# 15

# Embedding Finite Metric Spaces Into Normed Spaces

## 15.1 Introduction: Approximate Embeddings

We recall that a *metric space* is a pair $(X, \rho)$, where $X$ is a set and $\rho: X \times X \to [0, \infty)$ is a *metric*, satisfying the following axioms: $\rho(x, y) = 0$ if and only if $x = y$, $\rho(x, y) = \rho(y, x)$, and $\rho(x, y) + \rho(y, z) \geq \rho(x, z)$.

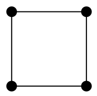
A metric $\rho$ on an $n$-point set $X$ can be specified by a $n \times n$ matrix of real numbers (actually $\binom{n}{2}$ numbers suffice because of the symmetry). Such tables really arise, for example, in microbiology: $X$ is a collection of bacterial strains, and for every two strains, one can obtain their *dissimilarity*, which is some measure of how much they differ. Dissimilarity can be computed by assessing the reaction of the considered strains to various tests, or by comparing their DNA, and so on.[1] It is difficult to see any structure in a large table of numbers, and so we would like to represent a given metric space in a more comprehensible way.

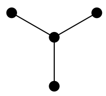
For example, it would be very nice if we could assign to each $x \in X$ a point $f(x)$ in the plane in such a way that $\rho(x, y)$ equals the Euclidean distance of $f(x)$ and $f(y)$. Such representation would allow us to see the structure of the metric space: tight clusters, isolated points, and so on. Another advantage would be that the metric would now be represented by only $2n$ real numbers, the coordinates of the $n$ points in the plane, instead of $\binom{n}{2}$ numbers as before. Moreover, many things concerning a point set in the plane can be computed by efficient geometric algorithms, which are not available for an arbitrary metric space.

This sounds very good, and indeed it is too good to be generally true: it is easy to find examples of small metric spaces which cannot be represented in this way by a planar point set. One example are 4 points, each two of them at distance 1: such 4 points cannot be found in the plane. On the other hand, they exist in the 3-dimensional Euclidean space.

[1] There are various measures of dissimilarity, and not all of them yield a metric, but many do.

Perhaps less obviously, there are 4-point metric spaces that cannot be represented (exactly) in *any* Euclidean space. Here are two examples:





The metric on these 4-point set is given by the indicated graphs; that is, the distance of two points is the number of edges of a shortest path connecting them in the graph. For example, in the second picture, the center has distance 1 from the leaves and the mutual distances of the leaves are 2.

So far we have considered *isometric embeddings.* A mapping $f: X \to Y$, where $X$ is a metric space with a metric $\rho$ and $Y$ is a metric space with a metric $\sigma$, is called an isometric embedding if it preserves distances, i.e. if $\sigma(f(x), f(y)) = \rho(x, y)$ for all $x, y \in X$. But in many applications, we need not insist on preserving the distance exactly; rather, we can allow some distortion, say by 10%. A notion of an approximate embedding is captured by the following definition.

**15.1.1 Definition ($D$-embedding of metric spaces).** *A mapping $f: X \to Y$, where $X$ is a metric space with a metric $\rho$ and $Y$ is a metric space with a metric $\sigma$, is called a $D$-*embedding, *where $D \geq 1$ is a real number, if there exists a number $r > 0$ such that for all $x, y \in X$,*

$$r \cdot \rho(x, y) \leq \sigma(f(x), f(y)) \leq D \cdot r \cdot \rho(x, y).$$

*The infimum of the numbers $D$ such that $f$ is a $D$-embedding is called the* distortion *of $f$.*

Note that this definition permits scaling of all distances in the same ratio $r$, in addition to the distortion of the individual distances by factors between 1 and $D$. If $Y$ is a Euclidean space (or a normed space), we can re-scale the image at will, and so we can choose the scaling factor $r$ at our convenience.

Mappings with a bounded distortion are sometimes called *bi-Lipschitz mappings.* This is because the distortion of $f$ can be equivalently defined using the Lipschitz constants of $f$ and of the inverse mapping $f^{-1}$. Namely, if we define the *Lipschitz norm* of $f$ by $\|f\|_{Lip} = \sup\{\sigma(f(x), f(y))/\rho(x, y) \colon x, y \in X, x \neq y\}$ then the distortion of $f$ equals $\|f\|_{Lip} \cdot \|f^{-1}\|_{Lip}$.

We are going to study the possibility of $D$-embedding of $n$-point metric spaces into Euclidean spaces and into various normed spaces. As usual, we cover only a small sample of results. Many of them are negative, showing that certain metric spaces cannot be embedded too well. But in the subsequent Section 15.2, we start on an optimistic note: we present a positive result of considerable theoretical and practical importance. Before that, we review a few definitions concerning $\ell_p$-spaces.

**The spaces $\ell_p$ and $\ell_p^d$.** For a point $x \in \mathbf{R}^d$ and $p \in [1, \infty)$, let

$$\|x\|_p = \left( \sum_{i=1}^{d} |x_i|^p \right)^{1/p}$$

denote the $\ell_p$-*norm* of $x$. Most of the time, we will consider the case $p = 2$, i.e. the usual Euclidean norm $\|x\|_2 = \|x\|$. Another particularly important case is $p = 1$, the $\ell_1$-norm (sometimes called the Manhattan distance). The $\ell_\infty$-norm, or *maximum norm*, is given by $\|x\|_\infty = \max_i |x_i|$.

Let $\ell_p^d$ denote the space $\mathbf{R}^d$ equipped with the $\ell_p$ norm. In particular, we write $\ell_2^d$ in order to stress that we mean $\mathbf{R}^d$ with the usual Euclidean norm.

Sometimes we are interested in embeddings into *some* space $\ell_p^d$, with $p$ given but without restrictions on the dimension $d$; for example, we can ask if a given metric space embeds isometrically into any Euclidean space. Then it is convenient to speak about $\ell_p$, which is the space of all infinite sequences $x = (x_1, x_2, \ldots)$ of real numbers $\|x\|_p < \infty$, where $\|x\|_p = \left(\sum_{i=1}^\infty |x_i|^p\right)^{1/p}$. In particular, $\ell_2$ is the *(separable) Hilbert space.* The space $\ell_p$ contains each $\ell_p^d$ isometrically, and it can be shown that any finite metric space isometrically embeddable into $\ell_p$ can be isometrically embedded into $\ell_p^d$ for some $d$. (In fact, every $n$-point subspace of $\ell_p$ can be isometrically embedded into $\ell_p^d$ with $d \le \binom{n}{2}$; see Exercise 15.5.2.)

Although the spaces $\ell_p$ are interesting mathematical objects, we will not really study them; we only use embeddability into $\ell_p$ as a convenient shorthand for embeddability into $\ell_p^d$ for some $d$.

**Bibliography and Remarks.** This chapter aims at providing an overview of important results concerning low-distortion embeddings of finite metric spaces. The scope is relatively narrow and we almost do not discuss even closely related areas, such as isometric embeddings. Another recent survey, with fewer proofs and mainly focused on algorithmic aspects, is Indyk [Ind01].

For studying approximate embeddings, it may certainly be helpful to understand isometric embeddings, and here extensive theory is available. For example, several ingenious characterizations of isometric embeddability into $\ell_2$ can be found in old papers of Schoenberg (e.g. [Sch38], building on the work of mathematicians like Menger and von Neumann). A recent book concerning isometric embeddings, and embeddings into $\ell_1$ in particular, is Deza and Laurent [DL97].

Another closely related area is the investigation of bi-Lipschitz maps, usually $(1+\varepsilon)$-embeddings with $\varepsilon > 0$ small, defined on an open subset of a Euclidean space (or a Banach space) and being local homeomorphisms. These mappings are called *quasi-isometries* (the definition of a quasi-isometry is slightly more general, though), and the main question is how close to an isometry such a mapping has to be, in terms of the dimension and $\varepsilon$; see Benyamini and Lindenstrauss [BL99], Chap. 14 and 15, for an introduction.

## Exercises

1. Consider the two 4-point examples (the square and the star); prove that they cannot be isometrically embedded into $\ell_2^2$. $\boxed{2}$ Can you determine the minimum necessary distortion for embedding into $\ell_2^2$?

2. (a) Prove that a bijective mapping $f$ between metric spaces is a $D$-embedding if and only if $\|f\|_{Lip} \cdot \|f^{-1}\|_{Lip} \le D$. $\boxed{1}$

   (b) Let $(X, \rho)$ be a metric space, $|X| \ge 3$. Prove that the distortion of an embedding $f: X \to Y$, where $(Y, \sigma)$ is a metric space, equals the supremum of the factors by which $f$ "spoils" the ratios of distances; that is,

$$\sup\left\{\frac{\sigma(f(x),f(y))/\sigma(f(z),f(t))}{\rho(x,y)/\rho(z,t)}: x,y,z,t\in X,\, x\neq y, z\neq t\right\}.$$

$\boxed{2}$

## 15.2 The Johnson–Lindenstrauss Flattening Lemma

It is easy to show that there is no isometric embedding of the vertex set $V$ of an $n$-dimensional regular simplex into a Euclidean space of dimension $k<n$. In this sense, the $(n{+}1)$-point set $V\subset\ell_2^n$ is truly $n$-dimensional. The situation changes drastically if we do not insist on exact isometry: as we will see, the set $V$, and any other $(n{+}1)$-point set in $\ell_2^n$, can be almost isometrically embedded into $\ell_2^k$ with $k=O(\log n)$ only!

**15.2.1 Theorem (Johnson–Lindenstrauss Flattening Lemma).** *Let $X$ be an $n$-point set in a Euclidean space (i.e. $X\subset\ell_2$), and let $\varepsilon\in(0,1]$ be given. Then there exists a $(1{+}\varepsilon)$-embedding of $X$ into $\ell_2^k$, where $k=O(\varepsilon^{-2}\log n)$.*

This result shows that any metric question about $n$ points in $\ell_2^n$ can be considered for points in $\ell_2^{O(\log n)}$, if we do not mind a distortion of the distances by at most 10%, say. For example, to represent $n$ points in $\ell_2^n$ in a computer, we need to store $n^2$ numbers. To store all their distances, we need about $n^2$ numbers as well. But by the Flattening Lemma, we can store only $O(n\log n)$ numbers and still reconstruct any of the $n^2$ distances with 10% accuracy.

Various proofs of the Flattening Lemma, including the one below, provide efficient randomized algorithms that find the almost isometric embedding into $\ell_2^k$ quickly. Numerous algorithmic applications were found recently: for example, in fast clustering of high-dimensional point sets, in approximate searching for nearest neighbors, in approximate multiplication of matrices, and also in purely graph-theoretical problems, such as approximating the bandwidth of a graph or multicommodity flows.

The proof of Theorem 15.2.1 is based on the following lemma, of considerable independent interest.

**15.2.2 Lemma (Concentration of the length of the projection).** *For a unit vector $x\in S^{n-1}$, let*

$$f(x)=\sqrt{x_1^2+x_2^2+\cdots+x_k^2}$$

*be the length of the projection of $x$ on the subspace $L_0$ spanned by the first $k$ coordinates. Consider $x\in S^{n-1}$ chosen at random. Then $f(x)$ is sharply concentrated around a suitable number $m=m(n,k)$:*

$$\mathrm{P}[f(x)\geq m+t]\leq 2e^{-t^2n/2} \text{ and } \mathrm{P}[f(x)\leq m-t]\leq 2e^{-t^2n/2},$$

*where $\mathrm{P}$ is the uniform probability measure on $S^{n-1}$. For $n$ larger than a suitable constant and $k\geq 10\ln n$, we have $m\geq\frac{1}{2}\sqrt{\frac{k}{n}}$.*

In the lemma, the $k$-dimensional subspace is fixed and $x$ is random. Equivalently, if $x$ is a fixed unit vector and $L$ is a random $k$-dimensional subspace of $\ell_2^n$ (as introduced in Section 14.3), the length of the projection of $x$ on $L$ obeys the bounds in the lemma.

**Proof of Lemma 15.2.2.** The orthogonal projection $p\colon \ell_2^n \to \ell_2^k$ given by $(x_1,\ldots,x_n) \mapsto (x_1,\ldots,x_k)$ is 1-Lipschitz, and so $f$ is 1-Lipschitz as well. Lévy's Lemma (Theorem 14.3.2) gives the tail estimates as in the lemma with $m = \mathrm{med}(f)$. It remains to show the lower bound for $m$. It is not impossibly difficult to do it by elementary calculation (we need to find the measure of a simple region on $S^{n-1}$). But we can also avoid the calculation by a trick combined with a general measure concentration result.

For random $x \in S^{n-1}$, we have $1 = \mathbf{E}\left[\|x\|^2\right] = \sum_{i=1}^n \mathbf{E}\left[x_i^2\right]$. By symmetry, $\mathbf{E}\left[x_i^2\right] = \frac{1}{n}$, and so $\mathbf{E}\left[f^2\right] = \frac{k}{n}$. We now show that since $f$ is tightly concentrated, $\mathbf{E}\left[f^2\right]$ cannot be much larger than $m^2$ and so $m$ is not too small.

For any $t \geq 0$, we can estimate

$$\begin{aligned}\frac{k}{n} = \mathbf{E}\left[f^2\right] &\leq \mathrm{P}[f \leq m+t] \cdot (m+t)^2 + \mathrm{P}[f > m+t] \cdot \max_x (f(x)^2)\\ &\leq (m+t)^2 + 2e^{-t^2 n/2}.\end{aligned}$$

Let us set $t = \sqrt{\frac{k}{5n}}$. Since $k \geq 10 \ln n$, we have $2e^{-t^2n/2} \leq \frac{2}{n}$, and from the above inequality we calculate $m \geq \sqrt{\frac{k-2}{n}} - t \geq \frac{1}{2}\sqrt{\frac{k}{n}}$.

Let us remark that a more careful calculation shows that for all $k$, $m = \sqrt{\frac{k}{n}} + O(\frac{1}{\sqrt{n}})$. □

**Proof of the Flattening Lemma (Theorem 15.2.1).** We may assume that $n$ is sufficiently large. Let $X \subset \ell_2^n$ be a given $n$-point set. We set $k = 200\varepsilon^{-2} \ln n$ (the constant can be improved). If $k \geq n$ there is nothing to prove, so we assume $k < n$. Let $L$ be a random $k$-dimensional linear subspace of $\ell_2^n$ (obtained by a random rotation of $L_0$).

The chosen $L$ is a copy of $\ell_2^k$. We let $p\colon \ell_2^n \to L$ be the orthogonal projection onto $L$. Let $m$ be the number around which $\|p(x)\|$ is concentrated, as in Lemma 15.2.2. We prove that for any two distinct points $x, y \in \ell_2^n$, the condition

$$(1 - \tfrac{\varepsilon}{3})m\,\|x-y\| \leq \|p(x) - p(y)\| \leq (1 + \tfrac{\varepsilon}{3})m\,\|x-y\| \tag{15.1}$$

is violated with probability at most $n^{-2}$. Since there are fewer than $n^2$ pairs of distinct $x, y \in X$, there exist some $L$ such that (15.1) holds for all $x, y \in X$. In such a case, the mapping $p$ is a $D$-embedding of $X$ into $\ell_2^k$ with $D \leq \frac{1+\varepsilon/3}{1-\varepsilon/3} < 1+\varepsilon$ (for $\varepsilon \leq 1$).

Let $x$ and $y$ be fixed. First we reformulate the condition (15.1). Let $u = x - y$; since $p$ is a linear mapping, we have $p(x) - p(y) = p(u)$, and (15.1) can be rewritten to $(1-\frac{\varepsilon}{3})m\,\|u\| \leq \|p(u)\| \leq (1+\frac{\varepsilon}{3})m\,\|u\|$. This is invariant under scaling, and so we may suppose that $\|u\| = 1$. The condition thus becomes

$$\Big|\|p(u)\| - m\Big| \leq \tfrac{\varepsilon}{3} m. \tag{15.2}$$

By Lemma 15.2.2 and the remark following it, the probability of violating (15.2), for $u$ fixed and $L$ random, is at most

$$4e^{-\varepsilon^2 m^2 n/18} \leq 4e^{-\varepsilon^2 k/72} < n^{-2}.$$

This proves the Johnson–Lindenstrauss Flattening Lemma. □

**Alternative proofs.** There are several variations of the proof, which are more suitable from the computational point of view, i.e. if we really want to produce the embedding into $\ell_2^{O(\log n)}$.

In the above proof, we project the set $X$ on a random $k$-dimensional subspace $L$. Such an $L$ can be chosen by selecting an orthonormal basis $(b_1, b_2, \ldots, b_k)$, where $b_1, \ldots, b_k$ is a random $k$-tuple of unit orthogonal vectors. The coordinates of the projection of $x$ to $L$ are the scalar products $\langle b_1, x\rangle, \ldots, \langle b_k, x\rangle$. It turns out that the condition of orthogonality of the $b_i$ can be dropped. That is, we can pick unit vectors $b_1, \ldots, b_k \in S^{n-1}$ independently at random and define a mapping $p: X \to \ell_2^k$ by $x \mapsto (\langle b_1, x\rangle, \ldots, \langle b_k, x\rangle)$. Using suitable concentration results, one can verify that $p$ is a $(1+\varepsilon)$-embedding with probability close to 1. The procedure of picking the $b_i$ is computationally much simpler.

Another way is to choose each component of each $b_i$ from the normal distribution $N(0,1)$, all the $nk$ choices of the components being independent. The distribution of each $b_i$ in $\mathbf{R}^n$ is rotationally symmetric (as was mentioned in Section 14.1). Therefore, for every fixed $u \in S^{n-1}$, the scalar product $\langle b_i, u\rangle$ also has the normal distribution $N(0,1)$ and $\|p(u)\|^2$, the squared length of the image, has the distribution of $\sum_{i=1}^k Z_i^2$, where the $Z_i$ are independent $N(0,1)$. This is the well-known Chi-Square distribution with $k$ degrees of freedom, and a strong concentration result analogous to Lemma 15.2.2 can be found in books on probability theory (or derived from general measure-concentration results for the Gaussian measure or from Chernoff-type tail estimates). A still different method, particularly easy to implement but with a more difficult proof, uses independent random vectors $b_i \in \{-1, 1\}^n$.

