**Computer Science 2: Advanced Algorithms** Fall 2026

Lecture 8 — 17 September, 2026

*Prof. Jarosław Błasiok* *Scribe: Jakub Przewoski, Samuele N. Straccialini*

# 1 Overview

In the last few lectures we developed concentration tools for sub-Gaussian and sub-exponential random variables, and we used them to study dimensionality reduction through the Johnson-Lindenstrauss lemma and subspace embeddings.

In this lecture instead, we will shift to linear programming: we will first review the basic geometry of linear programs, then we will study it as a tool for designing approximation algorithms. In particular, we will carry out the program in full on one example: the *sparsest cut problem*. A cut is first encoded as a 0/1 vector and then by its pairwise separation distances. Exact cuts form the cut cone; relaxing this cone to all finite semimetrics gives a polynomial size linear program. The relaxed solution is rounded in two geometric steps:

$$\text{arbitrary metric} \xrightarrow[\text{distortion } O(\log n)]{\text{Bourgain}} \ell_1 \text{ metric} \xrightarrow[\text{no loss}]{\text{threshold decomposition}} \text{cut metric.}$$

The result is a polynomial time $O(\log n)$-approximation for sparsest cut.

# 2 Linear Programming

## 2.1 The standard form and its geometry

A *linear program* (LP) is the problem of optimizing a linear function over a region defined by finitely many linear inequalities:

$$\max_{x \in \mathbb{R}^n} \langle c, x \rangle \quad \text{subject to} \quad Ax \le b, \qquad A \in \mathbb{R}^{m \times n},\ b \in \mathbb{R}^m,\ c \in \mathbb{R}^n. \tag{1}$$

The inequality is coordinate-wise: writing $a_1, \dots, a_m \in \mathbb{R}^n$ for the rows of $A$, then

$$Ax \le b \quad \iff \quad \langle a_i, x \rangle \le b_i \quad \text{for every } i.$$

Each of the constraints binds $x$ to lie in a half-space, so the *feasible set*

$$P := \{x \in \mathbb{R}^n : Ax \le b\}$$

is the intersection of $m$ half-spaces, hence a convex polyhedron. The *objective* $\langle c, x \rangle$ is constant on hyperplanes orthogonal to $c$, and increases as we translate in direction $c$. Solving the LP means pushing the hyperplane $\langle c, x \rangle = \text{const}$ as far as possible in the direction $c$ while still touching $P$.

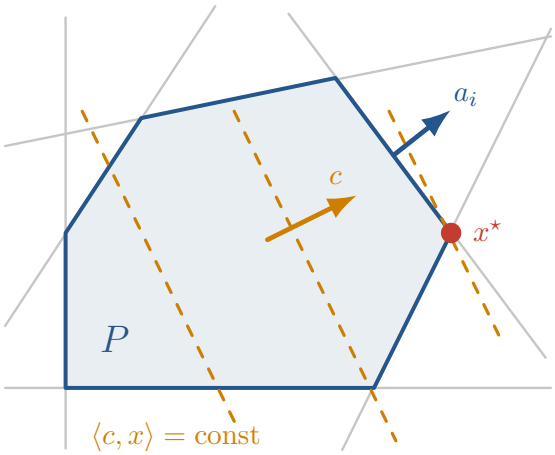



Figure 1: A bounded feasible polytope and parallel objective level sets.

**Where the optimum sits.** If $P$ is nonempty and bounded, a linear objective attains its maximum at some vertex of $P$. The optimum need not be unique: the set of maximizers is a face of $P$, which may contain several vertices and all points between them. Thus, an optimal vertex always exists under these assumptions, although not every optimal solution is a vertex (this geometric makes the simplex method possible: it is enough to search among vertices, of which there are finitely many). In particular, if all vertices of $P$ are integral, an integral optimal solution exists.

*Remark* 1. All the usual variants (minimization instead of maximization, equality constraints, sign constraints $x \geq 0$) are equivalent to the form (1) after simple transformations ($\min \langle c, x\rangle = -\max \langle -c, x\rangle$; an equality is two inequalities; $x \geq 0$ is $-x \leq 0$).

## 2.2 LPs can be solved in polynomial time

**Theorem 2** (Khachiyan 1979 [2]; Karmarkar 1984 [1])**.** *There exists an algorithm which outputs the optimal $x^*$ (or "infeasible"/"unbounded") of a linear program in time polynomial in the bit length of $(A, b, c)$.*

Two remarks about this theorem will be useful. First, the running time is polynomial in the size of the LP description. The sparsest-cut relaxation has $\binom{n}{2} = O(n^2)$ variables and $O(n^3)$ explicitly listed constraints. It therefore has polynomial size, so Theorem 2 applies directly. From now on, we may regard solving this LP as a polynomial-time step.

Second, an LP does not always need to list all of its constraints explicitly. The ellipsoid method [3] can instead use a *separation oracle*: given a candidate point $x$, the oracle either confirms that $x$ is feasible or returns a constraint violated by $x$. Consequently, an LP with exponentially many constraints can still be solved in polynomial time when it admits a polynomial-time separation oracle and the constraints have polynomially bounded encoding length.

## 2.3 Certifying that there is no solution: Farkas' Lemma

Consider the feasibility question: does there exist $x \in \mathbb{R}^n$ such that $Ax \leq b$? Suppose that one can find coefficients $\alpha \geq 0$ satisfying

$$\alpha^\top A = 0, \quad \alpha^\top b < 0.$$

Any feasible $x$ would then obey

$$0 = \alpha^\top Ax \leq \alpha^\top b < 0$$

which is a contradiction. Thus, finding $\alpha$ means certifying that the system is infeasible. This is the weak-duality direction of Farkas' lemma below (Theorem 4).

