15

Embedding Finite Metric Spaces into Normed Spaces

15.1 Introduction: Approximate Embeddings

We recall that a *metric space* is a pair (X, ρ), 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 $\rho(x, y) + \rho(y, z) \geq \rho(x, z)$.

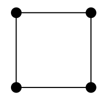
A metric ρ on an n-point set X can be specified by an $n \times n$ matrix of real numbers (actually $\binom{n}{2}$ numbers suffice because of the symmetry). Such tables really arise, for example, in microbiology: X is a collection of bacterial strains, and for every two strains, one can obtain their *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.

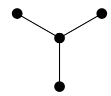
For example, it would be very nice if we could assign to each $x \in X$ a point $f(x)$ in the plane in such a way that $\rho(x, y)$ equals the Euclidean distance of $f(x)$ and $f(y)$. Such representation would allow us to see the structure of the metric space: tight clusters, isolated points, and so 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 quantities concerning a point set in the plane can be computed by efficient geometric algorithms, which are not available for an arbitrary metric space.

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

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

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





The metrics on these 4-point sets are 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 $f\colon X \to Y$, where X is a metric space with a metric ρ and Y is a metric space with a metric σ, 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 $f\colon X \to Y$, where X is a metric space with a metric ρ and Y is a metric space with a metric σ, 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 rescale 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 $\|f\|_{\mathrm{Lip}} = \sup\{\sigma(f(x), f(y))/\rho(x, y)\colon x, y \in X,\, x \neq y\}$, then the distortion of f equals $\|f\|_{\mathrm{Lip}} \cdot \|f^{-1}\|_{\mathrm{Lip}}$.

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 Section 15.2

we start on an optimistic note: We present a surprising positive result of considerable theoretical and practical importance. Before that, we review a few definitions concerning ℓ_p-spaces.

The spaces ℓ_p and ℓ_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 ℓ_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 ℓ_1-norm (sometimes called the Manhattan distance). The ℓ_∞-norm, or *maximum norm*, is given by $\|x\|_\infty = \max_i |x_i|$. It is the limit of the ℓ_p-norms as $p \to \infty$.

Let ℓ_p^d denote the space $\mathbf{R}^d$ equipped with the ℓ_p-norm. In particular, we write ℓ_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 ℓ_p^d, with p given but without restrictions on the dimension d; for example, we can ask whether there exists some Euclidean space into which a given metric space embeds isometrically. Then it is convenient to speak about ℓ_p, which is the space of all infinite sequences $x = (x_1, x_2, \ldots)$ of real numbers with $\|x\|_p < \infty$, where $\|x\|_p = \left(\sum_{i=1}^{\infty} |x_i|^p\right)^{1/p}$. In particular, ℓ_2 is the *(separable) Hilbert space*. The space ℓ_p contains each ℓ_p^d isometrically, and it can be shown that any finite metric space isometrically embeddable into ℓ_p can be isometrically embedded into ℓ_p^d for some d. (In fact, every n-point subspace of ℓ_p can be isometrically embedded into ℓ_p^d with $d \le \binom{n}{2}$; see Exercise 15.5.2.)

Although the spaces ℓ_p are interesting mathematical objects, we will not really study them; we only use embeddability into ℓ_p as a convenient shorthand for embeddability into ℓ_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. A survey with a similar range is Indyk and Matoušek [IM04], and one mainly focused on algorithmic aspects is Indyk [Ind01]; however, both are already outdated because of a very rapid development of the field.

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 ℓ_2 can be found in old papers of Schoenberg (e.g., [Sch38], building on the work of mathematicians like Menger and von Neumann). A book devoted mainly to isometric embeddings, and embeddings into ℓ_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 ε; see Benyamini and Lindenstrauss [BL99], Chapters 14 and 15, for an introduction.