**Bibliography and Remarks.** The Flattening Lemma is from Johnson and Lindenstrauss [JL84]. They were interested in the following question: given a metric space $Y$, an $n$-point subspace $X \subset Y$, and a 1-Lipschitz mapping $f: X \to \ell_2$, what is the smallest $C = C(n)$ such that there is always a $C$-Lipschitz mapping $\bar{f}: Y \to \ell_2$ extending $f$? They obtained the upper bound $C = O(\sqrt{\log n})$, together with an almost matching lower bound.

The alternative proof of the Flattening Lemma using independent normal random variables was given by Indyk and Motwani [IM98]. A streamlined exposition of a similar proof can be found in Dasgupta and Gupta [DG99]. For more general concentration results and techniques using the Gaussian distribution see e.g. [Pis89], [MS86].

Achlioptas [Ach01] proved that the components of the $b_i$ can also be chosen as independent uniform $\pm 1$ random variables. Here the distribution of $\langle b_i, u\rangle$ does depend on $u$ but the proof shows that for every $u \in S^{n-1}$, the concentration of $\|p(u)\|^2$ is at least as strong as in the case of the normally distributed $b_i$. This is established by analyzing higher moments of the distribution.

The best known upper bound on the dimension needed for a $(1+\varepsilon)$-embedding of an $n$-point Euclidean metric is $\frac{16}{\varepsilon^2}(1 + o(1)) \ln n$, where $o(1)$ is with respect to

$\varepsilon \to 0$ [IM98], [DG99], [Ach01]. The main term is optimal for the current proof method; see Exercises 3 and 15.3.4.

The Johnson–Lindenstrauss Flattening Lemma was applied in many algorithms, both in theory and practice; see the survey [Ind01] or, for example, Kleinberg [Kle97], Indyk and Motwani [IM98], Borodin, Ostrovsky, and Rabani [BOR99].

### Exercises

1. Let $x, y \in S^{n-1}$ be two points chosen independently and uniformly at random. Estimate their expected (Euclidean) distance, assuming that $n$ is large. $\boxed{3}$
2. Let $L \subseteq \mathbf{R}^n$ be a fixed $k$-dimensional linear subspace and let $x$ be a random point of $S^{n-1}$. Estimate the expected distance of $x$ from $L$, assuming that $n$ is large. $\boxed{3}$
3. (Lower bound for the Flattening Lemma)

   (a) Consider the $n{+}1$ points $0, e_1, e_2, \ldots, e_n \in \mathbf{R}^n$ (where the $e_i$ are the vectors of the standard orthonormal basis). Check that if these points with their Euclidean distances are $(1{+}\varepsilon)$-embemdded into $\ell_2^k$ then there exist unit vectors $v_1, v_2, \ldots, v_n \in \mathbf{R}^k$ with $|\langle v_i, v_j\rangle| \le 100\varepsilon$ for all $i \ne j$ (the constant can be improved). $\boxed{2}$

   (b) Let $A$ be an $n{\times}n$ symmetric real matrix with $a_{ii} = 1$ for all $i$ and $|a_{ij}| \le n^{-1/2}$ for all $j, j$, $i \ne j$. Prove that $A$ has rank at least $\frac{n}{2}$. $\boxed{4}$

   (c) Let $A$ be an $n{\times}n$ real matrix of rank $d$, let $k$ be a positive integer, and let $B$ be the $n{\times}n$ matrix with $b_{ij} = a_{ij}^k$. Prove that the rank of $B$ is at most $\binom{k+d}{k}$. $\boxed{3}$

   (d) Using (a)–(c), prove that if the set as in (a) is $(1{+}\varepsilon)$-embedded into $\ell_2^k$, where $100n^{-1/2} \le \varepsilon \le \frac{1}{2}$, then $k = \Omega(\frac{1}{\varepsilon^2 \log \frac{1}{\varepsilon}} \log n)$. $\boxed{3}$

   This proof is due to Alon (unpublished manuscript, Tel Aviv University).

## 15.3 Lower Bounds By Counting

In this section, we explain a construction providing many “essentially different” $n$-point metric spaces, and we derive a general lower bound on the minimum distortion required to embed all these spaces into a $d$-dimensional normed space. The key ingredient are constructions of graphs without short cycles.

**Graphs without short cycles.** The *girth* of a graph $G$ is the length of the shortest cycle in $G$. Let $m(\ell, n)$ denote the maximum possible number of edges of a simple graph on $n$ vertices containing no cycle of length $\ell$ or shorter, i.e. with girth at least $\ell{+}1$.

We have $m(2, n) = \binom{n}{2}$, since the complete graph $K_n$ has girth 3, and $m(3, n) = \lfloor \frac{n}{2} \rfloor \cdot \lceil \frac{n}{2} \rceil$ by Turán's theorem; the extremal example is the complete bipartite graph $K_{\lfloor n/2\rfloor, \lceil n/2\rceil}$. Another simple observation is that for all $k$, $m(2k{+}1, n) \ge \frac{1}{2} m(2k, n)$. This is because any graph $G$ has a bipartite subgraph $H$ that contains at least half of the edges of $G$.[2] So it suffices to care about

[2] To see this, divide the vertices of $G$ into two classes $A$ and $B$ arbitrarily, and while there is a vertex in one of the classes having more neighbors in its class than in the other class, move such vertex to the other class; the number of edges between $A$ and $B$ increases in each step. For another proof, assign each vertex randomly to $A$ or $B$ and check that the expected number of edges between $A$ and $B$ is $\frac{1}{2}|E(G)|$.

even cycles and to consider $\ell$ even, remembering that the bounds for $\ell = 2k$ and $\ell = 2k+1$ are almost the same up to a factor of 2.

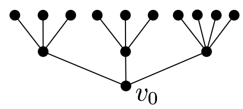
Here is a simple general upper bound on $m(\ell, n)$.

**15.3.1 Lemma.** *For all $n$ and $\ell$,*

$$m(\ell, n) \le n^{1+1/\lfloor \ell/2 \rfloor} + n.$$

**Proof.** It suffices to consider even $\ell = 2k$. Let $G$ be a graph with $n$ vertices and $m = m(2k, n)$ edges. The average degree is $\bar{d} = \frac{2m}{n}$. There is a subgraph $H \subseteq G$ with *minimum* degree at least $\delta = \frac{1}{2}\bar{d}$. Indeed, by deleting a vertex of degree smaller than $\delta$ the average degree does not decrease, and so $H$ can be obtained by a repeated deletion of such vertices.

Let $v_0$ be a vertex of $H$. The crucial observation is that since $H$ has no cycle of length $2k$ or shorter, the subgraph of $H$ induced by all vertices at distance at most $k$ from $v_0$ is a tree:




The number of vertices in this tree is at least $1+\delta+\delta(\delta-1)+\cdots+\delta(\delta-1)^{k-1} \ge (\delta-1)^k$, and this is no more than $n$. So $\delta \le n^{1/k}+1$ and $m = \frac{1}{2}\bar{d}n \le \delta n \le n^{1+1/k} + n$. $\square$

This simple argument remains essentially the best known upper bound. But it was asymptotically matched only for a few small values of $\ell$, namely for $\ell \in \{4, 5, 6, 7, 10, 11\}$. For $m(4, n)$ and $m(5, n)$, we need bipartite graphs without $K_{2,2}$; these were briefly discussed in Section 4.5, and we recall that they can have up to $n^{3/2}$ edges, as is witnessed by the finite projective plane. The remaining listed cases use clever algebraic constructions.

For the other $\ell$, the record is also held by algebraic constructions; they are not difficult to describe but proving that they work needs quite a deep mathematics. For all $\ell \equiv 1 \pmod 4$ (and not on the list above), they yield $m(\ell, n) = \Omega(n^{1+4/(3\ell-7)})$, while for $\ell \equiv 3 \pmod 4$, they lead to $m(\ell, n) = \Omega(n^{1+4/(3\ell-9)})$.

Here we prove a weaker but simple lower bound by the probabilistic method.

**15.3.2 Lemma.** *For all $\ell \ge 3$ and $n \ge 2$, we have*

$$m(\ell, n) \ge \tfrac{1}{9}\, n^{1+1/(\ell-1)}.$$

Of course, for odd $\ell$ we obtain an $\Omega(n^{1+1/(\ell-2)})$ bound by using the lemma for $\ell-1$.

**Proof.** First we note that we may assume $n \ge 4^{\ell-1} \ge 16$, for otherwise the bound in the lemma is verified by a path, say.

We consider the random graph $G(n, p)$ with $n$ vertices, where each of the $\binom{n}{2}$ possible edges is present with probability $p$, $0 < p < 1$, and these choices are mutually independent. The value of $p$ is going to be chosen later.

Let $E$ be the set of edges of $G(n,p)$ and let $F \subseteq E$ be the edges contained in cycles of length $\ell$ or shorter. By deleting all edges of $F$ from $G(n,p)$, we obtain a graph with no cycles of length $\ell$ or shorter. If we manage to show, for some $m$, that the expectation $\mathbf{E}[|E \setminus F|] \geq m$, then there is an instance of $G(n,p)$ with $|E \setminus F| \geq m$, and so there exists a graph with $n$ vertices, $m$ edges, and of girth greater than $\ell$.

We have $\mathbf{E}[|E|] = \binom{n}{2}p$. What is the probability that a fixed pair $e = \{u,v\}$ of vertices is an edge of $F$? First, $e$ must be an edge of $G(n,p)$, which has probability $p$, and second, there must be path of length between 2 and $\ell-1$ connecting $u$ and $v$. The probability that all the edges of a given potential path of length $k$ are present is $p^k$, and there are fewer than $n^{k-1}$ possible $u$-$v$ paths of length $k$. Therefore the probability of $e \in F$ is at most $\sum_{k=2}^{\ell-1} p^{k+1}n^{k-1}$, which can be bounded by $2p^\ell n^{\ell-2}$ provided that $np \geq 2$. Then $\mathbf{E}[|F|] \leq \binom{n}{2} \cdot 2p^\ell n^{\ell-2}$ and, by the linearity of expectation, we have

$$\mathbf{E}[|E \setminus F|] = \mathbf{E}[|E|] - \mathbf{E}[|F|] \geq \binom{n}{2}p\left(1 - 2p^{\ell-1}n^{\ell-2}\right).$$

Now we maximize this expression as a function of $p$; a somewhat rough but simple choice is $p = \frac{n^{1/(\ell-1)}}{2n}$, which leads to $\mathbf{E}[|E \setminus F|] \geq \frac{1}{9}n^{1+1/(\ell-1)}$ (the constant can be improved somewhat). The assumption $np \geq 2$ follows from $n \geq 4^{\ell-1}$. Lemma 15.3.2 is proved. □

There are several ways of proving a lower bound for $m(\ell,n)$ similar to that in Lemma 15.3.2, i.e. roughly $n^{1+1/\ell}$; one of the alternatives is indicated in Exercise 1 below. But obtaining a significantly better bound in an elementary way, or improving on the best known bounds (of roughly $n^{1+4/3\ell}$), remain challenging open problems.

We now use the knowledge about graphs without short cycles in lower bounds for distortion.

**15.3.3 Proposition (Distortion versus dimension).** *Let $Z$ be a $d$-dimensional normed space, such as some $\ell_p^d$, and suppose that all $n$-point metric spaces can be $D$-embedded into $Z$. Let $\ell$ be an integer with $D < \ell \leq 5D$ (it is essential that $\ell$ be strictly larger than $D$, while the upper bound is only for technical convenience). Then*

$$d \geq \frac{1}{\log_2 \frac{16D\ell}{\ell-D}} \cdot \frac{m(\ell,n)}{n}.$$

**Proof.** Let $G$ be a graph with vertex set $V = \{v_1, v_2, \ldots, v_n\}$ and with $m = m(\ell,n)$ edges. Let $\mathcal{G}$ denote the class of all subgraphs $H \subseteq G$ obtained from $G$ by deleting some edges (but retaining all vertices). For each $H \in \mathcal{G}$, we define a metric $\rho_H$ on the set $V$ by $\rho_H(u,v) = \min(\ell, d_H(u,v))$, where $d_H(u,v)$ is the length of a shortest path connecting $u$ and $v$ in $H$.

The idea of the proof is that $\mathcal{G}$ contains many essentially different metric spaces, and if the dimension of $Z$ were small then there would not be sufficiently many essentially different placements of $n$ points in $Z$.

Suppose that for every $H \in \mathcal{G}$ there exists a $D$-embedding $f_H\colon (V,\rho_H) \to Z$. By re-scaling, we make sure that $\frac{1}{D}\rho_H(u,v) \leq \|f_H(u) - f_H(v)\|_Z \leq \rho_H(u,v)$

for all $u, v \in V$. We may also assume that the images of all points are contained in the $\ell$-ball $B_Z(0,\ell) = \{x \in Z\colon \|x\|_Z \le \ell\}$.

Set $\beta = \frac{1}{4}(\frac{\ell}{D}-1)$. We have $0 < \beta \le 1$. Let $N$ be a $\beta$-net in $B_Z(0,\ell)$. The notion of $\beta$-net was defined above Lemma 13.1.1, and that lemma showed that a $\beta$-net in the $(d-1)$-dimensional Euclidean sphere has cardinality at most $(\frac{4}{\beta})^d$. Exactly the same volume argument proves that in our case $|N| \le (\frac{4\ell}{\beta})^d$.

For every $H \in \mathcal{G}$, we define a new mapping $g_H\colon V \to N$ by letting $g_H(v)$ be the nearest point to $f_H(v)$ in $N$ (ties resolved arbitrarily). We prove that for distinct $H_1, H_2 \in \mathcal{G}$, the mappings $g_{H_1}$ and $g_{H_2}$ are distinct.

The edge sets of $H_1$ and $H_2$ differ, so we can choose a pair $u$, $v$ of vertices that form an edge in one of them, say in $H_1$, and not in the other one ($H_2$). We have $\rho_{H_1}(u,v) = 1$ while $\rho_{H_2}(u,v) = \ell$, for otherwise an $u$–$v$ path in $H_2$ of length smaller than $\ell$ and the edge $\{u,v\}$ would induce a cycle of length at most $\ell$ in $G$. Thus

$$\|g_{H_1}(u) - g_{H_1}(v)\|_Z < \|f_{H_1}(u) - f_{H_1}(v)\|_Z + 2\beta \le 1 + 2\beta$$

and

$$\|g_{H_2}(u) - g_{H_2}(v)\|_Z > \|f_{H_2}(u) - f_{H_2}(v)\|_Z - 2\beta \ge \frac{\ell}{D} - 2\beta = 1 + 2\beta.$$

Therefore $g_{H_1}(u) \ne g_{H_2}(u)$ or $g_{H_1}(v) \ne g_{H_2}(v)$.

We have shown that there are at least $|\mathcal{G}|$ distinct mappings $V \to N$. The number of all mappings $V \to N$ is $|N|^n$, and so

$$|\mathcal{G}| = 2^m \le |N|^n \le \left(\frac{4\ell}{\beta}\right)^{nd}.$$

The bound in the proposition follows by calculation. $\square$

**15.3.4 Corollary ("Incompressibility" of general metric spaces).** *If $Z$ is a normed space such that all $n$ point metric spaces can be $D$-embedded into $Z$, where $D > 1$ is considered fixed and $n \to \infty$, then we have*

- $\dim Z = \Omega(n)$ *for* $D < 3$,
- $\dim Z = \Omega(\sqrt{n})$ *for* $D < 5$,
- $\dim Z = \Omega(n^{1/3})$ *for* $D < 7$.

This follows from Proposition 15.3.3 by substituting the asymptotically optimal bounds for $m(3,n)$, $m(5,n)$, and $m(7,n)$. The constant of proportionality in the first bound goes to 0 as $D \to 3$, and similarly for the other bounds.

The corollary shows that there is no normed space of dimension significantly smaller than $n$ in which one could represent all $n$-point metric spaces with distortion smaller than 3. So, for example, one cannot save much space by representing a general $n$-point metric space by the coordinates of points in some suitable normed space.

It is very surprising that, as we will see later, it *is* possible to 3-embed all $n$-point metric spaces into a particular normed space of dimension close to $\sqrt{n}$. So the value 3 for the distortion is a real threshold! Similar thresholds occur at the values 5 and 7. Most likely this continues for all odd integers $D$, but we cannot prove this because of the lack of tight bounds for the number of edges in graphs without short cycles.

Another consequence of Proposition 15.3.3 concerns embedding into Euclidean spaces, without any restriction on dimension.

**15.3.5 Proposition (Lower bound on embedding into Euclidean spaces).** *For all $n$, there exist $n$-point metric spaces that cannot be embedded into $\ell_2$ (i.e. into any Euclidean space) with distortion smaller than $c \log n / \log\log n$, where $c > 0$ is a suitable positive constant.*

**Proof.** If an $n$-point metric space is $D$-embedded into $\ell_2^n$ then, by the Johnson–Lindenstrauss Flattening Lemma 15.2.1, it can be $(2D)$-embedded into $\ell_2^d$ with $d \le C \log n$ for some specific constant $C$.

For contradiction, suppose that $D \le c_1 \log n / \log\log n$ with a sufficiently small $c_1 > 0$. Set $\ell = 4D$ and assume that $\ell$ is an integer. By Lemma 15.3.2, we have $m(\ell, n) \ge \frac{1}{9} n^{1+1/(\ell-1)} \ge C_1 n \log n$, where $C_1$ can be made as large as we wish by adjusting $c_1$. So Proposition 15.3.3 gives $d \ge \frac{C_1}{5} \log n$. If $C_1 > 5C$, we have a contradiction. $\square$

In the subsequent sections, the lower bound in Proposition 15.3.5 will be improved to $\Omega(\log n)$ by a completely different method, and then we will see that this latter bound is tight.