*Example* 3. Consider the system $x_1 \geq 1$, $x_2 \geq 1$, $x_1 + x_2 \leq 1$, written in standard form as

$$-x_1 \leq -1, \quad -x_2 \leq -1, \quad x_1 + x_2 \leq 1.$$

Take $\alpha = (1, 1, 1)$. Adding the three inequalities we have $(-x_1) + (-x_2) + (x_1 + x_2) \leq -1 - 1 + 1$, that is $0 \leq -1$. Thus, the system is infeasible.

Farkas' lemma tells us that such a certificate is not merely sufficient for infeasibility: whenever the system is infeasible, a certificate of this form exists.

**Theorem 4** (Farkas' lemma[4])**.** *Exactly one of the following alternatives holds:*

1. $Ax \leq b$ *for some* $x \in \mathbb{R}^n$*;*
2. $\alpha^\top A = 0$ *and* $\alpha^\top b < 0$ *for some* $\alpha \geq 0$*.*

Geometrically: either $b$ lies in the (convex, closed) set of right-hand sides for which the system is feasible, or it can be separated from it by a hyperplane, and the separating hyperplane has $\alpha$ as normal vector. Farkas' lemma is the source of LP duality.

**Theorem 5** (LP duality)**.** *If the primal maximum is finite and standard feasibility assumptions hold, then*

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

The $\leq$ inequality is a one line computation (weak duality):

$$\langle c, x\rangle = \left\langle A^\top y, x \right\rangle = \langle y, Ax\rangle \leq \langle y, b\rangle .$$

The nontrivial assertion of strong duality is that equality is attained by a pair of optimal solutions. Its proof is beyond the scope of this lecture.

## 2.4 Linear Programming in algorithm design

Linear Programming has three main uses in algorithm design:

1. **Direct modeling.** A combinatorial problem may be formulated as an LP which has an exact solution (i.e. whose vertices of the polytope are integral). An example is the *minimum cost perfect matching in bipartite graph*:

*Let $G = (U, E, V)$, where $U$ and $V$ are two sets of vertices and $c_{ij}$ = cost of $(i, j) \in E$. Let $x_{ij} = \mathbb{1}[\text{edge } (i, j) \text{ is chosen}]$. Then the min cost bipartite matching problem is:*

$$\begin{aligned} \min \quad & \sum_{(i,j)\in E} c_{ij} x_{ij} \\ \text{s.t.} \quad & \sum_{j:(i,j)\in E} x_{ij} = 1 \quad \forall i \in U, \\ & \sum_{i:(i,j)\in E} x_{ij} = 1 \quad \forall j \in V, \\ & x_{ij} \in \{0, 1\} \qquad \forall (i, j) \in E. \end{aligned}$$

The LP relaxation already has matching vertices, so solving the LP provides an integral solution to the problem.

2. **Implicit LPs.** The formulation may have exponentially many constraints, but there might exist an efficient separation oracle. Such a problem can be solved using the ellipsoid method.

3. **Relaxation and rounding.** The LP may not solve the problem, but its optimum may be a good enough bound: such solution can be transformed into an approximate optimal solution for the original problem. This is the case we will consider in the following sections.

### 2.5 Relaxation and Rounding

This paradigm can be summarized in four main passages:

- Problem: we want to solve $\max \langle c, x \rangle$ s.t. $x \in H$, where optimization over $H$ is difficult.
- Relax: we find a polytope $H' \supseteq H$, which is a larger set containing the original one, over which optimization is easy. Notice that over $H'$, optimum is at least as good as over $H$.
- Solve: find the optimum $x^*$ over $H'$. This is an easy LP.
- Round: map $x^* \mapsto \widetilde{x} \in H$, whose value produced is comparable to that of $x^*$.

The goal of the fourth step is to obtain a guarantee against the optimum of the original problem. In particular, for maximization we aim to $\langle c, \widetilde{x} \rangle \geq \gamma \langle c, x^* \rangle$ in order to get

$$\gamma \cdot \text{OPT} \underset{H \subseteq H'}{\leq} \gamma \langle c, x^* \rangle \underset{\text{rounding}}{\leq} \langle c, \tilde{x} \rangle \underset{\tilde{x} \in H}{\leq} \text{OPT} \qquad \text{for } 0 \leq \gamma \leq 1.$$

Analogously, in the minimization case, $\langle c, \widetilde{x} \rangle \leq \gamma \langle c, x^* \rangle$ and

$$\text{OPT} \underset{\tilde{x} \in H}{\leq} \langle c, \tilde{x} \rangle \underset{\text{rounding}}{\leq} \gamma \langle c, x^* \rangle \underset{H \subseteq H'}{\leq} \gamma \cdot \text{OPT} \qquad \text{for } \gamma \geq 1.$$

# 3 Sparsest cut problem as LP relaxation

We first define the graph given by $G = (V, E)$, with $n = |V|$. For a set $S$ and its complement $\bar{S}$, $E(S, \bar{S})$ defines the number of edges with one endpoint on each side. The sparsity of the cut $S$ is

$$\phi(S) := \frac{E(S, \bar{S})}{|S| \cdot |\bar{S}|}, \qquad \phi^* := \min_{\varnothing \neq S \subsetneq V} \phi(S)$$

We consider a setting in which we have few crossing edges and both sides of the graph are large. Our objective is to prove:

**Theorem 6** (Leighton–Rao [5]; Linial–London–Rabinovich [6])**.** *There is a polynomial-time algorithm that finds a nontrivial $S \subseteq V$ with $\phi(S) \leq O(\log n) \cdot \phi^*$.*

Which, guarantees that we can find a solution whose value is no larger than the $O(\log n)$ times the value of the optimal solution.

The approach proceeds by optimizing the problem via LP re-statement. Then we will relax the set of cuts to the set of all metrics over which we will optimize, and finally we will round up via the use of [7].