Exercises

1. Consider the two 4-point examples presented above (the square and the star); prove that they cannot be isometrically embedded into ℓ_2^2. [2] Can you determine the minimum necessary distortion for embedding into ℓ_2^2?
2. (a) Prove that a bijective mapping f between metric spaces is a D-embedding if and only if $\|f\|_{\rm Lip} \cdot \|f^{-1}\|_{\rm Lip} \le D$. [1]
   (b) Let (X,ρ) be a metric space, $|X| \ge 3$. Prove that the distortion of an embedding $f\colon X \to Y$, where (Y,σ) is a metric space, equals the supremum of the factors by which f "spoils" the ratios of distances; that is,
$$\sup\left\{\frac{\sigma(f(x),f(y))/\sigma(f(z),f(t))}{\rho(x,y)/\rho(z,t)}\colon x,y,z,t\in X,\ x\neq y, z\neq t\right\}.$$
   [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 ℓ_2^n, can be almost isometrically embedded into ℓ_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 ℓ_2^k, where $k = O(\varepsilon^{-2}\log n)$.*

This result shows that any metric question about n points in ℓ_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 of ℓ_2^n in a computer, we need to store n^2 numbers. To store all of 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 error at most 10%.

Various proofs of the flattening lemma, including the one below, provide efficient randomized algorithms that find the almost isometric embedding into ℓ_2^k quickly. Numerous algorithmic applications have recently been found: in fast clustering of high-dimensional point sets, in approximate searching for nearest neighbors, in approximate multiplication of matrices, and also in purely graph-theoretic 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 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 $m = m(n,k)$:

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

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

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

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 ℓ_2^n (obtained by a random rotation of L_0).

The chosen L is a copy of ℓ_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

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

is violated with probability at most n^{-2}. Since there are fewer than n^2 pairs of distinct $x, y \in X$, there exists 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 ℓ_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 as $(1-\frac{\varepsilon}{3})m\,\|u\| \leq \|p(u)\| \leq (1+\frac{\varepsilon}{3})m\,\|u\|$. This is invariant under scaling, and so we may suppose that $\|u\| = 1$. The condition thus becomes

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

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 (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 $p\colon X \to \ell_2^k$ 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 $f\colon X \to \ell_2$, what is the smallest $C = C(n)$ such that there is always a C-Lipschitz mapping $\bar{f}\colon Y \to \ell_2$ 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 [DG03]. For more general concentration results and techniques using the Gaussian distribution see, e.g., [Pis89], [MS86].

Achlioptas [Ach03] proved that the components of the b_i can also be chosen as independent uniform ± 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 sharpest known upper bound on the dimension needed for a $(1+\varepsilon)$-embedding of an n-point Euclidean metric is $\frac{4}{\varepsilon^2}(1+o(1))\ln n$, where $o(1)$ is with respect to $\varepsilon \to 0$ [IM98], [DG03], [Ach03]. The main term is optimal for the current proof method; see Exercises 3 and 15.3.4.

The Johnson–Lindenstrauss flattening lemma has been 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. [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. [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 $(1{+}\varepsilon)$-embedded into ℓ_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 $i \neq j$ (the constant can be improved). [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, $i \neq j$. Prove that A has rank at least $\frac{n}{2}$. [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 $\binom{k+d}{k}$. [4]
   (d) Using (a)–(c), prove that if the set as in (a) is $(1{+}\varepsilon)$-embedded into ℓ_2^k, where $100n^{-1/2} \le \varepsilon \le \frac{1}{2}$, then
   $$k = \Omega\left(\frac{1}{\varepsilon^2 \log \frac{1}{\varepsilon}} \log n\right).$$
   [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 is a construction 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 ℓ 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. Next, $m(3, n)$ is the maximum number of edges of a triangle-free graph on n vertices, and it equals $\lfloor \frac{n}{2} \rfloor \cdot \lceil \frac{n}{2} \rceil$ 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 even cycles and to consider ℓ even, remembering that the bounds for $\ell = 2k$ and $\ell = 2k+1$ are almost the same up to a factor of 2.

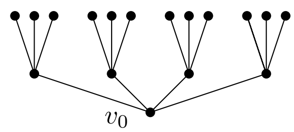
Here is a simple general upper bound on $m(\ell, n)$.

15.3.1 Lemma. *For all n and ℓ,*

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

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 δ 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 contains a tree of height k like this:



The root has δ successors and the other inner vertices of the tree have $\delta - 1$ successors (H may contain additional edges connecting the leaves of the 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 $m = \frac{1}{2}\bar{d}n = \delta n \le n^{1+1/k} + n$. □

This simple argument yields essentially the best known upper bound. But it was asymptotically matched only for a few small values of ℓ, 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 ℓ, the record is also held by algebraic constructions; they are not difficult to describe, but proving that they work needs quite 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.