**Bibliography and Remarks.** The problem of constructing small graphs with given girth and minimum degree has a rich history; see e.g. Bollobás [Bol85] for most of the earlier results.

In the proof of Lemma 15.3.1, we have derived that any graph of minimum degree $\delta$ and girth $2k+1$ has at least $1 + \delta \sum_{i=0}^{k-1} (\delta-1)^i$ vertices, and a similar lower bound for girth $2k$ is $2 \sum_{i=0}^{k-1} (\delta-1)^i$. Graphs attaining these bounds (they are called *Moore graphs* for odd girth and *generalized polygon graphs* for even girth) are known to exist only in very few cases (see e.g. Biggs [Big93] for a nice exposition). Alon, Hoory, and Linial [AHL00] proved by a neat argument using random walks that the same formulas still bound the number of vertices from below if $\delta$ is the *average* degree (rather than minimum degree) of the graph. But none of this helps improve the bound on $m(\ell, n)$ by any substantial amount.

The proof of Lemma 15.3.2 is a variation on well-known proofs by Erdős.

The constructions mentioned in the text attaining the asymptotically optimal value of $m(\ell, n)$ for several small $\ell$ are due to Benson [Ben66] (constructions with similar properties appeared earlier in Tits [Tit59], where they were investigated for different reasons). As for the other $\ell$, graphs with the parameters given in the text were constructed by Lazebnik, Ustimenko, and Woldar [LUW95],[LUW96] by algebraic methods, improving on earlier bounds (such as those in Lubotzky, Phillips, Sarnak [LPS88]; also see the notes to Section 15.5).

Proposition 15.3.5 and the basic idea of Proposition 15.3.3 were invented by Bourgain [Bou85]. The explicit use of graphs without short cycles and the detection of the "thresholds" in the behavior of the dimension as a function of the distortion appeared in Matoušek [Mat96b].

Proposition 15.3.3 implies that a normed space that should accommodate *all* $n$-point metric spaces with a given small distortion must have large dimension. But what if we consider just one $n$-point metric space $M$, and we ask for the minimum dimension of a normed space $Z$ such that $M$ can be $D$-embedded into $Z$? Here $Z$ can be "customized" to $M$ and the counting argument as in the proof of Proposition 15.3.3 cannot work. By a nice different method, using the rank of certain matrices, Arias-de-Reyna and Rodríguez-Piazza [AR92] proved that for each $D < 2$, there are $n$-point metric spaces that do not $D$-embed into any normed space of dimension below $c(D)n$, for some $c(D) > 0$. In [Mat96b], their technique was extended and it was shown that for any $D > 1$, the required dimension is at least $c(\lfloor D \rfloor)n^{1/2\lfloor D \rfloor}$, so for a fixed $D$ it is at least a fixed power of $n$. The proof again uses graphs without short cycles. An interesting open problem is whether the possibility of selecting the norm in dependence on the metric can ever help substantially. For example, we know that if we want one normed space for all $n$-point metric spaces, then a linear dimension is needed for all distortions below 3. But the lower bounds in [AR92], [Mat96b] for a customized normed space only force linear dimension for distortion $D < 2$. Can every $n$-point metric space $M$ be 2.99-embedded, say, into some normed space $Z = Z(M)$ of dimension $o(n)$?

We have examined the tradeoff between dimension and distortion when the distortion is a fixed number. One may also ask for the minimum distortion if the dimension $d$ is fixed; this was considered in Matoušek [Mat90b]. For fixed $d$, all $\ell_p$-norms on $\mathbf{R}^d$ are equivalent up to a constant, and so it suffices to consider embeddings into $\ell_2^d$. Considering the $n$-point metric space with all distances equal to 1, a simple volume argument shows that an embedding into $\ell_2^d$ has distortion at least $\Omega(n^{1/d})$. The exponent can be improved by a factor of roughly 2; more precisely, for any $d \geq 1$, there exist $n$-point metric spaces requiring distortion $\Omega(n^{1/\lfloor (d+1)/2 \rfloor})$ for embedding into $\ell_2^d$ (these spaces are even isometrically embeddable into $\ell_2^{d+1}$). They are obtained by taking a $q$-dimensional simplicial complex that cannot be embedded into $\mathbf{R}^{2q}$ (a Van Kampen–Flores complex; for modern treatment see e.g. [Sar91] or [Živ97]), considering its geometric realization in $\mathbf{R}^{2q+1}$, and filling it with points uniformly (taking an $\eta$-net within it for a suitable $\eta$, in the metric sense); see Exercise 3 below for the case $q = 1$. For $d = 1$ and $d = 2$, this bound is asymptotically tight, as can be shown by an inductive argument [Mat90b]. It is also almost tight for all even $d$. An upper bound of $O(n^{2/d} \log^{3/2} n)$ for the distortion is obtained by first embedding the considered metric space into $\ell_2^n$ (Theorem 15.7.1) and then projecting on a random $d$-dimensional subspace; the analysis is similar to the proof of the Johnson–Lindenstrauss Flattening Lemma 15.2.1. It would be interesting to close the gap for odd $d \geq 3$; the case $d = 1$ suggests that perhaps the lower bound might be the truth. It is also rather puzzling that the (suspected) bound for the distortion for fixed dimension, $D \approx n^{1/\lfloor (d+1)/2 \rfloor}$, looks optically similar to the (suspected) bound for dimension given the distortion (Corollary 15.3.4), $d \approx n^{1/\lfloor (D+1)/2 \rfloor}$. Is this a pure coincidence, or is it trying to tell us something?

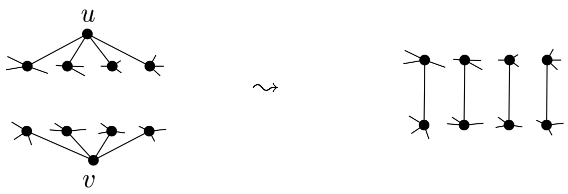
## Exercises

1. (Erdős–Sachs construction) This exercise indicates an elegant proof, by Erdős and Sachs [ES63], of the existence of graphs without short cycles whose number of edges is not much smaller than in Lemma 15.3.2 and that are *regular*. Let $\ell \geq 3$ and $\delta \geq 3$.

   (a) (Starting graph) For all $\delta$ and $\ell$, construct a finite $\delta$-regular graph $G(\delta, \ell)$ with no cycles of length $\ell$ or shorter; the number of vertices does not matter. One possibility is by double induction: construct $G(\delta+1, \ell)$ using $G(\delta, \ell)$ and

$G(\delta', \ell{-}1)$ with a suitable $\delta'$. $\boxed{4}$

(b) Let $G$ be a $\delta$-regular graph of girth at least $\ell{+}1$ and let $u$ and $v$ be two vertices of $G$ at distance at least $\ell{+}2$. Delete them together with their incident edges, and connect their neighbors by a matching:




Check that the resulting graph still does not contain any cycle of length at most $\ell$. $\boxed{2}$

(c) Show that starting with a graph as in (a) and reducing it by the operations as in (b), we arrive at a $\delta$-regular graph of girth $\ell{+}1$ and with at most $1 + \delta + \delta(\delta{-}1) + \cdots + \delta(\delta{-}1)^\ell$ vertices. What is the resulting asymptotic lower bound for $m(n, \ell)$, with $\ell$ fixed and $n \to \infty$? $\boxed{1}$

2. (Sparse spanners) Let $G$ be a graph with $n$ vertices and with positive real weights on edges which represent the edge lengths. A subgraph $H$ of $G$ is called a *$t$-spanner* of $G$ if the distance of any two vertices $u, v$ in $H$ is no more than $t$-times their distance in $G$ (both the distances are measured in the shortest-path metric). Using Lemma 15.3.1, prove that for every $G$ and every integer $t \geq 2$, there exists a $t$-spanner with $O(n^{1+1/\lfloor t/2 \rfloor})$ edges. $\boxed{4}$

3. Let $G_n$ denote the graph arising from $K_5$, the complete graph on 5 vertices, by subdividing each edge $(n{-}1)$ times, i.e. every two of the original vertices of $K_5$ are connected by a path of length $n$. Prove that the vertex set of $G_n$, considered as a metric space with the graph-theoretic distance, cannot be embedded into the plane with distortion smaller than $const \cdot n$. $\boxed{3}$

4. (Another lower bound for the Flattening Lemma)

   (a) Given $\varepsilon \in (0, \frac{1}{2})$ and $n$ sufficiently large in terms of $\varepsilon$, construct a collection $\mathcal{V}$ of ordered $n$-tuples of points of $\ell_2^n$ such that the distance of every two points in each $V \in \mathcal{V}$ is between two suitable constants, no two $V \neq V' \in \mathcal{V}$ can have the same $(1{+}\varepsilon)$-embedding (that is, there are $i, j$ such that the distances between the $i$th point and the $j$th point in $V$ and in $V'$ differ by a factor of at least $1{+}\varepsilon$), and $\log |\mathcal{V}| = \Omega(\varepsilon^{-2} n \log n)$. $\boxed{4}$

   (b) Use (a) and the method of this section to prove a lower bound of $\Omega(\frac{1}{\varepsilon^2 \log \frac{1}{\varepsilon}} \log n)$ for the dimension in the Johnson–Lindenstrauss Flattening Lemma 15.2.1. $\boxed{2}$

## 15.4 A Lower Bound for the Hamming Cube

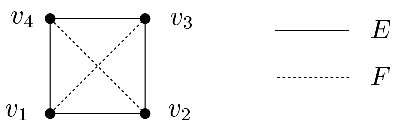
We have established the existence of $n$-point metric spaces requiring the distortion close to $\log n$ for embedding into $\ell_2$ (Proposition 15.3.5) but we have not constructed any specific metric space with this property. In this section, we prove a weaker lower bound, only $\Omega(\sqrt{\log n}\,)$, but for a specific and very simple space—the Hamming cube. Later on, we extend the proof method and exhibit metric spaces with $\Omega(\log n)$ lower bound, which turns out to be optimal. We recall that $C_m$ denotes the space $\{0,1\}^m$ with the Hamming (or $\ell_1$) metric, where the distance of two 0/1 sequences is the number of places where they differ.

**15.4.1 Theorem.** *Let $m \geq 2$ and $n = 2^m$. Then there is no $D$-embedding of the Hamming cube $C_m$ into $\ell_2$ with $D < \sqrt{m} = \sqrt{\log_2 n}$. That is, the natural embedding, where we regard $\{0,1\}^m$ as a subspace of $\ell_2^m$, is optimal.*

The reader may remember, perhaps with some dissatisfaction, that at the beginning of this chapter, we mentioned the 4-cycle as an example of a metric space which cannot be isometrically embedded into any Euclidean space, but we gave no reason. Now we are obliged to rectify this, because the 4-cycle is just the 2-dimensional Hamming cube.

The intuitive reason why the 4-cycle cannot be embedded isometrically is that if we embed the vertices so that the edges have the right length then at least one of the diagonals is too short. We make this precise using a slightly more complicated notation than necessary, in anticipation of later developments.

Let $V$ be a finite set, let $\rho$ be a metric on $V$, and let $E, F \subseteq \binom{V}{2}$ be nonempty sets of pairs of points of $V$. As our running example, $V = \{v_1, \ldots, v_4\}$ is the set of vertices of the 4-cycle, $\rho$ is the graph metric on it, $E = \{\{v_1, v_2\}, \{v_2, v_3\}, \{v_3, v_4\}, \{v_4, v_1\}\}$ are the edges, and $F = \{\{v_1, v_3\}, \{v_2, v_4\}\}$ are the diagonals.




Let us introduce the abbreviated notation

$$\rho^2(E) = \sum_{\{u,v\} \in E} \rho(u,v)^2.$$

We consider the ratio

$$R_{E,F}(\rho) = \sqrt{\frac{\rho^2(F)}{\rho^2(E)}};$$

the subscripts $E, F$ will be omitted unless there is danger of confusion. For our 4-cycle, $R(\rho)$ is a kind of ratio of "diagonals to edges" but with quadratic averages of distances, and it equals $\sqrt{2}$ (right?).

Next, let $f: V \to \ell_2^d$ be a $D$-embedding of the considered metric space into a Euclidean space. This defines another metric $\sigma$ on $V$: $\sigma(u,v) = \|f(u) - f(v)\|$. With the same $E$ and $F$, let us now look at the ratio $R(\sigma)$.

If $f$ is a $D$-embedding then $R(\sigma) \geq R(\rho)/D$. But, according to the idea mentioned above, in any embedding of the 4-cycle into a Euclidean space, the diagonals are always too short, and so $R(\sigma)$ can be expected to be smaller than $\sqrt{2}$ in this case. This is confirmed by the following lemma, which (with $x_i = f(v_i)$) shows that $R(\sigma) \leq 1$ and therefore $D \geq \sqrt{2}$.

**15.4.2 Lemma (Short diagonals).** *Let $x_1, x_2, x_3, x_4$ be arbitrary points in a Euclidean space. Then*

$$\|x_1 - x_3\|^2 + \|x_2 - x_4\|^2 \leq \|x_1 - x_2\|^2 + \|x_2 - x_3\|^2 + \|x_3 - x_4\|^2 + \|x_4 - x_1\|^2.$$

**Proof.** Four points can be assumed to lie in $\mathbf{R}^3$, so one could start some stereometric calculations. But a better way is to observe that it suffices to prove the lemma for points on the real line! Indeed, for the $x_i$ in some $\mathbf{R}^d$ we can write the 1-dimensional inequality for each coordinate and then add these inequalities together. (This is the reason for using squares in the definition of the ratio $R(\sigma)$: squares of Euclidean distances split into the contributions of individual coordinates and so they are easier to handle here than the distances themselves.)

If the $x_i$ are real numbers, we calculate

$$(x_1-x_2)^2+(x_2-x_3)^2+(x_3-x_4)^2+(x_4-x_1)^2-(x_1-x_3)^2-(x_2-x_4)^2$$

$$=(x_1-x_2+x_3-x_4)^2\geq 0,$$

and this is the desired inequality. □

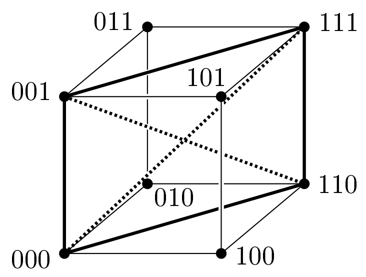
**Proof of Theorem 15.4.1.** We proceed similar to in the 2-dimensional case. Let $V=\{0,1\}^m$ be the vertex set of $C_m$, let $\rho$ be the Hamming metric, let $E$ be the set of edges of the cube (pairs of points at distance 1), and let $F$ be the set of the long diagonals. The long diagonals are pairs of points at distance $m$ or, in other words, pairs $\{u,\overline{u}\}$, $u\in V$, where $\overline{u}$ is the vector arising from $u$ by changing 0s to 1s and 1s to 0s.

We have $|E|=m2^{m-1}$ and $|F|=2^{m-1}$, and we calculate $R_{E,F}(\rho)=\sqrt{m}$. If $\sigma$ is a metric on $V$ induced by some embedding $f\colon V\to\ell_2^d$, we want to show $R_{E,F}(\sigma)\leq 1$; this will give the theorem. So we need to prove that $\sigma^2(F)\leq\sigma^2(E)$. This follows from the inequality for the 4-cycle (Lemma 15.4.2) by a convenient induction.

The basis for $m=2$ is directly Lemma 15.4.2. For larger $m$, divide the vertex set $V$ into two parts $V_0$ and $V_1$, where $V_0$ are the vectors with the last component 0, i.e. of the form $u0$, $u\in\{0,1\}^{m-1}$, and similarly for $V_1$. The set $V_0$ induces an $(m-1)$-dimensional subcube. Let $E_0$ be its edge set and $F_0$ the set of its long diagonals, i.e. $F_0=\{\{u0,\overline{u}0\}\colon u\in\{0,1\}^{m-1}\}$, and similarly for $E_1$ and $F_1$. Let $E_{01}=E\setminus(E_0\cup E_1)$ be the edges of the $m$-dimensional cube going between the two subcubes. By induction, we have

$$\sigma^2(F_0)\leq\sigma^2(E_0) \text{ and } \sigma^2(E_1)\leq\sigma^2(F_1).$$

For $u\in\{0,1\}^{m-1}$, consider the quadrilateral with vertices $u0$, $\overline{u}0$, $\overline{u}1$, $u1$; for $u=00$, it is indicated in the picture:



Its sides are two edges of $E_{01}$, one diagonal from $F_0$ and one from $F_1$, and its diagonals are from $F$. If we write the inequality of Lemma 15.4.2 for this quadrilateral and sum up over all such quadrilaterals (they are $2^{m-2}$, since $u$ and $\overline{u}$ yield the same quadrilaterals), we get

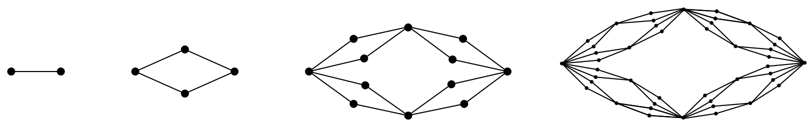
$$\sigma^2(F) \leq \sigma^2(E_{01}) + \sigma^2(F_0) + \sigma^2(F_1).$$

By the inductive assumption for the two subcubes, the right-hand side is at most $\sigma^2(E_{01}) + \sigma^2(E_0) + \sigma^2(E_1) = \sigma^2(E)$. □

**Bibliography and Remarks.** Theorem 15.4.1, found by Enflo [Enf69], is probably the first result showing an unbounded distortion for embeddings into Euclidean spaces. Enflo considered the problem of uniform embeddability among Banach spaces, and the distortion was an auxiliary device in his proof.