## 3.1 Cuts as 0/1 vectors

First, we re-state the problem. We encode each vertex in $S$ as $x = \mathbf{1}_S \in \{0, 1\}^n$. As this does not yet produce a linear problem since by the definition of

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

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

We define

$$d_{ij} = |x_i - x_j|,$$

which allows us to write (for a non-trivial set $S$)

$$\phi(S) \leq \gamma \Leftrightarrow \sum_{ij \in E} d_{ij} - \gamma \sum_{i<j} d_{ij} \leq 0 \quad d \neq 0 \tag{2}$$

This has the desired linear form, however this set is not convex, so it can't be easily described as an LP problem. This is because $d$ ranges over

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

Therefore, we will try to relax the set of cuts to the set of all metrics. To do that we need to introduce a few objects.

### 3.2 Metrics, $\ell_1$ metrics, cut metrics

**Definition 7** (Pseudometric)**.** *A function $d : X \times X \to R_{\geq 0}$ is a pseudometric if it satisfies the following properties:*

1. *Reflexivity: $d(x, x) = 0$*
2. *Symmetry: $d(x, y) = d(y, x)$*
3. *Triangle inequality: $d(x, y) + d(y, z) \geq d(x, z)$*

*Remark* 8. A pseudometric differs from a metric since it does not satisfy the identity of indiscernibles. Indeed, for a pseudometric, $d(x, y) = 0 \not\Longleftrightarrow x = y$. For convenience, we will refer to pseudometrics simply as metrics in the rest of the notes.

**Definition 9** ($\ell_1$ metric)**.** *A finite metric $d : X \times X \to \mathbb{R}_{\geq 0}$ is $\ell_1$ if there exists an integer $M$ and a map $f : X \to \mathbb{R}^M$ such that*

$$d(x, y) = \|f(x) - f(y)\|_1 = \sum_{i=1}^{M} |f_i(x) - f_i(y)| .$$

**Definition 10** (Cut metric)**.** *For $S \subseteq X$, the cut metric $d_s$ is defined as*

$$d_s(x, y) = |\mathbb{1}_S(x) - \mathbb{1}_S(y)| ,$$

*i.e. $d_S$ is 1 when the cut separates $x$ and $y$, 0 otherwise.*

Note that every cut metric is an $\ell_1$ metric (just take $M = 1$ and $f = \mathbb{1}_S$). Also, every $\ell_1$ metric is a metric because $\|\cdot\|_1$ satisfies the triangular inequality. Indeed, $d_S(x, y) + d_S(y, z) \geq d_S(x, z)$ can fail only if both terms on the left are 0, i.e. if both $x, y$ and $y, z$ are on the same side of the cut; but then $x, z$ are on the same side of the cut as well and the right-hand side is 0 as well.

Thus, summing up, we have that

$$\{\text{cut metrics}\} \subseteq \{\ell_1 \text{ metrics}\} \subseteq \{\text{metrics}\}.$$

### 3.3 Sparsest cut as an optimum

Since the objective (Equation 2) is scale invariant, we will work with cones rather than with individual metrics. We will therefore find useful the following definitions and fact.

**Definition 11** (Cut cone and metric cone)**.**

$$\text{CUT}_n := \left\{ \sum_S \alpha_S d_S : \alpha \geq 0 \right\},$$

$$\text{MET}_n := \{ d : d \textit{ is a pseudometric on } V \} .$$

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

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

Then, the following fact holds:

**Fact 12.**

$$\lambda(\mathrm{CUT}_n) = \phi^*.$$

*Proof.* A scaled cut metric $\dfrac{d_S}{|S|\,|\bar{S}|}$ is feasible in (2) and has objective $\phi(S)$, so $\lambda(\mathrm{CUT}_n) \le \phi^*$.

Conversely, let $d = \sum_S \alpha_S d_S \in \mathrm{CUT}_n$ be normalized by $\sum_{u<v} d(u,v) = 1$. Then,

$$\begin{aligned} 1 &= \sum_S \alpha_S\, |S|\, |\bar{S}| \\ \sum_{uv\in E} d(u,v) &= \sum_S \alpha_S \left|E(S,\bar{S})\right| \\ &= \sum_S \alpha_S\, |S|\, |\bar{S}| \phi(S) \ge \phi^* \end{aligned}$$

□

### 3.4 Relaxation

Since every cut is a metric, $\mathrm{CUT}_n \subset \mathrm{MET}_n$. So, the relaxed optimum $\lambda^*$ satisfies

$$\lambda^* := \lambda(\mathrm{MET}_n) \le \lambda(\mathrm{CUT}_n) = \phi^*, \tag{3}$$

where the $\lambda^*$ can be found by solving the following LP:

$$\begin{aligned} \min \quad & \sum_{uv\in E} d_{uv} \\ \text{s.t.} \quad & \sum_{u<v} d_{uv} = 1, \\ & d_{uv} \ge 0 && \text{for all } u < v, \\ & d_{uw} \le d_{uv} + d_{vw} && \text{for all } u, v, w. \end{aligned} \tag{4}$$

This problem has $\binom{n}{2} = O(n^2)$ variables and $O(n^3)$ constraints, so can be solved efficiently according to Theorem 2 applies, as remarked in Section 2.2.

We will now devote the following sections to the converse direction of Equation 3: how large can the gap between $\lambda^*$ and $\phi^*$ be, and how can we recover a cut from the optimal $d$?

### 3.5 Rounding

After solving the relaxed LP problem, we use two results to round the output of the previous problem to the best cut over $S$:

**Lemma 13** (A: $\ell_1$ metrics are made of cuts)**.** *Every $\ell_1$ metric on $V$ is a non-negative combination of cut metrics: $d = \sum_k \alpha_k d_{S_k}$ with $\alpha_k \geq 0$. Consequently $\mathrm{CUT}_n$ is exactly the set of $\ell_1$ metrics on $V$.*

**Theorem 14** (B: Bourgain 1985 [7])**.** *Every metric $d$ on $n$ points embeds into $\ell_1$ with distortion $\eta = O(\log n)$: there is $f : V \to \mathbb{R}^M$ with*

$$\frac{d(u,v)}{\eta} \;\leq\; \|f(u) - f(v)\|_1 \;\leq\; d(u,v) \qquad \textit{for all } u, v.$$

Lemma 13 is proved in Section 4 and Theorem 14 in Section 5. Taking them for granted, we can already describe the algorithm and prove Theorem 6.