[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 a 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 $\frac{1}{2}|E(G)|$.

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

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

Of course, for odd ℓ 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 $n \geq 4^{\ell-1} \geq 16$, 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 ℓ or shorter. By deleting all edges of F from $G(n,p)$, we obtain a graph with no cycles of length ℓ or shorter. If we manage to show, for some m, that the expectation $\mathbf{E}[|E \setminus F|]$ is at least 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 ℓ.

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 paths from u to v 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 and 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 ℓ_p^d, and suppose that all n-point metric spaces can be D-embedded into Z. Let ℓ be an integer with $D < \ell \leq 5D$ (it is essential that ℓ 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 set 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 ρ_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 $f_H : (V, \rho_H) \to Z$. By rescaling, 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 ℓ-ball $B_Z(0, \ell) = \{x \in Z : \|x\|_Z \leq \ell\}$.

Set $\beta = \frac{1}{4}(\frac{\ell}{D} - 1)$. We have $0 < \beta \leq 1$. Let N be a β-net in $B_Z(0, \ell)$. The notion of β-net was defined above Lemma 13.1.1, and that lemma showed that a β-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 $|N| \leq (\frac{4\ell}{\beta})^d$.

For every $H \in \mathcal{G}$, we define a new mapping $g_H : V \to N$ 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, a u–v path in H_2 of length smaller than ℓ and the edge $\{u, v\}$ would induce a cycle of length at most ℓ in G. Thus

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

and

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

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

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

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

The bound in the proposition follows by calculation. □

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 ℓ_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 ℓ_2^n, then by the Johnson–Lindenstrauss flattening lemma, it can be $(2D)$-embedded into ℓ_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 ℓ 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. □

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 δ 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 [AHL02] proved by a neat argument using random walks that the same formulas still bound the number of vertices from below if δ 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 ℓ 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 ℓ, 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).

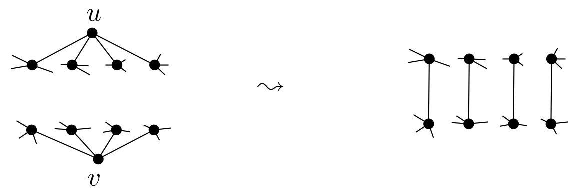
Proposition 15.3.5 and the basic idea of Proposition 15.3.3 were invented by Bourgain [Bou85]. The explicit use of graphs without short cycles and the detection of the "thresholds" in the behavior of the dimension as a function of the distortion appeared in Matoušek [Mat96b].

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 force linear dimension only 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 ℓ_p-norms on $\mathbf{R}^d$ are equivalent up to a constant, and so it suffices to consider embeddings into ℓ_2^d. Considering the n-point metric space with all distances equal to 1, a simple volume argument shows that an embedding into ℓ_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 $\Omega\left(n^{1/\lfloor(d+1)/2\rfloor}\right)$ for embedding into ℓ_2^d (these spaces are even isometrically embeddable into ℓ_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 a geometric realization of such a complex in $\mathbf{R}^{2q+1}$, and filling it with points uniformly (taking an η-net within it for a suitable η, 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 ℓ_2^n (Theorem 15.8.1), and then projecting on a random d-dimensional subspace; the analysis is similar to the proof of the Johnson–Lindenstrauss flattening lemma. 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 δ and ℓ, construct a finite δ-regular graph $G(\delta, \ell)$ with no cycles of length ℓ 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 δ'. [4]
   (b) Let G be a δ-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 ℓ. [2]
(c) Show that starting with a graph as in (a) and reducing it by the operations as in (b), we arrive at a δ-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 ℓ fixed and $n\to\infty$? [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 $t\ge 2$, there exists a t-spanner with $O\left(n^{1+1/\lfloor t/2\rfloor}\right)$ edges. [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; that is, 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 $\mathit{const}\cdot n$. [3]
4. (Another lower bound for the flattening lemma)
(a) Given $\varepsilon\in(0,\frac{1}{2})$ and n sufficiently large in terms of ε, construct a collection $\mathcal{V}$ of ordered n-tuples of points of ℓ_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 ith point and the jth point in V and in V' differ by a factor of at least $1+\varepsilon$), and $\log|\mathcal{V}|=\Omega(\varepsilon^{-2}n\log n)$. [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. [2]