### Exercises

1. Consider the second graph in the introductory section, the star with 3 leaves, and prove a lower bound of $\frac{2}{\sqrt{3}}$ for the distortion required to embed into a Euclidean space. Follow the method used for the 4-cycle. $\boxed{3}$
2. (Planar graphs badly embeddable into $\ell_2$) Let $G_0, G_1, \ldots$ be the following graphs:



$G_0$ $G_1$ $G_2$ $G_3$

$G_{i+1}$ is obtained from $G_i$ by replacing each edge by a square with two new vertices. Using the Short Diagonals Lemma 15.4.2 and the method of this section, prove that any Euclidean embedding of $G_m$ (with the graph metric) requires distortion at least $\sqrt{m+1}$. $\boxed{4}$

This result is due to Newman and Rabinovich [NR01].
3. (Almost-Euclidean subspaces) Prove that for every $k$ and $\varepsilon > 0$ there exists $n = n(k, \varepsilon)$ such that every $n$-point metric space $(X, \rho)$ contains a $k$-point subspace that is $(1+\varepsilon)$-embeddable into $\ell_2$. Use Ramsey's Theorem. $\boxed{5}$

This result is due to Bourgain, Figiel, and Milman [BFM86]; it is a kind of analogue of Dvoretzky's Theorem for metric spaces.

## 15.5 A Tight Lower Bound via Expanders

Here we provide an explicit example of an $n$-point metric space that requires distortion $\Omega(\log n)$ for embedding into any Euclidean space. It is the vertex set of a constant-degree expander $G$ with the graph metric. In the proof, we are going to use bounds on the second eigenvalue of $G$, but for the readers not familiar with the important notion of expander graphs, we first include a little wider background.

Roughly speaking, expanders are graphs which are sparse but well connected. If a model of an expander is made with vertices being little balls and edges being thin strings, it is difficult to tear off any subset of vertices, and the more vertices we want to tear off the larger effort is needed.

More formally, we define the *edge expansion* (also called the *conductance*) $\Phi(G)$ of a graph $G = (V, E)$ as

$$\min\left\{\frac{e(A, V \setminus A)}{|A|}\colon A \subset V,\ 1 \le |A| \le \tfrac{1}{2}|V|\right\},$$

where $e(A, B)$ is the number of edges of $G$ going between $A$ and $B$. One can say, still somewhat imprecisely, that a graph $G$ is a good expander if $\Phi(G)$ is not very small compared to the average degree of $G$.

In this section, we consider $r$-regular graphs for a suitable constant $r \ge 3$, say $r = 3$. We need $r$-regular graphs with an arbitrary large number $n$ of vertices and with edge expansion bounded below by a positive constant independent of $n$. Such graphs are usually called *constant-degree expanders*.[3]

It is useful to note that, for example, the edge expansion of the $n\times n$ planar square grid tends to 0 as $n \to \infty$. More generally, it is known constant-degree expanders cannot be planar; they must be much more tangled than planar graphs.

The existence of constant-degree expanders is not difficult to prove by the probabilistic method; for every fixed $r \ge 3$, random $r$-regular graphs provide very good expanders. With considerable effort, explicit constructions have been found as well; see the notes to this section.

Let us remark that several notions similar to edge expansion appear in the literature, and each of them can be used for quantifying how good an expander a given graph is (but they usually lead to an equivalent notion of a family of constant-degree expanders). Often it is also useful to consider non-regular expanders or expanders with larger than constant degree, but regular constant-degree expanders are probably used most frequently.

Now we pass to the second eigenvalue. For our purposes it is most convenient to talk about eigenvalues of the Laplacian of the considered graph. Let $G = (V, E)$ be an $r$-regular graph. The *Laplacian matrix* $L_G$ of $G$ is an $n \times n$ matrix, $n = |V|$, with both rows and columns indexed by the vertices of $G$, defined by

$$(L_G)_{uv} = \begin{cases} r & \text{for } u = v \\ -1 & \text{if } u \ne v \text{ and } \{u, v\} \in E(G) \\ 0 & \text{otherwise.} \end{cases}$$

It is a symmetric positive semidefinite real matrix, and it has $n$ real eigenvalues $\mu_1 = 0 \le \mu_2 \le \cdots \le \mu_n$. The *second eigenvalue* $\mu_2 = \mu_2(G)$ is a fundamental parameter of the graph $G$.[4]

[3] A rigorous definition should be formulated for an infinite *family* of graphs. A family $\{G_1, G_2, \ldots\}$ of $r$-regular graphs with $|V(G_i)| \to \infty$ as $i \to \infty$ is a family of constant-degree expanders if the edge expansion of all $G_i$ is bounded below by a positive constant independent of $i$.

[4] The notation $\mu_i$ for the eigenvalues of $L_G$ is not standard. We use it in order to distinguish these eigenvalues from the eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$ of the *adjacency matrix* $A_G$ usually considered in the literature, where $(A_G)_{uv} = 1$ if $\{u, v\} \in E(G)$ and $(A_G)_{uv} = 0$ otherwise. Here we deal exclusively with regular graphs, for which the eigenvalues of $A_G$ are related to those of $L_G$ in a very simple way: $\lambda_i = r - \mu_i$, $i = 1, 2 \ldots, n$, for any $r$-regular graph.

Somewhat similar to edge expansion, $\mu_2(G)$ describes how much $G$ "holds together" but in a different way. The edge expansion and $\mu_2(G)$ are related but they do *not* determine each other. For every $r$-regular graph $G$, we have $\mu_2(G) \geq \frac{\Phi(G)^2}{4r}$ (see e.g. Lovász [Lov93], Ex. 11.31 for a proof) and $\mu_2(G) \leq 2\Phi(G)$ (Exercise 6). Both the lower and the upper bound can almost be attained for some graphs.

For our application below, we need the following fact: there are constants $r$ and $\beta > 0$ such that for sufficiently many values of $n$ (say for at least one $n$ between $10^k$ and $10^{k+1}$), there exists an $n$-vertex $r$-regular graph $G$ with $\mu_2(G) \geq \beta$. This follows from the existence results for constant-degree expanders mentioned above (random 3-regular graphs will do, for example), and actually most of the known explicit constructions of expanders bound the second eigenvalue directly.

We are going to use the lower bound on $\mu_2(G)$ via the following inequality:

For all real vectors $(x_v)_{v\in V}$ with $\sum_{v\in V} x_v = 0$, we have $x^T L_G x \geq \mu_2 \|x\|$. (15.3)

To understand what is going on here, we recall that every symmetric real $n \times n$ matrix has $n$ real eigenvalues (not necessarily distinct) and the corresponding $n$ unit eigenvectors $b_1, b_2, \ldots, b_n$ form an orthonormal basis of $\mathbf{R}^n$. For the matrix $L_G$, the unit eigenvector $b_1$ belonging to the eigenvalue $\mu_1 = 0$ is $n^{-1/2}(1, 1, \ldots, 1)$. So the condition $\sum_{v\in V} x_v = 0$ means the orthogonality of $x$ to $b_1$ and we have $x = \sum_{i=1}^n \alpha_i b_i$ for suitable real $\alpha_i$ with $\alpha_1 = 0$. We calculate, using $x^T b_i = \alpha_i$,

$$x^T L_G x = \sum_{i=2}^n x^T(\alpha_i L_G b_i) = \sum_{i=2}^n \alpha_i \mu_i x^T b_i = \sum_{i=2}^n \alpha_i^2 \mu_i \geq \mu_2 \sum_{i=2}^n \alpha_i^2 = \mu_2 \|x\|^2.$$

This proves (15.3), and we can also see that $x = b_2$ yields equality in (15.3). So we can write $\mu_2 = \min\{x^T L_G x\colon \|x\| = 1, \sum_{v\in V} x_v = 0\}$ (this is a special case of the variational definition of eigenvalues discussed in many textbooks of linear algebra).

Now we are ready to prove the main result of this section.

**15.5.1 Theorem (Expanders are badly embeddable into $\ell_2$).** *Let $G$ be an $r$-regular graph on an $n$-element vertex set $V$ with $\mu_2(G) \geq \beta$, where $r \geq 3$ and $\beta > 0$ are constants, and let $\rho$ be the shortest-path metric on $V$. Then the metric space $(V, \rho)$ cannot be $D$-embedded into a Euclidean space for $D \leq c \log n$, where $c = c(r, \beta) > 0$ is independent of $n$.*

**Proof.** We again consider the ratios $R_{E,F}(\rho)$ and $R_{E,F}(\sigma)$ as in the proof for the cube (Theorem 15.4.1). This time we let $E$ be the edge set of $G$, and $F = \binom{V}{2}$ are all pairs of distinct vertices. In the graph metric, all pairs in $E$ have distance 1, while most pairs in $F$ have distance about $\log n$, as we will check below. On the other hand, it turns out that in any embedding into $\ell_2$ such that all the distances in $E$ are at most 1, a typical distance in $F$ is only $O(1)$. The calculations follow.

We have $\rho^2(E) = |E| = \frac{nr}{2}$. To bound $\rho^2(F)$ from below, we observe that for each vertex $v_0$, there are at most $1+r+r(r-1)+\cdots+r(r-1)^{k-1} \le r^k+1$ vertices at distance at most $k$ from $v_0$. So for $k = \log_r \frac{n-1}{2}$, at least half of the pairs in $F$ have distance more than $k$, and we obtain $\rho^2(F) = \Omega(n^2k^2) = \Omega(n^2\log^2 n)$. Thus

$$R_{E,F}(\rho) = \Omega\left(\sqrt{n}\cdot\log n\right).$$

Let $f: V \to \ell_2^d$ be an embedding into a Euclidean space, and let $\sigma$ be the metric induced by it on $V$. To prove the theorem, it suffices to show that $R_{E,F}(\sigma) = O(\sqrt{n})$, i.e.

$$\sigma^2(F) = O(n\sigma^2(E)).$$

By the observation in the proof of Lemma 15.4.2 about splitting into coordinates, it is enough to prove this inequality for a one-dimensional embedding. So for every choice of real numbers $(x_v)_{v\in V}$, we want to show

$$\sum_{\{u,v\}\in F}(x_u-x_v)^2 = O(n)\sum_{\{u,v\}\in E}(x_u-x_v)^2. \tag{15.4}$$

By adding a suitable number to all the $x_v$, we may assume that $\sum_{v\in V} x_v = 0$. This does not change anything in (15.4) but it allows us to relate both sides to the Euclidean norm of the vector $x$.

We calculate, using $\sum_{v\in V} x_v = 0$,

$$\sum_{\{u,v\}\in F}(x_u-x_v)^2 = (n-1)\sum_{v\in V}x_v^2 - \sum_{u\neq v}x_ux_v = n\sum_{v\in V}x_v^2 - \Big(\sum_{v\in V}x_v\Big)^2 = n\|x\|^2.$$

For the right-hand side of (15.4), the Laplace matrix enters:

$$\sum_{\{u,v\}\in E}(x_u-x_v)^2 = r\sum_{v\in V}x_v^2 - 2\sum_{\{u,v\}\in E}x_ux_v = x^TL_Gx \ge \mu_2\|x\|^2,$$

the last inequality being (15.3). This establishes (15.4) and concludes the proof of Theorem 15.5.1. □

The proof actually shows that the maximum of $R_{E,F}(\sigma)$ over all Euclidean metrics $\sigma$ equals $\sqrt{\mu_2/n}$ (which an interesting geometric interpretation of $\mu_2$). The maximum is attained for the $\sigma$ induced by the mapping $V \to \mathbf{R}$ specified by $b_2$, the eigenvector belonging to $\mu_2$.

**The cone of squared $\ell_2$-metrics and universality of the lower-bound method.** For the Hamming cubes, we obtained the exact minimum distortion required for a Euclidean embedding. This was due to the lucky choice of the sets $E$ and $F$ of point pairs. As we will see below, a "lucky" choice, leading to an exact bound, exists for every finite metric space if we allow for sets of *weighted* pairs. Let $(V,\rho)$ be a finite metric space and let $\eta,\varphi: \binom{V}{2} \to [0,\infty)$ be weight functions. We define

$$\rho^2(\eta) = \sum_{\{u,v\}\in\binom{V}{2}}\eta(u,v)\rho(u,v)^2$$

and similarly for $\rho^2(\varphi)$, and we let

$$R_{\eta,\varphi}(\rho) = \sqrt{\frac{\rho^2(\varphi)}{\rho^2(\eta)}}.$$

**15.5.2 Proposition.** *Let $(V,\rho)$ be a finite metric space and let $D \geq 1$ be the smallest number such that $(V,\rho)$ can be $D$-embedded into $\ell_2$. Then there are weight functions $\eta, \varphi: \binom{V}{2} \to [0,\infty)$ such that $R_{\eta,\varphi}(\rho) \geq D$ and $R_{\eta,\varphi}(\sigma) \leq 1$ for any metric $\sigma$ induced on $V$ by an embedding into $\ell_2$.*

Thus, the exact lower bound for the embeddability into Euclidean spaces always has an "easy" proof, provided that we can guess the right weight functions $\eta$ and $\varphi$. (As we will see below, there is even an efficient algorithm for deciding $D$-embeddability into $\ell_2$.)

Proposition 15.5.2 is included mainly because of generally useful concepts appearing in its proof.

Let $V$ be a fixed $n$-point set. An arbitrary function $\varphi: \binom{V}{2} \to \mathbf{R}$, assigning a real number to each unordered pair of points of $V$, can be represented by a point in $\mathbf{R}^N$, where $N = \binom{n}{2}$; the coordinates of such a point are indexed by pairs $\{u,v\} \in \binom{V}{2}$. For example, the set of all metrics on $V$ corresponds to a subset of $\mathbf{R}^N$ called the *metric cone* (also see the notes to Section 5.5). As is not difficult to verify, it is an $N$-dimensional convex polyhedron in $\mathbf{R}^N$. Its combinatorial structure has been studied intensively.

In the proof of Proposition 15.5.2, we will not work with the metric cone but rather with the *cone of squared Euclidean metrics*, denoted by $\mathcal{L}_2$. We define

$$\mathcal{L}_2 = \left\{ (\|f(u) - f(v)\|^2)_{\{u,v\}\in\binom{V}{2}} : f: V \to \ell_2 \right\} \subset \mathbf{R}^N.$$

**15.5.3 Observation.** *The set $\mathcal{L}_2$ is a convex cone.*

**Proof.** Clearly if $x \in \mathcal{L}_2$ then $\lambda x \in \mathcal{L}_2$ for all $\lambda \geq 0$, and so it suffices to verify that if $x, y \in \mathcal{L}_2$ then $x + y \in \mathcal{L}_2$. Let $x, y \in \mathcal{L}_2$ correspond to embeddings $f: V \to \ell_2^k$ and $g: V \to \ell_2^m$, respectively. We define a new embedding $h: V \to \ell_2^{k+m}$ by concatenating the coordinates of $f$ and $g$; that is,

$$h(v) = (f(v)_1, \ldots, f(v)_k, g(v)_1, \ldots, g(v)_m) \in \ell_2^{k+m}.$$

The point of $\mathcal{L}_2$ corresponding to $h$ is $x + y$. $\square$

**Proof of Proposition 15.5.2.** Suppose that $(V,\rho)$ cannot be $D$-embedded into any Euclidean space. We are going to exhibit $\eta$ and $\varphi$ with $R_{\eta,\varphi}(\rho) \geq D$ and $R_{\eta,\varphi}(\sigma) \leq 1$ for every Euclidean $\sigma$. The claim of the proposition is easily derived from this by a compactness argument.

Let $\mathcal{L}_2 \subset \mathbf{R}^N$ be the cone of squared Euclidean metrics on $V$ as above and let

$$\mathcal{K} = \Big\{ (x_{uv})_{\{u,v\}\in\binom{V}{2}} \in \mathbf{R}^N : \text{ there exists an } r > 0 \text{ with } \\ r^2\rho(u,v)^2 \leq x_{uv} \leq D^2 r^2 \rho(u,v)^2 \text{ for all } u,v \Big\}.$$

This $\mathcal{K}$ includes all squares of metrics arising by $D$-embeddings of $(V,\rho)$, but not all elements of $\mathcal{K}$ are necessarily squares of metrics, as the triangle inequality may be violated. Since there is no Euclidean $D$-embedding of $(V,\rho)$, we have $\mathcal{K} \cap \mathcal{L}_2 = \emptyset$. Both $\mathcal{K}$ and $\mathcal{L}_2$ are convex sets in $\mathbf{R}^N$, and so they can be separated by a hyperplane, by the Separation Theorem 1.2.4. Moreover, since $\mathcal{L}_2$ is a cone and $\mathcal{K}$ is a cone minus the origin 0, the separating hyperplane has to pass through 0. So there is an $a \in \mathbf{R}^N$ such that

$$\langle a,x\rangle \geq 0 \text{ for all } x \in \mathcal{K} \text{ and } \langle a,x\rangle \leq 0 \text{ for all } x \in \mathcal{L}_2. \tag{15.5}$$

Using this $a$, we define the desired $\eta$ and $\varphi$, as follows:

$$\eta(u,v) = \begin{cases} a_{uv} & \text{if } a_{uv} \geq 0 \\ 0 & \text{otherwise;} \end{cases}$$

$$\varphi(u,v) = \begin{cases} -a_{uv} & \text{if } a_{uv} < 0 \\ 0 & \text{otherwise.} \end{cases}$$

First we show that $R_{\eta,\varphi}(\rho) \geq D$. To this end, we employ the property (15.5) for the following $x \in \mathcal{K}$:

$$x_{uv} = \begin{cases} D^2\rho(u,v)^2 & \text{if } a_{uv} \geq 0 \\ \rho(u,v)^2 & \text{if } a_{uv} < 0. \end{cases}$$

Then $\langle a,x\rangle \geq 0$ boils down to $D^2\rho^2(\eta) - \rho^2(\varphi) \geq 0$, which means that $R_{\eta,\varphi}(\rho) \geq D$.