**Algorithm.**

1. Solve the LP (4); let $d$ be an optimal metric, so that $\sum_{u<v} d = 1$ and $\sum_{uv \in E} d = \lambda^*$.
2. Apply Bourgain Theorem (Theorem 14) to $(V, d)$ to get $f : V \to \mathbb{R}^M$, and set
$$d'(u,v) := \|f(u) - f(v)\|_1 ,$$
so that pointwise
$$\frac{d(u,v)}{\eta} \leq d'(u,v) \leq d(u,v).$$
3. For every coordinate $j \in [M]$ and every threshold $t$, form the cut $\{u : f_j(u) < t\}$; among all the nontrivial cuts obtained, output the one of smallest sparsity.

**Step 1** Let $R$ be the ratio of a nonzero metric, defined as

$$R(d) := \frac{\sum_{uv \in E} d(u,v)}{\sum_{u<v} d(u,v)}. \tag{5}$$

First notice that $R(d_S) = \phi(S)$. Then, we claim that

$$R(d') \leq \eta \cdot \lambda^*. \tag{6}$$

Indeed,

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

**Step 2** By Lemma 13, $d' = \sum_k \alpha_k d_{S_k}$ with $\alpha_k > 0$ and each $S_k$ nontrivial. Then, from (5), since a weighted average of ratios is at least their minimum, we have

$$\min_k \phi(S_k) \leq \frac{\sum_k \alpha_k E(S_k, \bar{S}_k)}{\sum_k \alpha_k |S_k| |\bar{S}_k|} = R(d') \underset{(6)}{\leq} \eta\lambda^* \underset{(3)}{\leq} \eta\phi^*.$$

Notice that the cuts $S_k$ are threshold cuts of the coordinates of $f$, so they are among the candidates enumerated in step 3 of the algorithm. The cut $S$ that the algorithm outputs is at least as sparse as every candidate

**Step 3** Step 3 needs only $M \cdot n$ candidate cuts: for a fixed coordinate $j$, sweeping $t$ through $\mathbb{R}$ produces at most $n$ distinct sets, namely the prefixes of the vertices sorted by $f_j$. Step 3 needs only $M \cdot n$ candidate cuts: for a fixed coordinate $j$, sweeping $t$ through $\mathbb{R}$ produces at most $n$ distinct sets, namely the prefixes of the vertices sorted by $f_j$. Both $M$ and the cost of evaluating $\phi$ on a cut are polynomial. The embedding construction is randomized: checking the required distance inequalities for all pairs and resampling if necessary gives an algorithm with expected polynomial running time.

**Putting everything together** The following corollary directly proves Theorem 6.

**Corollary 15.** *The cut $S$ output by the algorithm satisfies $\phi(S) \leq \eta \cdot \lambda^* \leq \eta \cdot \phi^* = O(\log n) \cdot \phi^*$.*

The rounding algorithm can therefore be summarized with the following scheme:

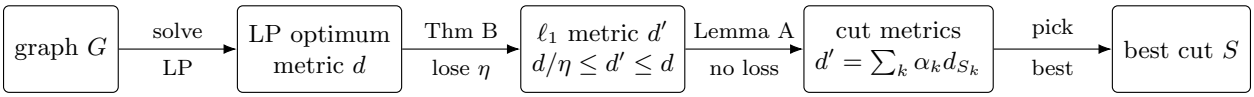



Figure 2: The three stages of the rounding algorithm. Only the middle stage costs a factor $\eta = O(\log n)$, while the decomposition into cuts is exact.

# 4 Proving Lemma A

Now we prove the Lemma 13 used in the algorithm. The idea is as follows. We define the line metric:

**Definition 16** (Line metric)**.** *A metric $d$ on $V$ is a* line metric *if $d(u, v) = |g(u) - g(v)|$ for some $g : V \to \mathbb{R}$.*

Then, we can use the line metric to express the $\ell_1$ norm between two values of $f(\cdot)$ as a sum of line metrics over each coordinate for some function $f = (f_1, \ldots, f_M) : V \to \mathbb{R}^M$:

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

Then, to prove the lemma, we proceed by looking at how random thresholds divide the line metrics on the real line. First, rescale the output of $g$ to the $[0, 1]$ interval. We can pick a random threshold $\tau \sim U[0, 1]$, such that it generates the set

$$S_\tau := \{u : g(u) < \tau\}$$

For distances $d(u, v)$, with $g(u) \leq g(v)$, we have that the $u$ and $v$ are separated by $S_\tau$, iff

$$g(u) < \tau \leq g(v)$$

Then, we can write that

$$\mathbb{E}[d_s(u, v)] = \Pr[g(u) < \tau \leq g(v))] = g(v) - g(u) = d(u, v),$$

which implies that an $\ell_1$ metric which is a sum of different line metrics can be written as a sum of probability weighted cut metrics. Indeed, let $L_k = \max_v f_k(v) - \min_v f_k(v)$

$$\begin{aligned}
\|f(u) - f(v)\|_1 &= \sum_{k=1}^{M} L_k \mathbb{E}\left[d_{S_\tau^k}(u, v)\right] \\
&= \sum_{k=1}^{M} L_k \sum_{A \subseteq V} \Pr\left(S_\tau^k = A\right) d_A(u, v) \\
&= \sum_{A \subseteq V} \left(\sum_{k=1}^{M} L_k \Pr\left(S_\tau^k = A\right)\right) d_A(u, v) \\
&= \sum_{A \subseteq V} \alpha_A d_A(u, v)
\end{aligned}$$

with $\alpha_A := \sum_{k:L_k>0} L_k \, \mathbb{P}[S_\tau^k = A]$. This proves the lemma.