15.4 A Lower Bound for the Hamming Cube

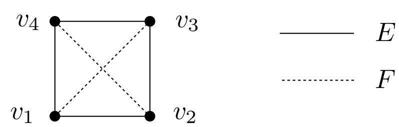
We have established the existence of n-point metric spaces requiring the distortion close to $\log n$ for embedding into ℓ_2 (Proposition 15.3.5), but we have not constructed any specific metric space with this property. In this section we prove a weaker lower bound, only $\Omega(\sqrt{\log n}\,)$, but 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 ℓ_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 ℓ_2 with $D < \sqrt{m} = \sqrt{\log_2 n}$. That is, the natural embedding, where we regard $\{0,1\}^m$ as a subspace of ℓ_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 that 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 notation slightly more complicated than necessary, in anticipation of later developments.

Let V be a finite set, let ρ 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, ρ 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

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

and let us write

$$\mathrm{ave}_2(\rho, E) = \sqrt{\frac{1}{|E|}\,\rho^2(E)}.$$

for the quadratic average of ρ over all pairs in E. We consider the ratio

$$R_{E,F}(\rho) = \frac{\mathrm{ave}_2(\rho, F)}{\mathrm{ave}_2(\rho, E)}.$$

For our 4-cycle, $R_{E,F}(\rho)$ is a kind of ratio of "diagonals to edges" but with quadratic averages of distances, and it equals 2 (right?).

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

If f is a D-embedding, then $R_{E,F}(\sigma) \geq R_{E,F}(\rho)/D$. 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 $R_{E,F}(\sigma)$ can be expected to be smaller than 2 in this case. This is confirmed by the following lemma, which (with $x_i = f(v_i)$) shows that $\sigma^2(F) \leq \sigma^2(E)$, which gives $R_{E,F}(\sigma) \leq \sqrt{2}$ and therefore, $D \geq \sqrt{2}$.

15.4.2 Lemma (Short diagonals lemma). *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 $R_{E,F}(\sigma)$: Squares of Euclidean distances split into the contributions of individual coordinates, and so they are easier to handle 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. □

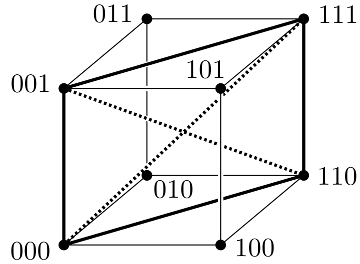
Proof of Theorem 15.4.1. We proceed as in the 2-dimensional case. Let $V = \{0,1\}^m$ be the vertex set of C_m, let ρ be the Hamming metric, let E be the set of edges of the cube (pairs of points at distance 1), and let F be the set of the long diagonals. The long diagonals are pairs of points at distance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

We have $|E| = m2^{m-1}$ and $|F| = 2^{m-1}$, and we calculate $R_{E,F}(\rho) = m$. If σ is a metric on V induced by some embedding $f\colon V \to \ell_2^d$, we want to show that $R_{E,F}(\sigma) \leq \sqrt{m}$; 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, we 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 $u0$, $u \in \{0,1\}^{m-1}$. 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; that is, $F_0 = \{\{u0, \overline{u}0\}\colon u \in \{0,1\}^{m-1}\}$, 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

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

For $u \in \{0,1\}^{m-1}$, we 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

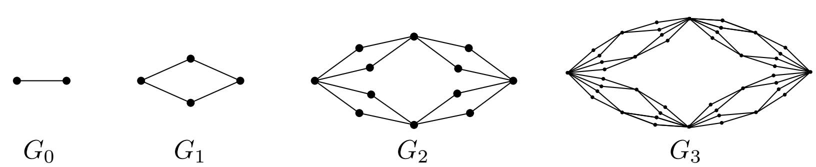
$$\sigma^2(F) \le \sigma^2(E_{01}) + \sigma^2(F_0) + \sigma^2(F_1).$$

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. [3]
2. (Planar graphs badly embeddable into ℓ_2) Let $G_0, G_1, \ldots$ be the following "diamond" graphs:



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

This result is due to Newman and Rabinovich [NR03].

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, ρ) contains a k-point subspace that is $(1+\varepsilon)$-embeddable into ℓ_2. Use Ramsey's theorem. [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 readers not familiar with the important notion of expander graphs, we first include a little wider background.