Next, let $\sigma$ be a metric induced by a Euclidean embedding of $V$. This time we apply $\langle a,x\rangle \leq 0$ with the $x \in \mathcal{L}_2$ corresponding to $\sigma$, i.e. $x_{uv} = \sigma(u,v)^2$. This yields $\sigma^2(\eta) - \sigma^2(\varphi) \leq 0$, and so $R_{\eta,\varphi}(\sigma) \leq 1$. This proves Proposition 15.5.2. □

**Algorithmic remark: Euclidean embeddings and semidefinite programming.** The problem of deciding whether a given $n$-point metric space $(V,\rho)$ admits a $D$-embedding into $\ell_2$ (i.e. into a Euclidean space without restriction on the dimension), for a given $D \geq 1$, can be solved by a polynomial-time algorithm. Let us stress that the dimension of the target Euclidean space cannot be prescribed in this method. If we insist that the embedding be into $\ell_2^d$, for some given $d$, we obtain a different algorithmic problem, and it is not known how hard it is. Many other similar-looking embedding problems are known to be NP-hard, such as the problem of $D$-embedding into $\ell_1$.

The algorithm for $D$-embedding into $\ell_2$ is based on a powerful technique called *semidefinite programming*, where the problem is expressed as the existence of a positive semidefinite matrix in a suitable convex set of matrices.

Let $(V,\rho)$ be an $n$-point metric space, let $f\colon V \to \mathbf{R}^n$ be an embedding, and let $X$ be the $n \times n$ matrix whose columns are indexed by the elements of $V$ and such that the $v$th column is the vector $f(v) \in \mathbf{R}^n$. The matrix $Q = X^TX$ has both rows and columns indexed by the points of $V$ and the entry $q_{uv}$ is the scalar product $\langle f(u), f(v)\rangle$.

The matrix $Q$ is positive semidefinite, since for any $x \in \mathbf{R}^n$, we have $x^T Q x = (x^T X^T)(Xx) = \|Xx\|^2 \geq 0$. (In fact, as is not too difficult to check, a real symmetric $n \times n$ matrix $P$ is positive semidefinite if *and only if* it can be written as $X^T X$ for some real $n \times n$ matrix $X$.)

Let $\sigma(u,v) = \|f(u) - f(v)\| = \langle f(u) - f(v), f(u) - f(v)\rangle^{1/2}$. We can express

$$\sigma(u,v)^2 = \langle f(u), f(u)\rangle + \langle f(v), f(v)\rangle - 2\langle f(u), f(v)\rangle = q_{uu} + q_{vv} - 2q_{uv}.$$

Therefore, the space $(V, \rho)$ can be $D$-embedded into $\ell_2$ if and only if there exists a symmetric real positive semidefinite matrix $Q$ whose entries satisfy the following constraints:

$$\rho(u,v)^2 \leq q_{uu} + q_{vv} - 2q_{uv} \leq D^2 \rho(u,v)^2$$

for all $u, v \in V$. These are linear inequalities for the unknown entries of $Q$.

The problem of finding a positive semidefinite matrix whose entries satisfy a given system of linear inequalities can be solved efficiently, in time polynomial in the size of the unknown matrix $Q$ and in the number of the linear inequalities. The algorithm is not simple; we say a little more about it in the remarks below.

**Bibliography and Remarks.** Theorem 15.5.1 was proved by Linial, London, and Rabinovich [LLR95]. This influential paper introduced methods and results concerning low-distortion embeddings, developed in local theory of Banach spaces, into theoretical computer science, and it gave several new results and algorithmic applications. It is very interesting that using low-distortion Euclidean embeddings, one obtains algorithmic results for certain graph problems which so far could not be attained by other methods, although the considered problems look purely graph-theoretic without any geometric structure. A simple but important example is presented at the end of Section 15.7.

The bad embeddability of expanders was formulated and proved in [LLR95] in connection with the problem of multicommodity flows in graphs. The proof was similar to the one shown above, but it established $\Omega(\log n)$ bound for embedding into $\ell_1$. The result for Euclidean spaces is a corollary, since every finite Euclidean metric space can be isometrically embedded into $\ell_1$ (Exercise 5). An inequality similar to (15.4) was used, but with squares of differences replaced by absolute values of differences. Such an inequality was well-known for expanders. The method of [LLR95] was generalized for embeddings to $\ell_p$-spaces with arbitrary $p$ in [Mat97b]; it was shown that the minimum distortion required to embed all $n$-point metric spaces into $\ell_p$ is of the order $\frac{\log n}{p}$, and a matching upper bound was proved by the method shown in Section 15.7.

The proof of Theorem 15.5.1 given in the text can easily be extended to show a lower bound for $\ell_1$-embeddability as well. It actually shows that distortion $\Omega(\log n)$ is needed for approximating the expander metric by a squared Euclidean metric, and every $\ell_1$-metric is a squared Euclidean metric. Squared Euclidean metrics do not generally satisfy the triangle inequality but that is not needed in the proof. Those squared Euclidean metrics that do satisfy the triangle inequality are sometimes called the *metrics of negative type*. Not all of these metrics are $\ell_1$-metrics, but a challenging conjecture (made by Linial and independently by Goemans) states that perhaps they are not very far from $\ell_1$-metrics: each metric of negative type might be embeddable into $\ell_1$ with distortion bounded by a universal constant. If true, this would have significant algorithmic consequences: many problems can be formulated as optimization over the cone of all $\ell_1$-metrics, which is computationally intractable, and the metrics of negative type would provide a good and algorithmically manageable approximation.

The formulation of the minimum distortion problem for Euclidean embeddings as semidefinite programming is also due to [LLR95], as well as Proposition 15.5.2. These ideas were further elaborated and applied in examples by Linial and Magen [LM00]. The proof of Proposition 15.5.2 given in the text is simpler than that in [LLR95] and it extends to $\ell_p$-embeddability (Exercise 4), unlike the formulation of the $D$-embedding problem as a semidefinite program. It was communicated to me by Yuri Rabinovich.

A further significant progress in lower bounds for $\ell_2$-embeddings of graphs was made by Linial, Magen, and Naor [LMN01]. They proved that the metric of every $r$-regular graph, $r > 2$, of girth $g$ requires distortion at least $\Omega(\sqrt{g}\,)$ for embedding into $\ell_2$ (an $\Omega(g)$ lower bound was conjectured in [LLR95]). They give two proofs, one based on the concept of Markov type of a metric space due to Ball [Bal92] and another which we now outline (adapted to the notation of this section). Let $G = (V, E)$ be an $r$-regular graph of girth $2t{+}1$ or $2t{+}2$ for some integer $t \geq 1$, and let $\rho$ be the metric of $G$. We set $F = \{\{u, v\} \in \binom{V}{2} : \rho(u, v) = t\}$; note that the graph $H = (V, F)$ is $s$-regular for $s = r(r{-}1)^{t-1}$. Calculating $R_{E,F}(\rho)$ is trivial and it remains to bound $R_{E,F}(\sigma)$ for all Euclidean metrics $\sigma$ on $V$, which amounts to finding the largest $\beta > 0$ such that $\sigma^2(E) - \beta \cdot \sigma^2(F) \geq 0$ for all $\sigma$. Here it suffices to consider line metrics $\sigma$; so let $x_v \in \mathbf{R}$ be the image of $v$ in the embedding $V \to \mathbf{R}$ inducing $\sigma$. We may assume $\sum_{v \in V} x_v = 0$ and, as in the proof in the text, $\sigma^2(E) = \sum_{\{u,v\} \in E} (x_u - x_v)^2 = x^T L_G x = x^T (rI - A_G) x^T$, where $I$ is the identity matrix and $A_G$ is the adjacency matrix of $G$, and similarly for $\sigma^2(F)$. So we require $x^T C x \geq 0$ for all $x$ with $\sum_{v \in V} x_v = 0$, where $C = (r{-}\beta s)I - A_G + \beta A_H$. It turns out that there is a degree-$t$ polynomial $P_t(x)$ such that $A_H = P_t(A_G)$ (here we need that the girth of $G$ exceeds $2t$). This $P_t(x)$ is called the *Geronimus polynomial* and it is not hard to derive a recurrence for it: $P_0(x) = 1$, $P_1(x) = x$, $P_2(x) = x^2 - r$, and $P_t(x) = xP_{t-1}(x) - (r{-}1)P_{t-2}(x)$ for $t{>}2$. So $C = Q(A)$ for $Q(x) = r - \beta s - x + P_t(x)$. As is well-known, all the eigenvalues of $A$ lie in the interval $[-r, r]$, and so if we make sure that $Q(x) \geq 0$ for all $x \in [-r, r]$, all eigenvalues of $C$ are nonnegative and our condition holds. This leaves us with a nontrivial but doable calculus problem whose discussion we omit.

*Semidefinite programming.* The general problem of semidefinite programming is to optimize a linear function over a set of positive definite $n \times n$ matrices defined by a system of linear inequalities. This is a convex set in the space of all real $n \times n$ matrices, and in principle it is not difficult to construct a polynomial-time membership oracle for it (see the explanation following Theorem 13.2.1). Then the *ellipsoid method* can solve the optimization problem in polynomial time; see Grötschel, Lovász and Schrijver [GLS88]. More practical algorithms are based on interior point methods. Semidefinite programming is an extremely powerful tool in combinatorial optimization and other areas. For example, it provides the only known polynomial-time algorithms for computing the chromatic number of perfect graphs and the best known approximation algorithms for several fundamental NP-hard graph-theoretical problems. Lovász' recent lecture notes [Lov] are a beautiful concise introduction. Here we outline at least one lovely application, concerning the approximation of the maximum cut in a graph, in Exercise 7 below.

*The second eigenvalue.* The investigation of graph eigenvalues constitutes a well-established part of graph theory; see e.g. Biggs [Big93] for a nice introduction. The second eigenvalue of the Laplace matrix as an important graph parameter was first considered by Fiedler [Fie73] (who called it the *algebraic connectivity*). Tanner

[Tan84] and Alon and Milman [AM85] established a lower bound for the so-called vertex expansion of a regular graph (a notion similar to edge expansion) in terms of $\mu_2(G)$, and a reverse relation was proved by Alon [Alo86a].

There are many useful analogies of graph eigenvalues with the eigenvalues of the Laplace operator on manifolds, whose theory is classical and well-developed; this is pursued to a considerable depth in Chung [Chu97]. This point of view prefers the eigenvalues of the Laplacian matrix of a graph, as considered in this section, to the eigenvalues of the adjacency matrix. In fact, for non-regular graphs, a still closer correspondence with the setting of manifolds is obtained with a differently normalized Laplacian matrix $\mathcal{L}_G$: $(\mathcal{L}_G)_{v,v} = 1$ for all $v \in V(G)$, $(\mathcal{L}_G)_{uv} = -(\deg_G(u)\deg_G(v))^{-1/2}$ for $\{u,v\} \in E(G)$, and $(\mathcal{L}_G)_{uv} = 0$ otherwise.
*Expanders* have been used to address many fundamental problems in computer science, in areas such as network design, theory of computational complexity, coding theory, on-line computation, and cryptography; see e.g. [RVW00] for references.

For random graphs, parameters such as edge expansion or vertex expansion are usually not too hard to estimate (the technical difficulty of the arguments depends on the chosen model of a random graph). On the other hand, estimating the second eigenvalue of a random $r$-regular graph is quite challenging and a satisfactory answer is known only for $r$ large (and even); see Friedman, Komlós, and Szemerédi [FKS89] or Friedman [Fri91]. Namely, with high probability, a random $r$-regular graph with $r$ even has $\lambda_2 \leq 2\sqrt{r-1} + O(\log r)$. Here the number of vertices $n$ is assumed to be sufficiently large in terms of $r$ and the $O(\cdot)$ notation is with respect to $r \to \infty$. At the same time, for every fixed $r \geq 3$ and any $r$-regular graph on $n$ vertices, $\lambda_2 \geq 2\sqrt{r-1} - o(1)$, where this time $o(\cdot)$ refers to $n \to \infty$. So random graphs are almost optimal for large $r$.

For many of the applications of expanders, random graphs are not sufficient, and explicit constructions are required. In fact, explicitly constructed expanders often serve as substitutes for truly random graphs; for example, they allow one to convert some probabilistic algorithms into deterministic ones (derandomization) or reduce the number of random bits required by a probabilistic algorithm.

Explicit construction of expanders was a big challenge and it led to an excellent research employing surprisingly deep results from classical areas of mathematics (group theory, number theory, harmonic analysis, etc.). In the analysis of such constructions, one usually bounds the second eigenvalue (rather than edge expansion or vertex expansion). After the initial breakthrough by Margulis in 1973 and several other works in this direction (see e.g. [Mor94] or [RVW00] for references), explicit families of constant-degree expanders matching the quality of random graphs in several parameters (and even superseding them in some respects) were constructed by Lubotzky, Phillips, and Sarnak [LPS88] and independently by Margulis [Mar88]. Later Morgenstern [Mor94] obtained similar results for many more values of the parameters (degree and number of vertices). In particular, these constructions achieve $\lambda_2 \leq 2\sqrt{r-1}$, which is asymptotically optimal, as was mentioned earlier.

For illustration, here is one of the constructions (from [LPS88]). Let $p \neq q$ be primes with $p, q \equiv 1 \pmod 4$ and such that $p$ is a quadratic non-residue modulo $q$, let $i$ be an integer with $i^2 \equiv -1 \pmod q$, and let $F$ denote the field of residue classes modulo $q$. The vertex set $V(G)$ consists of all $2\times 2$ nonsingular matrices over $F$. Two matrices $A, B \in V(G)$ are connected by an edge iff $AB^{-1}$ is a matrix of the form $\begin{pmatrix} a_0+ia_1 & a_2+ia_3 \\ -a_2+ia_3 & a_0-ia_1 \end{pmatrix}$, where $a_0, a_1, a_2, a_3$ are integers with $a_0^2+a_1^2+a_2^2+a_3^2 = p$, $a_0 > 0$, $a_0$ odd, and $a_1, a_2, a_3$ even. By a theorem of Jacobi, there are exactly $p+1$ such vectors $(a_0, a_1, a_2, a_3)$ and it follows that the graph is $(p+1)$-regular with $q(q^2-1)$ vertices. A family of constant-degree expanders is obtained by fixing $p$, say $p = 5$, and letting $q \to \infty$.

Reingold, Vadhan, and Wigderson [RVW00] discovered an explicit construction of a different type: expanders are obtained from a constant-size initial graph by iterating certain product operations. Their parameters are somewhat inferior to those from [Mar88], [LPS88], [Mor94] but the proof is relatively short and it uses only elementary linear algebra.

## Exercises

1. Show that every real symmetric positive semidefinite $n \times n$ matrix can be written as $X^T X$ for a real $n \times n$ matrix $X$. $\boxed{3}$

2. (Dimension for isometric $\ell_p$-embeddings)

   (a) Let $V$ be an $n$-point set and let $N = \binom{n}{2}$. Analogous to the set $\mathcal{L}_2$ defined in the text, let $\mathcal{L}_1^{(fin)} \subset \mathbf{R}^N$ be the set of all metrics on $V$ induced by embeddings $f\colon V \to \ell_1^k$, $k = 1, 2, \ldots$. Show that $\mathcal{L}_1^{(fin)}$ is the convex hull of *line (pseudo)metrics* , i.e. (pseudo)metrics induced by mappings $f\colon V \to \ell_1^1$. $\boxed{2}$

   (b) Prove that any metric from $\mathcal{L}_1^{(fin)}$ can be isometrically embedded into $\ell_1^N$. That is, any $n$-point set in some $\ell_1^k$ can be realized in $\ell_1^N$. $\boxed{4}$ (Examples show that one cannot do much better and that dimension $\Omega(n^2)$ is necessary, unlike for Euclidean embeddings where dimension $n-1$ always suffices.)

   (c) Let $\mathcal{L}_1 \subset \mathbf{R}^N$ be all metrics induced by embeddings of $V$ into $\ell_1$ (the space of infinite sequences with finite $\ell_1$-norm). Show that $\mathcal{L}_1 = \mathcal{L}_1^{(fin)}$, and thus that any $n$-point subset of $\ell_1$ can be realized in $\ell_1^N$. $\boxed{3}$

   (d) Extend the considerations in (a)–(c) to $\ell_p$-metrics with arbitrary $p \in [1, \infty)$. $\boxed{3}$

   See Ball [Bal90] for more on the dimension of isometric $\ell_p$-embeddings.

3. With the notation as in Exercise 2, show that every line pseudometric $\nu$ on an $n$-point set $V$ is a nonnegative linear combination of at most $n-1$ *cut pseudometrics*: $\nu = \sum_{i=1}^{n-1} \alpha_i \tau_i$, $\alpha_1, \ldots, \alpha_{n-1} \geq 0$, each $\tau_i$ being a cut pseudometric, i.e. a line pseudometric induced by a mapping $\psi_i\colon V \to \{0, 1\}$. (Consequently, by Exercise 2(a), every finite metric isometrically embeddable into $\ell_1$ is a nonnegative linear combination of cut pseudometrics.) $\boxed{3}$

4. (An $\ell_p$-analogue of Proposition 15.5.2) Let $p \in [1, \infty)$ be fixed. Using Exercise 2, formulate and prove an appropriate $\ell_p$-analogue of Proposition 15.5.2. $\boxed{3}$

5. (Finite $\ell_2$-metrics embed isometrically into $\ell_p$)

   (a) Let $p$ be fixed. Check that if for all $\varepsilon > 0$, a finite metric space $(V, \rho)$ can be $(1+\varepsilon)$-embedded into some $\ell_p^k$, $k = k(\varepsilon)$, then $(V, \rho)$ can be isometrically embedded into $\ell_p^N$, where $N = \binom{|V|}{2}$. Use Exercise 2. $\boxed{2}$

   (b) Prove that every $n$-point set in $\ell_2$ can be isometrically embedded into $\ell_p^N$. $\boxed{2}$

6. (The second eigenvalue and edge expansion) Let $G$ be an $r$-regular graph with $n$ vertices, and let $A, B \subseteq V$ be disjoint. Prove that the number of edges connecting $A$ to $B$ is at least $e(A, B) \geq \mu_2(G) \cdot \frac{|A| \cdot |B|}{n}$ (use (15.3) with a suitable vector $x$), and deduce that $\Phi(G) \geq \frac{1}{2} \mu_2(G)$. $\boxed{4}$

7. (The Goemans–Williamson approximation to MAXCUT) Let $G = (V, E)$ be a given graph and let $n = |V|$. The MAXCUT problem for $G$ is to find the maximum possible number of "crossing" edges for a partition $V = A \dot\cup B$ of the vertex set into two disjoint subsets, i.e. $\max_{A \subseteq V} e(A, V \setminus A)$. This is an

NP-complete problem. The exercise outlines a geometric randomized algorithm which finds an approximate solution using semidefinite programming.