# 5 Proving Bourgain's embedding theorem

Now, we prove the Theorem 14. The proof will be composed of two steps:

- *Averaging.* We construct a random map to the real line that never expands distances and on average, shrinks them by at most a factor $\eta$.
- *Bourgain's map.* We sample independently random copies of the map $x \mapsto d(x, S)$ and use them as coordinates of an $\ell_1$ embedding.

## 5.1 Random distance-to-set maps

Fix a nonempty set $S \subseteq X$ and define the mapping $\pi : x \mapsto d(x, S)$ as

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

Then, by triangle inequality, $d(x,S) \le d(x,y) + d(y,S)$ and vice-versa $d(y,S) \le d(x,y) + d(x,S)$ giving

$$|d(x,S) - d(y,S)| \le d(x,y)$$

This means that distance-to-set maps never expand distances (they are 1-Lipschitz).

A fixed choice of $S$ may collapse the distance between two points. We therefore choose $S$ at random and consider several sampling densities. As an illustrative case, fix $x, y$ with $D := d(x,y) > 0$. Suppose that $B(y, D/4)$ contains $M$ points and $B(x, D/2)$ contains at most $CM$ points, where $C$ is a fixed constant. Include each point independently with probability $1/(2M)$. The two balls are disjoint, so the events of hitting the first and missing the second are independent. Each event has probability bounded below by a positive constant depending only on $C$. When both occur,

$$d(y,S) \le \frac{D}{4}, \quad d(x,S) \ge \frac{D}{2} \qquad \Longrightarrow \qquad |d(x,S) - d(y,S)| \ge \frac{D}{4}.$$

This illustrates how an appropriate sampling density can separate a pair of points. For a general metric, however, we do not assume that a single density gives such a large contribution. Instead, we use the dyadic densities $2^{-1}, \dots, 2^{-m}$. The proof below associates an increment $\hat{r}_i - \hat{r}_{i-1}$ with each scale and shows that its expected contribution is at least $c_0(\hat{r}_i - \hat{r}_{i-1})$. These increments sum to $D/2$. Choosing the scale uniformly therefore gives expected separation at least $c_0 D/(2m)$, which yields logarithmic distortion because $m = O(\log n)$.

### 5.2 Step 1: averaging random one-dimensional maps

**Lemma 17** (Averaging)**.** *Let $\pi : X \to \mathbb{R}$ be a random map such that for all $x, y \in X$:*

*(a) $|\pi(x) - \pi(y)| \le d(x,y)$ always (with probability 1);*

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

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

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

The intuition behind this lemma is: a random map that is good on average, pair by pair can be converted into a single deterministic map that is good simultaneously for all pairs, at the cost of a factor 2. Before proving it, here is the idea in the easy case.

**Warm-up: $\pi$ has finite support** Let $\pi = \pi^{(k)}$ with probability $p_k$ for $k = 1, \dots, K$. Define $f$ by weighting outcomes and probabilities:

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

Then

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

By assumptions (a) and (b), then, $\frac{d(x,y)}{\eta} \leq (\star) \leq d(x,y)$. In this case distortion is exactly $\eta$; in general, $K$ may be exponential. The fix will be to sample instead of enumerating (and this is where the factor 2 is spent).

*Proof of Lemma 17.* Take $T$ independent copies of $\pi$ and let

$$f(x) := \frac{1}{T}\left(\pi_1(x), \dots \pi_T(x)\right)$$

Fix $x, y$ and write $D := d(x,y)$, $Z_t := |\pi_t(x) - \pi_t(y)|$. The random variables $Z_t$ are i.i.d, take value into $[0, D]$ by (a) and have mean $\mu := \mathbb{E}[Z_1] \geq D/\eta$ by (b). By construction

$$\|f(x) - f(y)\|_1 = \frac{1}{T}\sum_{t=1}^{T} Z_t =: \bar{Z}_t$$

Now the upper bound is deterministic: since $Z_t \leq D$ always, $\bar{Z}_t \leq D$.

For the lower bound, $0 \leq Z_1 \leq D \implies 0 \leq Z_1^2 \leq DZ_1 \implies \mathbb{E}\, Z_1^2 \leq D\mu$. Therefore,

$$\operatorname{Var} \bar{Z} = \frac{\operatorname{Var} Z_1}{T} \leq \frac{\mathbb{E}[Z_1^2]}{T} \leq \frac{D\mu}{T}.$$

By Chebyshev inequality,

$$\mathbb{P}\Big[\bar{Z} < \frac{\mu}{2}\Big] \;\leq\; \frac{4 \operatorname{Var} \bar{Z}}{\mu^2} \;\leq\; \frac{4D}{T\mu} \underset{\mu \geq D/\eta}{\leq} \frac{4\eta}{T}.$$

Choosing $T := \lceil 4\eta n^2 \rceil$, each pair fails with probability at most $1/n^2$. There are $\binom{n}{2} < n^2/2$ pairs, so by a union bound, with probability greater than $1/2$ every pair satisfies

$$\|f(x) - f(y)\|_1 = \bar{Z} \geq \frac{\mu}{2} \geq \frac{d(x,y)}{2\eta}.$$

In particular such an $f$ exists, which is what we claimed. □

*Remark* 18. Chebyshev's inequality would already be enough to show that the empirical average is close to its expectation with some probability. Since the coordinates are sampled independently and are bounded, we can obtain a more efficient guarantee over all pairs by using Chernoff bound instead, which gives an exponentially small failure probability. This would allow us to take only $T = O(\eta \log n)$ independent coordinates; this gives $T = O(\log^2 n)$.