Roughly speaking, expanders are graphs that 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 that 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 that 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

[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.

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 nonregular 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

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

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

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], Exercise 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 fact:

For all real vectors $(x_v)_{v \in V}$ with $\sum_{v \in V} x_v = 0$, we have $x^T L_G x \geq \mu_2 \|x\|^2$. (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

[4] The notation μ_i for the eigenvalues of L_G is not standard. We use it in order to distinguish these eigenvalues from the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ 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.

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 α_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 ℓ_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 ρ be the shortest-path metric on V. Then the metric space (V, ρ) 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 ℓ_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 $\mathrm{ave}_2(\rho, E) = 1$.To bound $\mathrm{ave}_2(\rho, F)$ from below, we observe that for each vertex v_0, there are at most $1 + r + r(r-1) + \cdots + r(r-1)^{k-1} \leq r^k + 1$ 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 $\mathrm{ave}_2(\rho, F) = \Omega(k) = \Omega(\log n)$. Thus

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

Let $f\colon V \to \ell_2^d$ be an embedding into a Euclidean space, and let σ be the metric induced by it on V. To prove the theorem, it suffices to show that $R_{E,F}(\sigma) = O(1)$; that is,

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

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 that

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

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

$$\begin{aligned}\sum_{\{u,v\}\in F} (x_u - x_v)^2 &= (n-1)\sum_{v\in V} x_v^2 - \sum_{u\neq v} x_u x_v \\ &= n\sum_{v\in V} x_v^2 - \Big(\sum_{v\in V} x_v\Big)^2 = n\|x\|^2.\end{aligned} \tag{15.5}$$

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

$$\sum_{\{u,v\}\in E} (x_u - x_v)^2 = r\sum_{v\in V} x_v^2 - 2\sum_{\{u,v\}\in E} x_u x_v = x^T L_G x \geq \mu_2\|x\|^2,$$

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 $R_{E,F}(\sigma)$ is attained for the σ induced by the mapping $V \to \mathbf{R}$ specified by b_2, the eigenvector belonging to μ_2.

The cone of squared ℓ_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,ρ) be a finite metric space and let $\eta, \varphi\colon \binom{V}{2} \to [0,\infty)$ be weight functions. Let us write

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

15.5.2 Proposition. *Let (V,ρ) be a finite metric space and suppose that (V,ρ) cannot be D-embedded into ℓ_2. Then there are weight functions $\eta, \varphi\colon \binom{V}{2} \to [0,\infty)$, not both identically zero, such that*

$$\rho^2(\varphi) \geq D^2\rho^2(\eta),$$

while

$$\sigma^2(\varphi) \leq \sigma^2(\eta)$$

for every metric σ induced on V by an embedding into ℓ_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 η and φ. (As we will see below, there is even an efficient algorithm for deciding D-embeddability into ℓ_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 $\varphi\colon \binom{V}{2} \to \mathbf{R}$, 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 pseudometrics 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

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

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 $f\colon V \to \ell_2^k$ and $g\colon V \to \ell_2^m$, respectively. We define a new embedding $h\colon V \to \ell_2^{k+m}$ by concatenating the coordinates of f and g; that is,

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

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

Proof of Proposition 15.5.2. Let $\mathcal{L}_2 \subset \mathbf{R}^N$ be the cone of squared Euclidean metrics on V as above and let

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

This $\mathcal{K}$ includes all squares of metrics arising by D-embeddings of (V, ρ). But not all elements of $\mathcal{K}$ are necessarily squares of metrics, since the triangle inequality may be violated. Since there is no Euclidean D-embedding of (V, ρ), 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 (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 a nonzero $a \in \mathbf{R}^N$ such that

$$\langle a, x\rangle \geq 0 \text{ for all } x \in \mathcal{K} \text{ and } \langle a, x\rangle \leq 0 \text{ for all } x \in \mathcal{L}_2. \tag{15.6}$$

Using this a, we define the desired η and φ, as follows:

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

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

First we show that $\rho^2(\varphi) \geq D^2\rho^2(\eta)$. To this end, we employ the property (15.6) for the following $x \in \mathcal{K}$:

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

Then $\langle a, x\rangle \geq 0$ boils down to $\rho^2(\varphi) - D^2\rho^2(\eta) \geq 0$.