(a) Check that the MAXCUT problem is equivalent to computing

$$M_{opt} = \max\Big\{ \tfrac{1}{2} \sum_{\{u,v\}\in E} (1 - x_u x_v)\colon x_v \in \{-1,1\},\, v \in V \Big\}.$$

$\boxed{2}$

(b) Let

$$M_{relax} = \max\Big\{ \tfrac{1}{2} \sum_{\{u,v\}\in E} (1 - \langle y_u, y_v\rangle)\colon y_v \in \mathbf{R}^n, \|y_v\| = 1,\, v \in V \Big\}.$$

Clearly $M_{relax} \geq M_{opt}$. Verify that this relaxed version of the problem is an instance of a semidefinite program, i.e. the maximum of a linear function over the intersection of a polytope with the cone of all symmetric positive semidefinite real matrices. $\boxed{2}$

(c) Let $(y_v\colon v \in V)$ be some system of unit vectors in $\mathbf{R}^n$ for which $M_{relax}$ is attained. Let $r \in \mathbf{R}^n$ be a random unit vector, and set $x_v = \operatorname{sgn}\langle y_v, r\rangle$, $v \in V$. Let $M_{approx} = \frac{1}{2}\sum_{\{u,v\}\in E}(1 - x_u x_v)$ for these $x_v$. Show that the expectation, with respect to the random choice of $r$, of $M_{approx}$ is at least $0.878 \cdot M_{relax}$ (consider the expected contribution of each edge separately). So we obtain a polynomial-time randomized algorithm producing a solution to MAXCUT whose expected value is at least about 88% of the optimal solution. $\boxed{4}$

*Remark.* This algorithm is due to Goemans and Williamson [GW95]. Later Håstad [Hås97] proved that no polynomial-time algorithm can produce better approximation in the worst case than about 94% unless $P = NP$ (also see Feige and Schechtman [FS01] for nice mathematics showing that the Goemans–Williamson value 0.878... is, in a certain sense, optimal for approaches based on semidefinite programming).

## 15.6 Upper Bounds for $\ell_\infty$-Embeddings

In this section, we explain a technique for producing low-distortion embeddings of finite metric spaces. Although we are mainly interested in Euclidean embeddings, here we begin with embeddings into the space $\ell_\infty$, which are somewhat simpler. We derive almost tight upper bounds.

Let $(V,\rho)$ be an arbitrary metric space. To specify an embedding

$$f\colon (V,\rho) \to \ell_\infty^d$$

means to define $d$ functions $f_1,\ldots,f_d\colon V \to \mathbf{R}$, the coordinates of the embedded points. If we aim at a $D$-embedding, without loss of generality we may require it to be nonexpanding, which means $|f_i(u) - f_i(v)| \leq \rho(u,v)$ for all $u,v \in V$ and all $i = 1,2,\ldots,d$. The $D$-embedding condition then means that for every pair $\{u,v\}$ of points of $V$, there is a coordinate $i = i(u,v)$ which "takes care" of the pair: $|f_i(u) - f_i(v)| \geq \frac{1}{D}\rho(u,v)$.

One of the key tricks in constructions of such embeddings is to take each $f_i$ as the distance to some suitable subset $A_i \subseteq V$; that is, $f_i(u) = \rho(u, A_i) =$

$\max_{a\in A_i} \rho(u,a)$. By the triangle inequality, we have $|\rho(u,A_i) - \rho(v,A_i)| \leq \rho(u,v)$ for any $u,v \in V$, and so such an embedding is automatically nonexpanding. We "only" have to choose a suitable collection of the $A_i$ that take care of all pairs $\{u,v\}$.

We begin with a simple case: an old observation showing that every finite metric space embeds isometrically into $\ell_\infty$.

**15.6.1 Proposition (Fréchet's embedding).** *Let $(V,\rho)$ be an arbitrary $n$-point metric space. Then there is an isometric embedding $f\colon V \to \ell_\infty^n$.*

**Proof.** Here the coordinates in $\ell_\infty^n$ are indexed by the points of $V$, and the $v$th coordinate is given by $f_v(u) = \rho(u,v)$. In the notation above, we thus put $A_v = \{v\}$. As we have seen, the embedding is nonexpanding by the triangle inequality. On the other hand, the coordinate $v$ takes care of the pairs $\{u,v\}$ for all $u \in V$:

$$\|f(u) - f(v)\|_\infty \geq |f_v(u) - f_v(v)| = \rho(u,v).$$

□

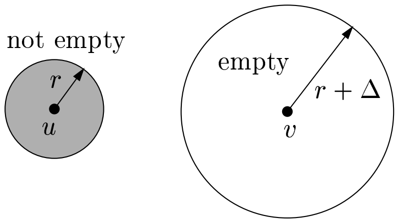
The dimension of the image in this embedding can be reduced a little; for example, we can choose some $v_0 \in V$ and remove the coordinate corresponding to $v_0$, and the above proof still works. To reduce the dimension significantly, though, we have to pay the price of distortion: for example, from Corollary 15.3.4 we know that for distortions below 3, the dimension must generally remain at least a fixed fraction of $n$. We prove an upper bound on the dimension needed for embeddings with a given distortion, which nearly matches the lower bounds in Corollary 15.3.4:

**15.6.2 Theorem.** *Let $D = 2q-1 \geq 3$ be an odd integer and let $(V,\rho)$ be an $n$-point metric space. Then there is a $D$-embedding of $V$ into $\ell_\infty^d$ with*

$$d = O(qn^{1/q} \ln n).$$

**Proof.** The basic scheme of the construction is as explained above: each coordinate is given by the distance to a suitable subset of $V$. This time the subsets are chosen at random with suitable densities.

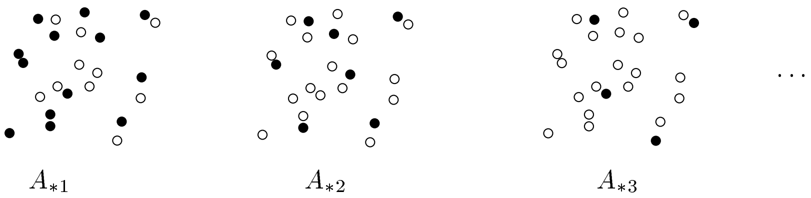
Let us consider two points $u,v \in V$. What are the sets $A$ such that $|\rho(u,A) - \rho(v,A)| \geq \Delta$, for a given real $\Delta > 0$? For some $r \geq 0$, they must intersect the $r$-ball around $u$ and avoid the $(r+\Delta)$-ball around $v$; schematically



or conversely (with the role of $u$ and $v$ interchanged).

In the favorable situation where the $r$-ball around $u$ does not contain much fewer points of $V$ than the $(r+\Delta)$-ball around $v$, a random $A$ with a suitable density has a reasonable chance to work. Generally we have no control over the distribution of points around $u$ and around $v$, but by considering several suitable balls simultaneously we can find a good pair of balls. We also do not know the right density needed for the sample to work, but since we have many coordinates we can take samples of essentially all possible densities.

Now we begin with the formal proof. We define an auxiliary parameter $p = n^{-1/q}$, and for $j = 1, 2, \ldots, q$, we introduce the probabilities $p_j = \min(\frac{1}{2}, p^j)$. Further let $m = \lceil 24n^{1/q} \ln n \rceil$. For $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, q$, we choose a random subset $A_{ij} \subseteq V$. The sets (and the corresponding coordinates in $\ell_\infty^{mq}$) now have double indices, and the index $j$ influences the "density" of $A_{ij}$. Namely, each point $v \in V$ has probability $p_j$ of being included into $A_{ij}$, and these events are mutually independent. The choices of the $A_{ij}$, too, are independent for distinct indices $i$ and $j$. Here is a schematic illustration of the sampling:




We divide the coordinates in $\ell_\infty^d$ into $q$ blocks by $m$ coordinates. For $v \in V$, we let

$$f(v)_{ij} = \rho(v, A_{ij}), \quad i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, q.$$

We claim that with a positive probability, this $f\colon V \to \ell_\infty^{mq}$ is a $D$-embedding. We have already noted that $f$ is nonexpanding, and the following lemma serves for showing that, with a positive probability, every pair $\{u, v\}$ is taken care of.

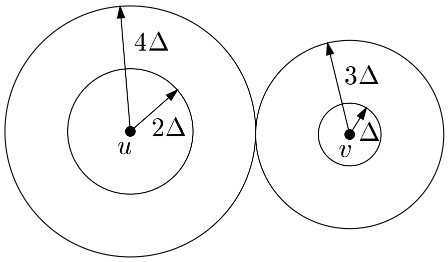
**15.6.3 Lemma.** *Let $u, v$ be two distinct points of $V$. Then there exists an index $j \in \{1, 2, \ldots, q\}$ such that if the set $A_{ij}$ is chosen randomly as above, then the probability of the event*

$$|\rho(u, A_{ij}) - \rho(v, A_{ij})| \geq \tfrac{1}{D}\, \rho(u, v) \tag{15.6}$$

*is at least $\frac{p}{12}$.*

First, assuming this lemma, we finish the proof of the theorem. To show that $f$ is a $D$-embedding, it suffices to show that, with a nonzero probability, for every pair $\{u, v\}$ there are $i, j$ such that the event (15.6) in the lemma occurs for the set $A_{ij}$. Consider a fixed pair $\{u, v\}$ and select the appropriate index $j$ as in the lemma. The probability that the event (15.6) does not occur for any of the $m$ indices $i$ is at most $(1 - \frac{p}{12})^m \leq e^{-pm/12} \leq n^{-2}$. Since there are $\binom{n}{2} < n^2$ pairs $\{u, v\}$, the probability that we fail to choose a good set for any of the pairs is smaller than 1. $\square$

**Proof of Lemma 15.6.3.** Set $\Delta = \frac{1}{D}\,\rho(u,v)$. Let $B_0 = \{u\}$, let $B_1$ be the (closed) $\Delta$-ball around $v$, let $B_2$ be the (closed) $2\Delta$-ball around $u$,..., finishing with $B_q$ which is a $q\Delta$-ball around $u$ (if $q$ is even) or around $v$ (if $q$ is odd). The parameters are chosen so that the radii of $B_{q-1}$ and $B_q$ add up to $\rho(u,v)$, i.e. the last two balls just touch (recall that $D = 2q-1$):




Let $n_t$ denote number of points of $V$ in $B_t$.

We want to select an index $j$ such that

$$n_t \geq n^{(j-1)/q} \quad \text{and} \quad n_{t+1} \leq n^{j/q}. \tag{15.7}$$

To this end, we divide the interval $[1,n]$ into $q$ intervals $I_1, I_2, \ldots, I_q$, where

$$I_j = \left[n^{(j-1)/q}, n^{j/q}\right].$$

If the sequence $(n_1, n_2, \ldots, n_q)$ is not monotone increasing, i.e. if $n_{t+1} < n_t$ for some $t$, then (15.7) holds for the $j$ such that $I_j$ contains $n_t$. On the other hand, if $1 = n_0 \leq n_1 \leq \ldots \leq n_q \leq n$, then by the pigeonhole principle, there exist $t$ and $j$ such that the interval $I_j$ contains both $n_t$ and $n_{t+1}$. Then (15.7) holds for this $j$ as well.

In this way, we have selected the index $j$ whose existence is claimed in the lemma. We will show that with probability at least $\frac{p}{12}$, the set $A_{ij}$, randomly selected with point probability $p_j$, includes a point of $B_t$ (event $E_1$) and is disjoint from the interior of $B_{t+1}$ (event $E_2$); such an $A_{ij}$ satisfies (15.6). Since $B_t$ and the interior of $B_{t+1}$ are disjoint, the events $E_1$ and $E_2$ are independent.

We calculate

$$\mathrm{Prob}[E_1] = 1 - \mathrm{Prob}[A_{ij} \cap B_t = \emptyset] = 1 - (1-p_j)^{n_t} \geq 1 - e^{-p_j n_t}.$$

Using (15.7), we have $p_j n_t \geq p_j n^{(j-1)/q} = p_j p^{-j+1} = \min(\frac{1}{2}, p^j)p^{-j+1} \geq \min(\frac{1}{2}, p)$. For $p \geq \frac{1}{2}$, we get $\mathrm{Prob}[E_1] \geq 1 - e^{-1/2} > \frac{1}{3} \geq \frac{p}{3}$, while for $p < \frac{1}{2}$, we have $\mathrm{Prob}[E_1] \geq 1 - e^{-p}$, and a bit of calculus verifies that the last expression is well above $\frac{p}{3}$ for all $p \in [0, \frac{1}{2})$.

Further

$$\mathrm{Prob}[E_2] \geq (1-p_j)^{n_{t+1}} \geq (1-p_j)^{n^{j/q}} \geq (1-p_j)^{1/p_j} \geq \tfrac{1}{4}$$

(since $p_j \leq \frac{1}{2}$). Thus $\mathrm{Prob}[E_1 \cap E_2] \geq \frac{p}{12}$, which proves the lemma. □

**Bibliography and Remarks.** The embedding method discussed in this section was found by Bourgain [Bou85], who used it to prove Theorem 15.7.1 explained in the subsequent section. Theorem 15.6.2 is from [Mat96b].

**Exercises**

1. (a) Find an isometric embedding of $\ell_1^d$ into $\ell_\infty^{2^d}$. $\boxed{3}$

   (b) Explain how an embedding as in (a) can be used to compute the diameter of an $n$-point set in $\ell_1^d$ in time $O(d2^d n)$. $\boxed{3}$

2. Show that if the unit ball $K$ of some finite-dimensional normed space is a convex polytope with $2m$ facets then that normed space embeds isometrically into $\ell_\infty^m$. $\boxed{2}$

   (Using results on approximation of convex bodies by polytopes, this yields useful approximate embeddings of arbitrary norms into $\ell_\infty^k$.)

3. Deduce from Theorem 15.6.2 that every $n$-point metric space can be $D$-embedded into $\ell_2^k$ with $D = O(\log^2 n)$ and $k = O(\log^2 n)$. $\boxed{2}$

## 15.7 Upper Bounds for Euclidean Embeddings

By a method similar to the one shown in the previous section, one can also prove a tight upper bound on Euclidean embeddings; the method was actually invented for this problem.

**15.7.1 Theorem (Bourgain's embedding into $\ell_2$).** *Every $n$-point metric space $(V, \rho)$ can be embedded into a Euclidean space with distortion at most $O(\log n)$.*

The overall strategy of the embedding is similar to the embedding into $\ell_\infty^d$ in the proof of Theorem 15.6.2: the coordinates in $\ell_2^d$ are given by distances to suitable subsets. The situation is slightly more complicated than before: for embedding into $\ell_\infty^d$, it was enough to exhibit one coordinate "taking care" of each pair, whereas for the Euclidean embedding, many of the coordinates will contribute significantly to every pair. Here is the appropriate analogue of Lemma 15.6.3.

**15.7.2 Lemma.** *Let $u, v \in V$ be two distinct points. Then there exist real numbers $\Delta_1, \Delta_2, \ldots, \Delta_q \geq 0$ with $\Delta_1 + \cdots + \Delta_q = \frac{1}{4}\rho(u,v)$, where $q = \lfloor \log_2 n \rfloor + 1$, and such that the following holds for each $j = 1, 2, \ldots, q$: if $A_j \subseteq V$ is a randomly chosen subset of $V$, with each point of $V$ included in $A_j$ independently with probability $2^{-j}$, then the probability $P_j$ of the event*

$$|\rho(u, A_j) - \rho(v, A_j)| \geq \Delta_j$$

*satisfies $P_j \geq \frac{1}{12}$.*

**Proof.** We fix $u$ and $v$. We define $r_q = \frac{1}{4}\rho(u,v)$ and for $j = 0, 1, \ldots, q-1$, we let $r_j$ be the smallest radius such that both $|B(u, r_j)| \geq 2^j$ and $|B(v, r_j)| \geq 2^j$ where, as usual, $B(x, r) = \{y \in V : \rho(x,y) \leq r\}$. We are going to show that the claim of the lemma holds with $\Delta_j = r_j - r_{j-1}$.