### 5.3 Step 2: Bourgain's random map

It remains to build a random 1-Lipschitz map that, in expectation, does not shrink the distance too much. By our earlier observation, $\pi(x) = d(x, S)$ is 1-Lipschitz for every $S$. Thus, property (a) holds and so it could be used as the random map if property (b) also holds. This turns out to be true.

**Lemma 19** (Random scale)**.** *Let $m := \lfloor \log_2 n \rfloor + 2$. Choose $k \in \{1, \dots, m\}$ uniformly at random, then include each point in $S \subseteq V$ independently with probability $2^{-k}$. Define $\pi(x) := d(x, S)$, using the convention $d(x, \varnothing) = 0$. Then for all $x, y$:*

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

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

Given the Lemma, Theorem 14 follows by simply setting $\eta = 2m/c_0$, so we obtain an $\ell_1$ embedding of distortion $2\eta = 4m/c_0 = O(\log n)$. Therefore, we now try to prove statement (b).

The key element of the proof is the choice of $S$. First, we can re-write the statement to a statement about scales. Let $S_i$ include every point independently with probability $2^{-i}$, $i = 1, \dots, m$. Conditioning on the choice of $k$,

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

so (b) is equivalent to the following statement:

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

### 5.3.1 Growing balls around $x$ and $y$

To prove ($\star$) we assign to each scale $i$ a range of distances that it is responsible for, and show that it recovers a constant fraction of that range; since sampling at density $2^{-i}$ reacts to how many points a region contains, these ranges are defined by growing balls around $x$ and $y$ until they contain $2^i$ points.

Now, fix $x, y$ with $D := d(x, y) > 0$ (if $D = 0$ there is nothing to prove). Let $B(x, r) := \{z : d(x, z) \le r\}$ the closed ball centered at $x$ of radius $r \ge 0$ and define

$$N(r) := |B(x, r)| + |B(y, r)|, \qquad r_i := \min\big\{r : N(r) \ge 2^i\big\}, \qquad \hat{r}_i := \min(r_i, D/2),$$

with $r_i := \infty$ if $2^i > 2n$ (in which case no radius achieves the count, since $N(r) \le 2n$ always). The truncation of $\hat{r}_i$ makes sure that the two balls are disjoint (see Remark 22). The sequences are nondecreasing: $r_0 \le r_1 \le \cdots \le r_m$ and likewise for $\hat{r}$.

**Fact 20** (Three facts about the radii)**.** *With the notation above, and writing $B^\circ(x, r) := \{z : d(x, z) < r\}$ for the* open *ball:*

*(F0)* $\hat{r}_0 = 0$ *and* $\hat{r}_m = D/2$;

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

*(F2)* $|B(x, r_{i-1})| + |B(y, r_{i-1})| \ge 2^{i-1}$, *whenever* $r_{i-1} < \infty$.

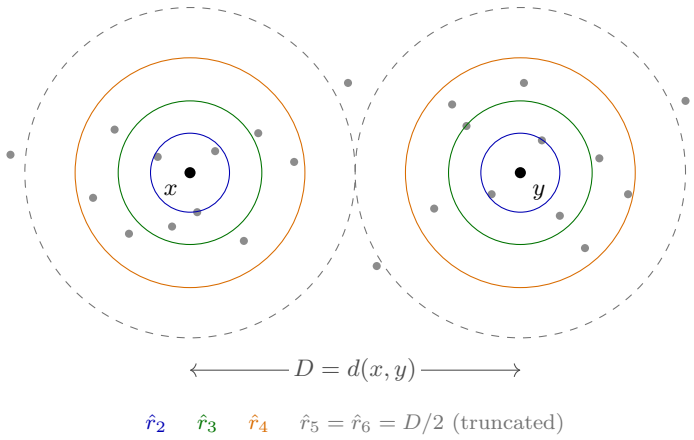



Figure 3: The radii $\hat{r}_i$ grow until they are cut off at $D/2$. Each circle marks the radius at which the two balls, counted together, contain $2^i$ points; beyond $D/2$ the radii are frozen, so the ball around $x$ can never reach $y$ and vice versa.

*Proof.* (F0) $N(0) \geq 2 \geq 2^0$, so $r_0 = 0$ and $\hat{r}_0 = 0$. For the second part, note $m = \lfloor \log_2 n \rfloor + 2 > \log_2 n + 1$, so $2^m > 2n \geq N(r)$ for every $r$; hence $r_m = \infty$ and $\hat{r}_m = D/2$.

(F1) If $\hat{r}_i = 0$ both open balls are empty and there is nothing to prove, so assume $\hat{r}_i > 0$. Suppose first $\hat{r}_i = r_i$. By minimality of $r_i$ we have $N(r) < 2^i$ for every $r < r_i$. Let $\rho$ be the largest distance smaller than $r_i$ realised from $x$ or from $y$ (it exists, since $V$ is finite and the distance 0 from $x$ to itself qualifies). Every point at distance $< r_i$ from $x$ is then at distance $\leq \rho$ from $x$, and likewise for $y$, so $|B^\circ(x, r_i)| + |B^\circ(y, r_i)| = N(\rho) < 2^i$. If instead $\hat{r}_i = D/2 < r_i$, then the open balls of radius $D/2$ are contained in the closed balls of the same radius, and $N(D/2) < 2^i$ by minimality of $r_i$ again.

(F2) If $r_{i-1} < \infty$, then $N(r_{i-1}) \geq 2^{i-1}$ by the definition of $r_{i-1}$ (the minimum is attained because $V$ is finite and $N$ is a nondecreasing step function that is right-continuous in $r$). □

### 5.3.2 Key claim: every scale contributes

**Claim 21.** *For every $i \in \{1, \ldots, m\}$,*

$$\mathbb{E}\big|d(x, S_i) - d(y, S_i)\big| \;\geq\; c_0\,(\hat{r}_i - \hat{r}_{i-1}), \qquad c_0 = \tfrac{1}{4}(1 - e^{-1/4}).$$

At scale $i$ we hope for two things to happen at once:

- the sample $S_i$ misses a small ball around $x$, which forces $d(x, S_i)$ to be large (at least the radius of that ball);
- the sample $S_i$ hits a larger ball around $y$, which forces $d(y, S_i)$ to be small (at most the radius of that ball).