Next, let σ 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 σ, i.e., $x_{uv} = \sigma(u,v)^2$. This yields $\sigma^2(\varphi) - \sigma^2(\eta) \leq 0$. Proposition 15.5.2 is proved. □

Algorithmic remark: Euclidean embeddings and semidefinite programming. The problem of deciding whether a given n-point metric space (V,ρ) admits a D-embedding into ℓ_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 ℓ_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 ℓ_1.

The algorithm for D-embedding into ℓ_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,ρ) 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 vth 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 $x^TQx = (x^TX^T)(Xx) = \|Xx\|^2 \geq 0$. (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^TX 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,ρ) can be D-embedded into ℓ_2 if and only if there exists a symmetric real positive semidefinite matrix Q whose entries satisfy the following constraints:

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

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 that until then 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 15.8.

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 an $\Omega(\log n)$ bound for embedding into ℓ_1. The result for Euclidean spaces is a corollary, since every finite Euclidean metric space can be isometrically embedded into ℓ_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 ℓ_p-spaces with arbitrary p in [Mat97]; it was shown that the minimum distortion required to embed all n-point metric spaces into ℓ_p is of order $\frac{\log n}{p}$, and a matching upper bound was proved by the method shown in Section 15.8.

The proof of Theorem 15.5.1 given in the text can easily be extended to prove a lower bound for ℓ_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 ℓ_1-metric is a squared Euclidean metric (see, e.g., Schoenberg [Sch38] for a proof)[5].

[5] Here is an outline of a beautiful proof communicated to me by Assaf Naor. We represent the given ℓ_1 metric by points in $L_1(\mathbf{R})$ (functions $\mathbf{R} \to \mathbf{R}$ with norm $\|f\|_1 = \int_{\mathbf{R}} |f(x)|\,\mathrm{d}x$). The embedding T maps $f \in L_1(\mathbf{R})$ to $g = Tf \in L_2(\mathbf{R}^2)$ defined by $g(x,y) = 1$ if $f(x) \in [0,y]$ and $g(x,y) = 0$ otherwise. Showing $\|f_1 - f_2\|_1 = \int_{\mathbf{R}^2} (Tf_1(x,y) - Tf_2(x,y))^2\,\mathrm{d}(x,y) = \|Tf_1 - Tf_2\|_2^2$ is easy, and it remains to check (or know) that $L_2(\mathbf{R}^2)$, being a Hilbert space of countable dimension, is isometric to ℓ_2.

Squared Euclidean metrics do not generally satisfy the triangle inequality, but that is not needed in the proof.

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 ℓ_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 ℓ_2-embeddings of graphs was made by Linial, Magen, and Naor [LMN02]. 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 ℓ_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 that 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 ρ be the metric of G. We set $F = \{\{u, v\} \in \binom{V}{2} \colon \rho(u, v) = t\}$; 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 σ on V, which amounts to finding the largest $\beta > 0$ such that $\sigma^2(E) - \beta \cdot \sigma^2(F) \geq 0$ for all σ. Here it suffices to consider line metrics σ; so let $x_v \in \mathbf{R}$ be the image of v in the embedding $V \to \mathbf{R}$ inducing σ. 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-theoretic problems. Lovász's recent lecture notes [Lov03] 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 8 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] gave 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 Δ 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 nonregular 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 of 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 $\lambda_2 \le 2\sqrt{r-1} + O(\log r)$. 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 $r \ge 3$ and any r-regular graph on n vertices, $\lambda_2 \ge 2\sqrt{r-1} - o(1)$, 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 has led to 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 nonresidue 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 2×2 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 $q \to \infty$. For an accessible exposition of some of the beautiful mathematics underlying this construction and a proof that the resulting graphs are expanders see Davidoff, Sarnak, and Valette [DSV03].