Fix $j \in \{1, 2, \ldots, q\}$ and let $A_j \subseteq V$ be a random sample with point probability $2^{-j}$. By the definition of $r_j$, $|B^\circ(u, r_j)| < 2^j$ or $|B^\circ(v, r_j)| < 2^j$, where $B^\circ(x, r) = \{y \in V : \rho(x,y) < r\}$ denotes the open ball (this holds for $j = q$ too because $|V| \leq 2^q$). Choose the notation $u, v$ so that $|B^\circ(u, r_j)| < 2^j$. A random set $A_j$ is good if it intersects $B(v, r_{j-1})$ and misses $B^\circ(u, r_j)$. The former set has cardinality at least $2^{j-1}$ and the latter at most $2^j$. The calculation of the probability that $A_j$ has these properties is identical to the calculation in the proof of Lemma 15.6.3 with $p = \frac{1}{2}$. □

In the subsequent proof of Theorem 15.7.1, we will construct the embedding in a slightly roundabout way, which sheds some light on what is really going on. Define a *line pseudometric* on $V$ to be any pseudometric[5] $\nu$ induced by a mapping $\varphi: V \to \mathbf{R}$, i.e. given by $\nu(u, v) = |\varphi(u) - \varphi(v)|$. For each $A \subseteq V$, let $\nu_A$ be the line pseudometric corresponding to the mapping $v \mapsto \rho(v, A)$. As we have noted, each $\nu_A$ is *dominated* by $\rho$, i.e. $\nu_A \leq \rho$ (the inequality between two (pseudo)metrics on the same point set means inequality for each pair of points).

The following easy lemma shows that if a metric $\rho$ on $V$ can be approximated by a convex combination of line pseudometrics, each of them dominated by $\rho$, then a good embedding of $(V, \rho)$ into $\ell_2$ exists.

**15.7.3 Lemma.** *Let $(V, \rho)$ be a finite metric space, and let $\nu_1, \ldots, \nu_N$ be line pseudometrics on $V$ with $\nu_i \leq \rho$ for all $i$ and such that*

$$\sum_{i=1}^{N} \alpha_i \nu_i \geq \frac{1}{D} \rho$$

*for some nonnegative $\alpha_1, \ldots, \alpha_N$ summing up to 1. Then $(V, \rho)$ can be $D$-embedded into $\ell_2^N$.*

**Proof.** Let $\varphi_i: V \to \mathbf{R}$ be a mapping inducing the line pseudometric $\nu_i$. We define the embedding $f: V \to \ell_2^N$ by

$$f(v)_i = \sqrt{\alpha_i} \cdot \varphi_i(v).$$

Then, on the one hand,

$$\|f(u) - f(v)\|^2 = \sum_{i=1}^{N} \alpha_i \nu_i(u, v)^2 \leq \rho(u, v)^2$$

because all $\nu_i$ are dominated by $\rho$ and $\sum \alpha_i = 1$. On the other hand,

$$\begin{aligned} \|f(u) - f(v)\| &= \Big( \sum_{i=1}^{N} \alpha_i \nu_i(u, v)^2 \Big)^{1/2} = \Big( \sum_{i=1}^{N} \alpha_i \Big)^{1/2} \Big( \sum_{i=1}^{N} \alpha_i \nu_i(u, v)^2 \Big)^{1/2} \\ &\geq \sum_{i=1}^{N} \alpha_i \nu_i(u, v) \end{aligned}$$

by Cauchy–Schwarz, and the latter expression is at least $\frac{1}{D} \rho(u, v)$ by the assumption. □

[5] A pseudometric $\nu$ satisfies all the axioms of a metric except that we may have $\nu(x, y) = 0$ even for two distinct points $x$ and $y$.

**Proof of Theorem 15.7.1.** As was remarked above, each of the line pseudometrics $\nu_A$ corresponding to the mapping $v \mapsto \rho(v, A)$ is dominated by $\rho$. It remains to observe that Lemma 15.7.2 provides a convex combination of these line pseudometrics that is bounded from below by $\frac{1}{48q} \cdot \rho$. The coefficient of each $\nu_A$ in this convex combination is given by the probability of $A$ appearing as one of the sets $A_j$ in Lemma 15.7.2. More precisely, write $\pi_j(A)$ for the probability that a random subset of $V$, with points picked independently with probability $2^{-j}$, equals $A$. Then the claim of Lemma 15.7.2 implies, for every pair $\{u, v\}$,

$$\sum_{A \subseteq V} \pi_j(A) \cdot \nu_A(u, v) \geq \tfrac{1}{12} \Delta_j.$$

Summing up over $j = 1, 2, \ldots, q$, we have

$$\sum_{A \subseteq V} \left( \sum_{j=1}^{q} \pi_j(A) \right) \cdot \nu_A(u, v) \geq \tfrac{1}{12} \cdot \sum_{j=1}^{q} \Delta_j = \tfrac{1}{48} \rho(u, v).$$

Dividing by $q$ and using $\sum_{A \subseteq V} \pi_j(A) = 1$, we arrive at

$$\sum_{A \subseteq V} \alpha_A \nu_A \geq \frac{1}{48q} \rho, \qquad \sum_{A \subseteq V} \alpha_A = 1,$$

with $\alpha_A = \frac{1}{q} \sum_{j=1}^{q} \pi_j(A)$. Lemma 15.7.3 now gives embeddability into $\ell_2$ with distortion at most $48q$. Theorem 15.7.1 is proved. □

**Remarks.** Almost the same proof with a slight modification of Lemma 15.7.3 shows that for each $p \in [1, \infty)$, every $n$-point metric space can be embedded into $\ell_p$ with distortion $O(\log n)$; see Exercise 1.

The proof as stated produces an embedding into space of dimension $2^n$, since there are $2^n$ subsets $A \subseteq V$, each of them yielding one coordinate. To reduce the dimension, one can argue that not all the sets $A$ are needed: by suitable Chernoff-type estimates, it follows that it is sufficient to choose $O(\log n)$ random sets with point probability $2^{-j}$, i.e. $O(\log^2 n)$ sets altogether (Exercise 2). Of course, for Euclidean embeddings, an even better dimension $O(\log n)$ is obtained using the Flattening Lemma 15.2.1, but for other $\ell_p$, no flattening lemma is available.

**An algorithmic application: approximating the sparsest cut.** We know that every $n$-point metric space can be $O(\log n)$-embedded into $\ell_1^d$ with $d = O(\log^2 n)$. By inspecting the proof, it is not difficult to give a randomized algorithm which computes such an embedding in polynomial expected time. We show a neat algorithmic application to a graph-theoretical problem.

Let $G = (V, E)$ be a graph. A *cut* in $G$ is a partition of $V$ into two nonempty subsets $A$ and $B = V \setminus A$. The *density* of the cut $(A, B)$ is $\frac{e(A,B)}{|A| \cdot |B|}$, where $e(A, B)$ is the number of edges connecting $A$ and $B$. Given $G$, we would like to find a cut of the smallest possible density. This problem is NP-hard, and here we discuss an efficient algorithm for finding an approximate answer: a cut whose density is at most $O(\log n)$-times larger than the density of the sparsest cut, where $n = |V|$ (this is the best known approximation guarantee for any polynomial-time algorithm). Note that this also allows us to approximate the edge expansion of $G$ (discussed in Section 15.5) within a multiplicative factor of $O(\log n)$.

First we re-formulate the problem equivalently using cut pseudometrics. A *cut pseudometric* on $V$ is a pseudometric $\tau$ corresponding to some cut $(A, B)$, with $\tau(u,v) = \tau(v,u) = 1$ for $u \in A$ and $v \in B$ and $\tau(u,v) = 0$ for $u, v \in A$ or $u, v \in B$. In other words, a cut pseudometric is a line pseudometric induced by a mapping $\psi: V \to \{0,1\}$ (excluding the trivial case where all of $V$ gets mapped to the same point). Letting $F = \binom{V}{2}$, the density of the cut $(A, B)$ can be written as $\frac{\tau(E)}{\tau(F)}$, where $\tau$ is the corresponding cut pseudometric and $\tau(E) = \sum_{\{u,v\}\in E} \tau(u,v)$ and similarly for $\tau(F)$. Therefore, we would like to minimize the ratio $R_1(\tau) = \frac{\tau(E)}{\tau(F)}$ over all cut pseudometrics $\tau$.

In the first step of the algorithm, we relax the problem and we find a pseudometric, not necessarily a cut one, minimizing the ratio $R_1(\rho) = \frac{\rho(E)}{\rho(F)}$. This can be done efficiently by linear programming. The minimized function looks nonlinear but we can get around this by a simple trick: we postulate the additional condition $\rho(F) = 1$ and minimize the linear function $\rho(E)$. The variables in the linear program are the $\binom{n}{2}$ numbers $\rho(u,v)$ for $\{u,v\} \in F$ and the constraints are $\rho(u,v) \geq 0$ (for all $u, v$), $\rho(F) = 1$, and those expressing the triangle inequalities for all triples $u, v, w \in V$.

Having computed a $\rho_0$ minimizing $R_1(\rho)$, we find a $D$-embedding $f$ of $(V, \rho_0)$ into some $\ell_1^d$ with $D = O(\log n)$. If $\sigma_0$ is the pseudometric induced on $V$ by this $f$, we clearly have $R_1(\sigma_0) \leq D \cdot R_1(\rho_0)$. Now since $\sigma_0$ is an $\ell_1$-pseudometric, it can be expressed as a nonnegative linear combination of suitable cut pseudometrics (Exercise 15.5.3): $\sigma_0 = \sum_{i=1}^{N} \alpha_i \tau_i$, $\alpha_1, \ldots, \alpha_N > 0$, $N \leq d(n-1)$. It is not difficult to check that $R_1(\sigma_0) \geq \min\{R_1(\tau_i): i = 1, 2, \ldots, N\}$ (Exercise 3). Therefore, at least one of the $\tau_i$ is a cut pseudometric satisfying $R_1(\tau_i) \leq R_1(\sigma_0) \leq D \cdot R_1(\rho_0) \leq D \cdot R_1(\tau_0)$, where $\tau_0$ is a cut pseudometric with the smallest possible $R_1(\tau_0)$. Therefore, the cut corresponding to this $\tau_i$ has density at most $O(\log n)$-times larger than the sparsest possible cut.

**Bibliography and Remarks.** Theorem 15.7.1 is due to Bourgain [Bou85]. The algorithmic application to approximating the sparsest cut uses the idea of an algorithm for a somewhat more complicated problem (multicommodity flow) found by Linial et al. [LLR95] and independently by Aumann and Rabani [AR98].

We will briefly discuss further results proved by variations of Bourgain's embedding technique. Many of them have been obtained in the study of approximation algorithms and imply strong algorithmic results.

*Tree metrics.* Let $\mathcal{G}$ be a class of graphs and consider a graph $G \in \mathcal{G}$. Each positive weight function $w: E(G) \to (0, \infty)$ defines a metric on $V(G)$, namely the shortest-path metric, where the length of a path is the sum of the weights of its edges. All subspaces of the resulting metric spaces are referred to as *$\mathcal{G}$-metrics*. A *tree metric* is a $\mathcal{T}$-metric for $\mathcal{T}$ being the class of all trees. Tree metrics generally behave much better than arbitrary metrics, but for embedding problems they are far from being trivial.

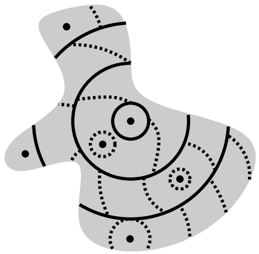
Bourgain [Bou86] proved, using martingales, a surprising lower bound for embedding tree metrics into $\ell_2$: a tree metric on $n$ points requires distortion $\Omega(\sqrt{\log\log n})$ in the worst case. His example is the complete binary tree with unit edge lengths, and for that example, he also constructed an embedding with $O(\sqrt{\log\log n})$ distortion. For embedding the complete binary tree into $\ell_p$, $p>1$, the distortion is $\Omega((\log\log n)^{\min(1/2,1/p)})$, with the constant of proportionality depending on $p$ and tending to 0 as $p\to 1$. (For Banach-space specialists, we also remark that all tree metrics can be embedded into a given Banach space $Z$ with bounded distortion if and only if $Z$ is not superreflexive.) In Matoušek [Mat99b], it was shown that the complete binary tree is essentially the worst example, i.e. every $n$-point tree metric can be embedded into $\ell_p$ with distortion $O((\log\log n)^{\min(1/2,1/p)})$, and an alternative, elementary proof was given for the matching lower bound (see Exercise 5 for a weaker version). Another proof of the lower bound, very short but applying only for embeddings into $\ell_2$, was found by Linial and Saks [LS01] (Exercise 6).

In the notes to Section 15.3, we mentioned that general $n$-point metric spaces required worst-case distortion $\Omega(n^{1/\lfloor(d+1)/2\rfloor})$ for embedding into $\ell_2^d$, $d\geq 2$ fixed. Gupta [Gup00] proved that for $n$-point tree metrics, $O(n^{1/(d-1)})$-embeddings into $\ell_2^d$ are possible. The best known lower bound is $\Omega(n^{1/d})$, from a straightforward volume argument.

*Planar-graph metrics and metrics with excluded minor.* A *planar-graph metric* is a $\mathcal{P}$-metric with $\mathcal{P}$ standing for the class of all planar graphs (in the literature, the shorter but potentially confusing term *planar metric* is used). Rao [Rao99] proved that every $n$-point planar-graph metric can be embedded into $\ell_2$ with distortion only $O(\sqrt{\log n}\,)$, as opposed to $\log n$ for general metrics. More generally, the same method shows that whenever $H$ is a fixed graph and $Excl(H)$ is the class of all graphs not containing $H$ as a minor, then $Excl(H)$-metrics can be $O(\sqrt{\log n}\,)$-embedded into $\ell_2$. For a matching lower bound, valid already for the class $Excl(K_4)$, i.e. series-parallel graphs, and consequently for planar-graph metrics too, see Exercise 15.4.2.

Now we outline Rao's method of embedding. We begin with graphs where all edges have unit weight (this is the setting in [Rao99] but our presentation differs in some details) and then we indicate how graphs with arbitrary edge weights can be treated. The main new ingredient in Rao's method, compared to Bourgain's approach, is a result of Klein, Plotkin, and Rao [KPR93] about a decomposition of graphs with an excluded minor into pieces of low diameter. Here is the decomposition procedure.

Let $G$ be a graph, let $\rho$ be the corresponding graph metric (with all edges having unit length), and let $\Delta$ be an integer parameter. We fix a vertex $v_0\in V(G)$ arbitrarily, we choose an integer $r\in\{0,1,\ldots,\Delta{-}1\}$ uniformly at random, and we let $B_1=\{v\in V(G)\colon \rho(v,v_0)\equiv r \pmod{\Delta}\}$. By deleting the vertices of $B_1$ from $G$, the remaining vertices are partitioned into connected components; this is the first level of the decomposition. For each of these components of $G\setminus B_1$, we repeat the same procedure; $\Delta$ remains unchanged and $r$ is chosen anew at random (but we can use the same $r$ for all the components). Let $B_2$ be the set of vertices deleted from $G$ in this second round, taken together for all the components. The second level of the decomposition consists of the connected components of $G\setminus(B_1\cup B_2)$, and decompositions of levels 3,4,... can be produced similarly. The following schematic drawing illustrates the two-level decomposition; the graph is marked as the gray area and the vertices of $B_1$ and $B_2$ are indicated by the full and dashed arcs, respectively.



For planar graphs, it suffices to use a 3-level decomposition, and for every fixed graph $H$, there is a suitable $k = k(H)$ such that a $k$-level decomposition is appropriate for all graphs $G \in Excl(H)$.

Let $B = B_1 \cup \cdots \cup B_k$; this can be viewed as the boundary of the components in the $k$-level decomposition. The key properties of the decomposition are:

(i) For each vertex $v \in V(G)$, we have $\rho(v, B) \geq c_1\Delta$ with probability at least $c_2$, for suitable constants $c_1, c_2 > 0$. The probability is with respect to the random choices of the parameters $r$ at each level of the decomposition. (This is not hard to see; for example, in the first level of the decomposition, for every fixed $v$, $\rho(v, v_0)$ is some fixed number and it has a good chance to be at least $c_1\Delta$ away, modulo $\Delta$, from a random $r$.)

(ii) Each component in the resulting decomposition has diameter at most $O(\Delta)$. (This is not so easy to prove and it is where one needs $k = k(H)$ sufficiently large. For $H = K_{3,3}$, which includes the case of planar graphs, the proof is a relatively simple case analysis.)

Next, we describe the embedding of $V(G)$ into $\ell_2$ in several steps. First we consider $\Delta$ and the decomposition as above fixed, and we let $C_1, \ldots, C_m$ be the components of $G \setminus B$. For all the $C_i$, we choose random signs $\sigma(C_i) \in \{-1, +1\}$ uniformly and independently. For a vertex $x \in V(G)$, we define $\sigma(x) = 0$ if $x \in B$ and $\sigma(x) = \sigma(C_i)$ if $x \in V(C_i)$. Then we define the mapping $\varphi_{B,\sigma}: V(G) \to \mathbf{R}$ by $\varphi_{B,\sigma}(v) = \sigma(x) \cdot \rho(x, B)$ (the distance of $x$ to the boundary signed by the component's sign). This $\varphi_{B,\sigma}$ induces a line pseudometric $\nu_{B,\sigma}$, and it is easy to see that $\nu_{B,\sigma}$ is dominated by $\rho$.

Let $C$ be a constant such that all the $C_i$ have diameter at most $C\Delta$, and let $x, y \in V(G)$ be such that $C\Delta < \rho(x, y) \leq 2C\Delta$. Such $x$ and $y$ certainly lie in distinct components, and $\sigma(x) \neq \sigma(y)$ with probability $\frac{1}{2}$. With probability at least $c_2$, we have $\rho(x, B) \geq c_1\Delta$, and so with a fixed positive probability, $\nu_{B,\sigma}$ places $x$ and $y$ at distance at least $c_1\Delta$.