The two balls are disjoint, so the two events are independent, and the density $2^{-i}$ is such that each of them has constant probability: the small ball has fewer than $2^i$ points, so a sample of density

$2^{-i}$ misses it with constant probability, while the large ball has at least $2^{i-2}$ points, so the same sample hits it with constant probability. Figure 4 shows the configuration.

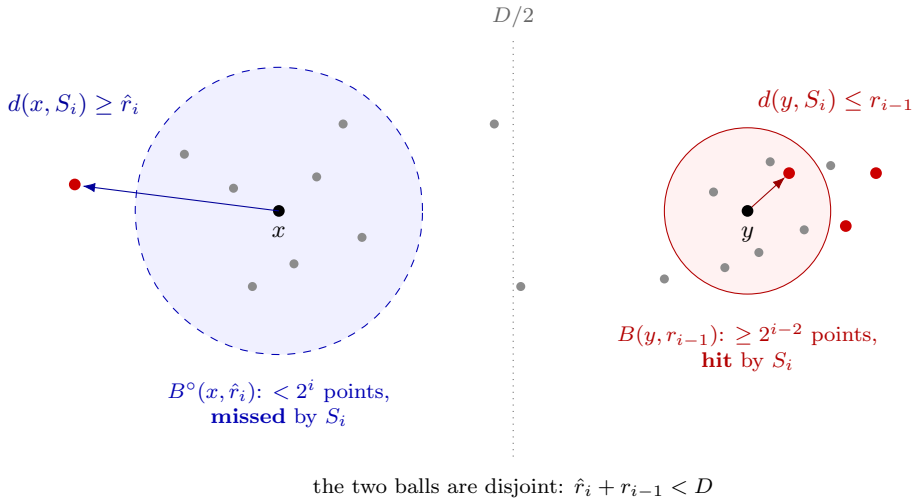



Figure 4: The good event $A$ at scale $i$. Red points belong to the sample $S_i$. Missing the small ball around $x$ makes $d(x, S_i)$ large; hitting the ball around $y$ makes $d(y, S_i)$ small. Disjointness of the two balls makes the two requirements independent.

*Proof of Claim 21.* If $\hat{r}_i = \hat{r}_{i-1}$ the right-hand side is 0 and there is nothing to prove, so assume $\hat{r}_i > \hat{r}_{i-1}$. Then $\hat{r}_{i-1} < D/2$, which forces $\hat{r}_{i-1} = r_{i-1} < D/2$; in particular $r_{i-1}$ is finite and (F2) applies. Since $|B(x, r_{i-1})| + |B(y, r_{i-1})| \geq 2^{i-1}$, at least one of the two balls contains $2^{i-2}$ points; as the quantity $|d(x, S_i) - d(y, S_i)|$ is symmetric in $x$ and $y$, and (F1) is symmetric too, we may assume without loss of generality that

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

Consider the small ball $B^\circ(x, \hat{r}_i)$, which has fewer than $2^i$ points by (F1), and the *large* ball $B(y, r_{i-1})$, which has at least $2^{i-2}$ points. They are disjoint: a common point $z$ would satisfy $d(x, z) < \hat{r}_i \leq D/2$ and $d(z, y) \leq r_{i-1} < D/2$, whence $D = d(x, y) \leq d(x, z) + d(z, y) < D$, a contradiction.

So, we define the good event as

$$A := \big\{S_i \cap B^\circ(x, \hat{r}_i) = \varnothing\big\} \ \cap \ \big\{S_i \cap B(y, r_{i-1}) \neq \varnothing\big\}.$$

On $A$ no sample point lies strictly within $\hat{r}_i$ of $x$, so $d(x, S_i) \geq \hat{r}_i$; and some sample point lies within $r_{i-1}$ of $y$, so $d(y, S_i) \leq r_{i-1} = \hat{r}_{i-1}$. Hence on $A$,

$$\big|d(x, S_i) - d(y, S_i)\big| \ \geq \ \hat{r}_i - \hat{r}_{i-1} > 0,$$

and since the random variable $|d(x, S_i) - d(y, S_i)|$ is non-negative,

$$\mathbb{E}\big|d(x, S_i) - d(y, S_i)\big| \ \geq \ \mathbb{P}[A] \cdot (\hat{r}_i - \hat{r}_{i-1}). \tag{7}$$

Now we see that the good event has constant probability. Write $p := 2^{-i} \le \frac{1}{2}$. The two events defining $A$ concern disjoint sets of points, and each point is included in $S_i$ by an independent coin flip, so they are independent and

$$\mathbb{P}[A] = \underbrace{(1-p)^{|B^\circ(x,\hat{r}_i)|}}_{\ge (1-p)^{2^i}} \cdot \underbrace{\Big(1-(1-p)^{|B(y,r_{i-1})|}\Big)}_{\ge 1-e^{-p2^{i-2}}}.$$

For the first factor we used $|B^\circ(x,\hat{r}_i)| < 2^i$ and that $(1-p)^u$ decreases in $u$; then

$$(1-p)^{2^i} = (1-p)^{1/p} \;\ge\; \tfrac{1}{4} \qquad \text{for } 0 < p \le \tfrac{1}{2},$$

because $p \mapsto (1-p)^{1/p}$ is decreasing on $(0, \frac{1}{2}]$ and equals $(\frac{1}{2})^2 = \frac{1}{4}$ at $p = \frac{1}{2}$. For the second factor we used $|B(y,r_{i-1})| \ge 2^{i-2}$, that $1-(1-p)^u$ increases in $u$, and $1-p \le e^{-p}$; since $p\,2^{i-2} = 2^{-i}2^{i-2} = \frac{1}{4}$, this factor is at least $1 - e^{-1/4}$. Altogether

$$\mathbb{P}[A] \;\ge\; \tfrac{1}{4}\big(1-e^{-1/4}\big) = c_0 \;>\; \tfrac{1}{20},$$

and the claim follows from (7). □

### 5.3.3 Telescoping finishes the proof

Summing Claim 21 over $i = 1, \dots, m$, the increments telescope, and (F0) evaluates the two ends:

