$, so

$$|d(x, S_i) - d(y, S_i)| \;\geq\; \hat{r}_i - \hat{r}_{i-1}.$$

$$\mathbb{E}\,|d(x, S_i) - d(y, S_i)| \;\geq\; \Pr[\mathcal{A}] \cdot (\hat{r}_i - \hat{r}_{i-1})$$

The good event has probability at least c_0

Disjoint balls $\Rightarrow$ independent events. $\quad p := 2^{-i} \le \frac{1}{2}$:

$$\Pr[\mathcal{A}] \;=\; \underbrace{(1-p)^{|B^\circ(x,\hat{r}_i)|}}_{\ge (1-p)^{2^i} \ge 1/4} \cdot \underbrace{\Big(1-(1-p)^{|B(y,r_{i-1})|}\Big)}_{\ge 1-e^{-p\,2^{i-2}} = 1-e^{-1/4}} \;\ge\; c_0$$

using $(1-p)^{1/p} \ge \frac{1}{4}$ for $p \le \frac{1}{2}$ and $1-p \le e^{-p}$

Key claim proved

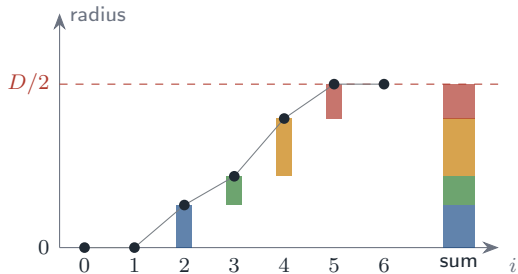
$$\mathbb{E}\,|d(x,S_i) - d(y,S_i)| \;\ge\; c_0\,(\hat{r}_i - \hat{r}_{i-1})$$

Telescoping finishes the proof

$$\sum_{i=1}^{m} \mathbb{E}\left|d(x,S_i) - d(y,S_i)\right| \;\geq\; c_0 \sum_{i=1}^{m} (\hat{r}_i - \hat{r}_{i-1}) \;=\; c_0\,(\hat{r}_m - \hat{r}_0) \underset{\text{(F0)}}{=} c_0\,\frac{D}{2}$$

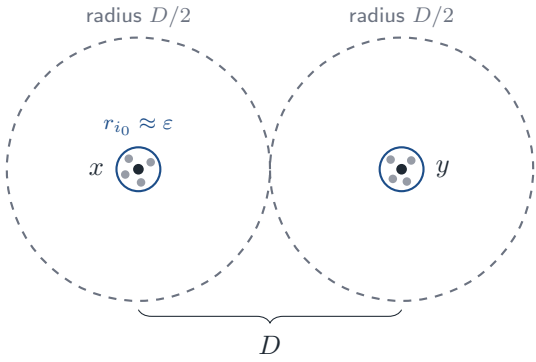
This is $(\star)$.

$\Rightarrow$ Bourgain's theorem $\square$



bars: increments $\hat{r}_i - \hat{r}_{i-1}$ (example point set)

Why truncate the radii at $D/2$?



× untruncated: the scales with $2r_i < D$ give only $c_0\, r_{i_0} \approx c_0\, \varepsilon$

✓ truncated: the next increment $\frac{D}{2} - r_{i_0}$ supplies the rest

Summary and outlook

- LP over all metrics: $\lambda^* \le \phi^*$
- Bourgain: metric $\to \ell_1$, lose $O(\log n)$
- $\ell_1 \to$ cuts, no loss $\Rightarrow$ $O(\log n)$-approximation

Outlook

- tight for this LP: integrality gap $\Omega(\log n)$ on expanders (LLR)
- SDP relaxation (ℓ_2^2 metrics): $O(\sqrt{\log n})$ (Arora–Rao–Vazirani 2004)