Now we still keep $\Delta$ fixed and consider $\nu_{B,\sigma}$ for all possible $B$ and $\sigma$. Letting $\alpha_{B,\sigma}$ be the probability that a particular pair $(B, \sigma)$ results from the decomposition procedure, we have

$$\sum_{B,\sigma} \alpha_{B,\sigma} \nu_{B,\sigma}(x, y) = \Omega(\rho(x, y))$$

whenever $C\Delta < \rho(x, y) \leq 2C\Delta$. As in the proof of Lemma 15.7.3, this yields a 1-Lipschitz embedding $f_\Delta: V(G) \to \ell_2^N$ (for some $N$) which shortens distances for pairs $x, y$ as above by at most a constant factor. (It is not really necessary to use all the possible pairs $(B, \sigma)$ in the embedding; it is easy to show that $const \cdot \log n$ independent random $B$ and $\sigma$ will do.)

To construct the final embedding $f: V(G) \to \ell_2$, we let $f(v)$ be the concatenation of the vectors $f_\Delta$ for $\Delta \in \{2^j : 1 \leq 2^j \leq \operatorname{diam}(G)\}$. No distance is expanded by more than $O(\sqrt{\log \operatorname{diam}(G)}\,) = O(\sqrt{\log n}\,)$ and the contraction is at most by a constant factor, and so we have an embedding into $\ell_2$ with distortion $O(\sqrt{\log n}\,)$.

Why do we get a better bound than for Bourgain's embedding? In both cases, we have about $\log n$ groups of coordinates in the embedding. In Rao's embedding we know that for every pair $(x,y)$, one of the groups contributes at least a fixed fraction of $\rho(x,y)$ (and no group contributes more than $\rho(x,y)$). Thus, the sum of squares of the contributions is between $\rho(x,y)^2$ and $\rho(x,y)^2 \log n$. In Bourgain's embedding (with a comparable scaling), no group contributes more than $\rho(x,y)$ and the sum of the contributions of all groups is at least a fixed fraction of $\rho(x,y)$. But since we do not know how the contributions are distributed among the groups, we can only conclude that the sum of squares of the contributions is between $\rho(x,y)^2/\log n$ and $\rho(x,y)^2 \log n$.

It remains to sketch the modifications of Rao's embedding for a graph $G$ with arbitrary nonnegative weights on edges. For the unweighted case, we defined $B_1$ as the vertices lying exactly at the given distances from $v_0$. In the weighted case, there need not be vertices exactly at these distances, but we can add artificial vertices by subdividing the appropriate edges; this is a minor technical issue. A more serious problem is that the distances $\rho(x,y)$ can be in a very wide range, not just from 1 to $n$. We let $\Delta$ run through all the relevant powers of 2 (i.e. such that $C\Delta < \rho(x,y) \leq 2C\Delta$ for some $x \neq y$) but for producing the decomposition for a particular $\Delta$, we use a modified graph $G_\Delta$ obtained from $G$ by contracting all edges shorter than $\frac{\Delta}{2n}$. In this way, we can have much more than $\log n$ values of $\Delta$ but only $O(\log n)$ of them are relevant for each pair $(x,y)$ and the analysis works as before.

Gupta, Newman, Rabinovich, and Sinclair [GNRS99] proved that any $\mathit{Excl}(K_4)$-metric, as well as any $\mathit{Excl}(K_{2,3})$-metric, can be $O(1)$-embedded into $\ell_1$, and they conjectured that for any $H$, $\mathit{Excl}(H)$-metrics might be $O(1)$-embeddable into $\ell_1$ (the constant depending on $H$).

*Volume-respecting embeddings.* Feige [Fei00] introduced an interesting strengthening of the notion of the distortion of an embedding, concerning embeddings into Euclidean spaces. Let $f\colon (V,\rho) \to \ell_2$ be an embedding which, for simplicity, we require to be 1-Lipschitz (nonexpanding). The usual distortion of $f$ is determined by looking at pairs of points, while Feige's notion takes into account all $k$-tuples for some $k \geq 2$. For example, if $V$ has 3 points, every two with distance 1, then the following two embeddings into $\ell_2^2$ have about the same distortion:

But while the left embedding is good in Feige's sense for $k = 3$, the right one is completely unsatisfactory. For a $k$-point set $P \subset \ell_2$, define $\mathrm{Evol}(P)$ as the $(k-1)$-dimensional volume of the simplex spanned by $P$ (so $\mathrm{Evol}(P) = 0$ if $P$ is affinely dependent). For a $k$-point metric space $(S,\rho)$, the *volume* $\mathrm{Vol}(S)$ is defined as $\sup_f \mathrm{Evol}(f(S))$, where the supremum is over all 1-Lipschitz $f\colon S \to \ell_2$. An embedding $f\colon (V,\rho) \to \ell_2$ is $(k,D)$ *volume-respecting* if for every $k$-point subset $S \subseteq V$, we have $D \cdot \mathrm{Evol}(f(S))^{1/(k-1)} \geq \mathrm{Vol}(S)^{1/(k-1)}$. For $D$ small, this means that the image of any $k$-tuple spans nearly as large volume as it possibly can for a 1-Lipschitz map.

Feige showed that $\mathrm{Vol}(S)$ can be approximated quite well by an intrinsic parameter of the metric space (not referring to embeddings), namely by the *tree volume* $\mathrm{Tvol}(S)$, which equals the products of the edge lengths in a minimum spanning tree on $S$ (with respect to the metric on $S$). Namely, $\mathrm{Vol}(S) \leq \frac{1}{(k-1)!}\,\mathrm{Tvol}(S) \leq 2^{(k-2)/2}\,\mathrm{Vol}(S)$. He proved that for any $n$-point metric space and all $k \geq 2$, the embedding as in the proof of Theorem 15.7.1 is $(k, O(\log n + \sqrt{k \log n \log k}\,))$ volume-respecting (the result in the conference version of his paper is slightly weaker).

The notion of volume-respecting embeddings currently still looks somewhat mysterious. In an attempt to convey some feeling about it, we outline Feige's application and indicate the use of the volume-respecting condition in it. He considered the problem of approximating the *bandwidth* of a given $n$-vertex graph $G$. The bandwidth is the minimum, over all bijective maps $\varphi: V(G) \to \{1, 2, \ldots, n\}$, of $\max\{|\varphi(u) - \varphi(v)|: \{u, v\} \in E(G)\}$ (so it has the flavor of an approximate embedding problem). Computing the bandwidth is NP-hard but Feige's ingenious algorithm approximates it within a factor of $O((\log n)^{const})$. The algorithm has two main steps: first, embed the graph (as a metric space) into $\ell_2^m$, with $m$ being some suitable power of $\log n$, by a $(k, D)$ volume-respecting embedding $f$, where $k = \log n$ and $D$ is as small as one can get. Second, let $\lambda$ be a random line in $\ell_2^m$ and let $\psi(v)$ denote the orthogonal projection of $f(v)$ on $\lambda$. This $\psi: V(G) \to \lambda$ is almost surely injective and so it provides a linear ordering of the vertices, i.e. a bijective map $\varphi: V(G) \to \{1, 2, \ldots, n\}$, and this is used for estimating the bandwidth.

To indicate the analysis, we need the notion of *local density* of the graph $G$: $\mathrm{ld}(G) = \max\{|B(v, r)|/r: v \in V(G), r = 1, 2, \ldots, n\}$, where $B(v, r)$ are all vertices at distance at most $r$ from $v$. It is not hard to see that $\mathrm{ld}(G)$ is a lower bound for the bandwidth, and Feige's analysis shows that $O(\mathrm{ld}(G)(\log n)^{const})$ is an upper bound.

One first verifies that, with high probability, if $\{u, v\} \in E(G)$ then the images $\psi(u)$ and $\psi(v)$ on $\lambda$ are close, concretely $|\psi(u) - \psi(v)| \le \Delta = O(\sqrt{(\log n)/m})$. For proving this, it suffices to know that $f$ is 1-Lipschitz and it is an immediate consequence of measure concentration on the sphere. If $b$ is the bandwidth obtained from the ordering given by $\psi$ then some interval of length $\Delta$ on $\lambda$ contains the images of $b$ vertices. Call a $k$-tuple $S \subset V(G)$ *squeezed* if $\psi(S)$ lies in an interval of length $\Delta$. If $b$ is large then there are many squeezed $S$. On the other hand, one proves that, not surprisingly, if $\mathrm{ld}(G)$ is small then $\mathrm{Vol}(S)$ is large for all but few $k$-tuples $S \subset V(G)$. Now the volume-respecting condition enters: if $\mathrm{Vol}(S)$ is large then $\mathrm{conv}(f(S))$ has large $(k-1)$-dimensional volume. It turns out that the projection of a convex set in $\ell_2^m$ with large $(k-1)$-dimensional volume on a random line is unlikely to be short, and so $S$ with large $\mathrm{Vol}(S)$ is unlikely to be squeezed. Thus, by estimating the number of squeezed $k$-tuples in two ways, one gets an inequality bounding $b$ from above in terms of $\mathrm{ld}(G)$.

Vempala [Vem98] applied volume-respecting embeddings in another algorithmic problem, this time concerning arrangement of graph vertices in the plane. Moreover, he also gave alternative proof of some of Feige's lemmas. Rao in the already mentioned paper [Rao99] also obtained improved volume-respecting embeddings for planar metrics.

*Bartal's trees.* As we have seen, in Bourgain's method, for a given metric $\rho$ one constructs a convex combination $\sum \alpha_i \nu_i \ge \frac{1}{D}\rho$, where $\nu_i$ are line pseudometrics dominated by $\rho$. An interesting "dual" result was found by Bartal [Bar96], following an earlier work in this direction by Alon, Karp, Peleg, and West [AKPW95]. He approximates a given $\rho$ by a convex combination $\sum_{i=1}^{N} \alpha_i \tau_i$, where this time the inequalities go in the opposite direction: $\tau_i \ge \rho$ and $\sum \alpha_i \tau_i \le D\rho$, with $D = O(\log^2 n)$ (later he improved this to $O(\log n \log\log n)$ in [Bar98]). The $\tau_i$ are not line metrics (and in general they cannot be), but they are tree metrics, and even of a special form, the so-called *hierarchically well-separated trees.* This means that $\tau_i$ is given as the shortest-path metric of a rooted tree with weighted edges, such that the distances from each vertex to all of its sons are the same, and if $v$ is a son of $u$ and $w$ a son of $v$ then $\tau_i(u, v) \ge K \cdot \tau_i(v, w)$, where $K \ge 1$ is a parameter which can be set at will (and the constant in the bound on $D$ depends on it).

This result has been used in approximation algorithms for problems involving metric spaces, according to the following scheme: choose $i \in \{1, 2, \ldots, N\}$ at random, with each $i$ having probability $\alpha_i$, solve the problem in question for the tree metric $\tau_i$, and show that the expected value of the solution is not very far from the optimal solution for the original metric $\rho$.

Since tree metrics embed isometrically into $\ell_1$, Bartal's result also implies $O(\log n \log\log n)$-embeddability of all $n$-point metric spaces into $\ell_1$, which is just a little weaker than Bourgain's approach (and it also implies that $\Omega(\log n)$ is a lower bound in Bartal's setting). For a simpler proof of a weaker version Bartal's result see Indyk [Ind01].

## Exercises

1. (Embedding into $\ell_p$) Prove that under the assumptions of Lemma 15.7.3, the metric space $(V,\rho)$ can be $D$-embedded into $\ell_p^N$, $1 \le p \le \infty$, with distortion at most $D$. (You may want to start with the rather easy cases $p = 1$ and $p = \infty$, and use Hölder's inequality for an arbitrary $p$.) $\boxed{3}$

2. (Dimension reduction for the embedding)

   (a) Let $E_1, \ldots, E_m$ be independent events, each of them having probability at least $\frac{1}{12}$. Prove that the probability of no more than $\frac{m}{24}$ of the $E_i$ occurring is at most $e^{-cm}$, for a sufficiently small positive constant $c$. Use suitable Chernoff-type estimates or direct estimates of binomial coefficients. $\boxed{3}$

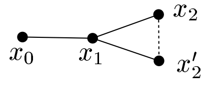
   (b) Modify the proof of Theorem 15.7.1 as follows: for each $j = 1, 2, \ldots, q$, pick sets $A_{ij}$ independently at random, $i = 1, 2, \ldots, m$, where the points are included in $A_{ij}$ with probability $2^{-j}$ and where $m = C \log n$ for a sufficiently large constant $C$. Using (a) and Lemmas 15.7.2 and 15.7.3, prove that with a positive probability, the embedding $f: V \to \ell_2^{qm}$ given by $f(v)_{ij} = \rho(v, A_{ij})$ has distortion $O(\log n)$. $\boxed{3}$

3. Let $a_1, a_2, \ldots, a_n$, $b_1, b_2, \ldots, b_n$, $\alpha_1, \alpha_2, \ldots, \alpha_n$ be positive real numbers. Show that $\frac{\alpha_1 a_1 + \alpha_2 a_2 + \cdots + \alpha_n a_n}{\alpha_1 b_1 + \alpha_2 b_2 + \cdots + \alpha_n b_n} \ge \min\{\frac{a_1}{b_1}, \frac{a_2}{b_2}, \ldots, \frac{a_n}{b_n}\}$. $\boxed{2}$

4. Let $P_n$ be the metric space $\{0, 1, \ldots, n\}$ with the metric inherited from $\mathbf{R}$ (or a path of length $n$ with the graph metric). Prove the following Ramsey-type result: for every $D > 1$ and every $\varepsilon > 0$ there exists an $n = n(D, \varepsilon)$ such that whenever $f: P_n \to (Z, \sigma)$ is a $D$-embedding of $P_n$ into some metric space, then there are $a < b < c$, $b = \frac{a+c}{2}$, such that $f$ restricted to the subspace $\{a, b, c\}$ of $P_n$ is a $(1+\varepsilon)$-embedding. That is, if a sufficiently long path is $D$-embedded then it contains a scaled copy of a path of length 2 embedded with distortion close to 1. $\boxed{4}$ Can you extend the proof so that it provides a scaled copy of a path of length $k$?

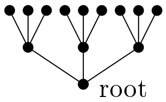
5. (Lower bound for embedding trees into $\ell_2$)

   (a) Show that for every $\varepsilon > 0$ there exists $\delta > 0$ with the following property. Let $x_0, x_1, x_2, x_2' \in \ell_2$ be points such that $\|x_0 - x_1\|, \|x_1 - x_2\|, \|x_1 - x_2'\| \in [1, 1+\delta]$ and $\|x_0 - x_2\|, \|x_0 - x_2'\| \in [2, 2+\delta]$ (so all the distances are almost like the graph distances in the following tree, except possibly for the one marked by dotted line).




   Then $\|x_2 - x_2'\| \le \varepsilon$, i.e. the remaining distance must be very short. $\boxed{3}$

   (b) Let $T_{k,m}$ denote the complete $k$-ary tree of height $m$; the following picture shows $T_{3,2}$:




Show that for every $r$ and $m$ there exists $k$ such that whenever the leaves of $T_{k,m}$ are colored by $r$ colors, there is a subtree of $T_{k,m}$ isomorphic to $T_{2,m}$ with all leaves having the same color. $\boxed{2}$

(c) Use (a),(b), and Exercise 4 to prove that for any $D > 1$, there exist $m$ and $k$ such that the tree $T_{k,m}$ considered as a metric space with the shortest-path metric cannot be $D$-embedded into $\ell_2$. $\boxed{4}$

6. (Another lower bound for embedding trees into $\ell_2$)

   (a) Let $x_0, x_1, \ldots, x_n$ be arbitrary points in a Euclidean space (thought of as images of the vertices of a path of length $n$ under some embedding). Let $\Gamma = \{(a, a+2^k, a+2^{k+1})\colon a = 0, 1, 2, \ldots, a+2^{k+1} \leq n, k = 0, 1, 2 \ldots\}$. Prove
$$\sum_{(a,b,c)\in\Gamma} \frac{\|x_a - 2x_b + x_c\|^2}{(c-a)^2} \leq \sum_{a=0}^{n-1} \|x_a - x_{a+1}\|^2;$$
   this shows that an average triple $(x_a, x_b, x_c)$ is "straight" (and provides an alternative solution to Exercise 4 for $Z = \ell_2$). $\boxed{3}$

   (b) Prove that the complete binary tree $T_{2,m}$ requires $\Omega(\sqrt{\log m}\,)$ distortion for embedding into $\ell_2$. Consider a nonexpanding embedding $f\colon V(T_{2,m}) \to \ell_2$ and sum up the inequalities as in (a) over all images of the root-to-leaf paths. $\boxed{4}$

7. (Bourgain's embedding of complete binary trees into $\ell_2$) Let $B_m = T_{2,m}$ be the complete binary tree of height $m$ (notation as in Exercise 5). We identify the vertices of $B_m$ with words of length at most $m$ over the alphabet $\{0, 1\}$: the root of $B_m$ is the empty word and the sons of a vertex $w$ are the vertices $w0$ and $w1$. We define the embedding $f\colon V(B_m) \to \ell_2^{|V(B_m)|-1}$, where the coordinates in the range of $f$ are indexed by the vertices of $B_m$ distinct from the root, i.e. by nonempty words. For a word $w \in V(B_m)$ of length $a$, let $f(w)_u = \sqrt{a-b+1}$ if $u$ is a nonempty initial segment of $w$ of length $b$, and $f(w)_u = 0$ otherwise. Prove that this embedding has distortion $O(\sqrt{\log m}\,)$. $\boxed{4}$

8. Prove that any finite tree metric can be isometrically embedded into $\ell_1$. $\boxed{3}$

9. (Low-dimensional embedding of trees)

   (a) Let $T$ be a tree (in the graph-theoretical sense) on $n \geq 3$ vertices. Prove that there exists subtrees $T_1$ and $T_2$ of $T$ which share a single vertex and no edge and together cover $T$, such that $\min(|V(T_1)|, |V(T_2)|) \leq 1+\frac{2}{3}n$. $\boxed{3}$

   (b) Using (a), prove that every tree metric space with $n$ points can be isometrically embedded into $\ell_\infty^d$ with $d = O(\log n)$. $\boxed{4}$

   This result is from [LLR95].