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

This is exactly ($\star$), which proves Lemma 19(b), hence (with Lemma 17) Theorem 14, hence Theorem 6. □

*Remark* 22 (Why truncate the radii at $D/2$?). The truncation does two jobs at once. First, it is what makes the two balls in Claim 21 disjoint, and disjointness is what makes "miss the ball around $x$" and "hit the ball around $y$" independent events. Untruncated radii would eventually exceed $D$, at which point the ball around $x$ swallows $y$, and the good event becomes unlikely. Second, it is what makes the telescoping sum reach $D/2$. Consider a space consisting of two tight clusters of $n/2$ points each, one within distance $\varepsilon$ of $x$ and the other within distance $\varepsilon$ of $y$, with $\varepsilon \ll D$. Already at $r = \varepsilon$ we have $N(r) = n$, so $r_i \approx \varepsilon$ for every $i$ with $2^i \le n$; and the next radii jump straight past $D/2$, to about $D$, the distance at which the ball around $x$ finally reaches the other cluster. Without truncation the usable scales – those with $r_i < D/2$, the only ones for which the disjointness step of Claim 21 survives – therefore contribute a total of about $c_0\varepsilon$, and the rest of the distance is lost. With truncation, the first scale $i$ whose radius exceeds $D/2$ has $\hat{r}_i = D/2$ and contributes the single increment $D/2 - \varepsilon$, which is essentially the whole bound; such a scale exists because $r_m = \infty$ forces $\hat{r}_m = D/2$ by (F0). And its contribution is legitimate: at that scale the density is $\approx 1/(2n)$, so the sample misses the cluster around $x$ (of size $n/2$) with constant probability and hits the cluster around $y$ (which has at least $2^{i-2}$ points, by (F2)) with constant probability – exactly the good event.

# 6 Summary

The full argument can be compressed into four inequalities and two geometric facts:

$$1.\ \lambda^\star \le \phi^\star, \qquad 2.\ R(d') \le O(\log n)\lambda^\star, \qquad 3.\ \min_k \phi(S_k) \le R(d'),$$

where

$$4.\ \text{every metric embeds into } \ell_1 \text{ with } O(\log n) \text{ distortion, and}$$
$$5.\ \ell_1 \text{ metrics are nonnegative sums of cuts.}$$

Consequently,

$$6.\ \phi(S) \le O(\log n)\phi^\star.$$

The conceptual takeaways are:

1. A linear program optimises a linear objective over a polyhedron; it is solvable in polynomial time, and the ellipsoid method needs only a separation oracle, so even exponentially many constraints are acceptable.
2. Infeasibility of $Ax \le b$ always has a short certificate (Farkas' lemma), and the same duality gives, for every LP, a dual whose feasible points certify bounds on the primal optimum.
3. Relaxation and rounding: enlarge the feasible set to something tractable, solve, and round back at a controlled loss $\gamma$; the guarantee obtained is really a bound on the integrality gap.
4. Sparsest cut is the minimisation of the ratio $R(d) = \sum_{uv\in E} d / \sum_{u<v} d$ over the cut cone $\mathrm{CUT}_n$. Relaxing $\mathrm{CUT}_n$ to the metric cone $\mathrm{MET}_n$ gives an LP with $\binom{n}{2}$ variables and $O(n^3)$ constraints, of value $\lambda^* \le \phi^*$.
5. Every $\ell_1$ metric is a non-negative combination of cut metrics – via coordinates and threshold cuts – and a combination of cuts is never sparser than its sparsest constituent. Passing from $\ell_1$ metrics to cuts is therefore lossless.
6. Bourgain's theorem embeds any $n$-point metric into $\ell_1$ with distortion $O(\log n)$, by averaging the random maps $x \mapsto d(x, S)$ over $\log n$ densities of a random set $S$. This is the only lossy step.
7. Together: a polynomial-time $O(\log n)$-approximation for sparsest cut.

**LLM disclosure** LLM were used to re-write our notes and the lecture slides together, to better understand the material, and to produce the basic draft of the document. The lecture notes were subsequently manually edited by the authors, and finally the model was used for proofreading and formatting. The TikZ figures from the presentation were recreated using the LLM.

## References


[1] Narendra Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984. DOI: 10.1007/BF02579150.

[2] Leonid G. Khachiyan. A polynomial algorithm in linear programming. *Soviet Mathematics Doklady*, 20(1):191–194, 1979. Russian original: *Doklady Akademii Nauk SSSR* 244 (1979), 1093–1096.

[3] Martin Grötschel, László Lovász, and Alexander Schrijver. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, 1(2):169–197, 1981. DOI: 10.1007/BF02579273. Corrigendum: *Combinatorica* 4 (1984), 291–295.

[4] Julius Farkas. Theorie der einfachen Ungleichungen. *Journal für die reine und angewandte Mathematik*, 124:1–27, 1902. DOI: 10.1515/crll.1902.124.1.

[5] F. Thomson Leighton and Satish Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *Journal of the ACM*, 46(6):787–832, 1999. DOI: 10.1145/331524.331526. Preliminary version in *FOCS* 1988, 422–431.

[6] Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995. DOI: 10.1007/BF01200757.

[7] Jean Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel Journal of Mathematics*, 52(1–2):46–52, 1985. DOI: 10.1007/BF02776078.