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 sophisticated product operations. Some of their parameters are inferior to those from ⇓ NEW ⇓ [Mar88], [LPS88], [Mor94], but the proof is relatively short, and it uses only elementary linear algebra. The ideas of the construction later led to a construction of constant-degree *lossless expanders*, whose edge expansion is much better than can be achieved through bounds on the second eigenvalue [CRVW02], and to several other applications. They also inspired a new proof by Dinur [Din05] of the *PCP theorem*, arguably one of the greatest theorems of all computer science, which was thus moved from the category "a booklet presentable in a one-semester advanced course" to "a paper presentable in a seminar." ⇑ NEW ⇑

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. [3]
2. (Dimension for isometric ℓ_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 $\mathcal{L}_1^{(\text{fin})} \subset \mathbf{R}^N$ be the set of all pseudometrics[6] on V induced by embeddings $f\colon V \to \ell_1^k$, $k = 1, 2, \ldots$. Show that $\mathcal{L}_1^{(\text{fin})}$ is the convex hull of *line pseudometrics*, i.e., pseudometrics induced by mappings $f\colon V \to \ell_1^1$. [2]
   (b) Prove that any metric from $\mathcal{L}_1^{(\text{fin})}$ can be isometrically embedded into ℓ_1^N. That is, any n-point set in some ℓ_1^k can be realized in ℓ_1^N. [4] (Examples show that one cannot do much better and that dimension $\Omega(n^2)$ is necessary, in contrast to Euclidean embeddings, where dimension $n-1$ always suffices.)
   (c) Let $\mathcal{L}_1 \subset \mathbf{R}^N$ be all pseudometrics induced by embeddings of V into ℓ_1 (the space of infinite sequences with finite ℓ_1-norm). Show that $\mathcal{L}_1 = \mathcal{L}_1^{(\text{fin})}$, and thus that any n-point subset of ℓ_1, can be realized in ℓ_1^N. [3]
   (d) Extend the considerations in (a)–(c) to ℓ_p-metrics with arbitrary $p \in [1, \infty)$. [3]
   See Ball [Bal90] for more on the dimension of isometric ℓ_p-embeddings.
3. With the notation as in Exercise 2, show that every line pseudometric ν on an n-point set V is a nonnegative linear combination of at most $n-1$ *cut pseudometrics*: $\nu = \sum_{i=1}^{n-1} \alpha_i \tau_i$, $\alpha_1, \ldots, \alpha_{n-1} \geq 0$, where each τ_i is 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 ℓ_1 is a nonnegative linear combination of cut pseudometrics.) [3]
4. (An ℓ_p-analogue of Proposition 15.5.2) Let $p \in [1, \infty)$ be fixed. Using Exercise 2, formulate and prove an appropriate ℓ_p-analogue of Proposition 15.5.2. [3]
5. (Finite ℓ_2-metrics embed isometrically into ℓ_p)

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

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

(b) Prove that every n-point set in ℓ_2 can be isometrically embedded into ℓ_p^N. $\boxed{2}$

6. (The second eigenvalue and edge expansion) Let G be an r-regular graph with n vertices, and let A be a nonempty proper subset of V. Prove that the number of edges connecting A to $V \setminus A$ is at least $e(A, V \setminus A) \geq \mu_2(G) \cdot \frac{|A| \cdot |V \setminus A|}{n}$ (use (15.3) with a suitable vector x), and deduce that $\Phi(G) \geq \frac{1}{2}\, \mu_2(G)$. $\boxed{4}$
7. (Expansion and measure concentration) Let us consider the vertex set of a graph G as a metric probability space, with the usual graph metric and with the uniform probability measure P (each vertex has measure $\frac{1}{n}$, $n = |V(G)|$). Suppose that $\Phi = \Phi(G) > 0$ and that the maximum degree of G is Δ. Prove the following measure concentration inequality: If $A \subseteq V(G)$ satisfies $\mathrm{P}[A] \geq \frac{1}{2}$, then $1 - \mathrm{P}[A_t] \leq \frac{1}{2} e^{-t\Phi/\Delta}$, where A_t denotes the t-neighborhood of A. $\boxed{3}$
8. (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 that finds an approximate solution using semidefinite programming.

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

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

$\boxed{2}$

(b) Let

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

Clearly, $M_{\mathrm{relax}} \geq M_{\mathrm{opt}}$. Verify that this relaxed version of the problem is an instance of a semidefinite program, that is, 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 $x_v = \mathrm{sgn}\langle y_v, r \rangle$, $v \in V$. Let $M_{\mathrm{approx}} = \frac{1}{2} \sum_{\{u,v\} \in E} (1 - x_u x_v)$ for these x_v. Show that the expectation, with respect to the random choice of r, of M_{approx} is at least $0.878 \cdot M_{\mathrm{relax}}$ (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 value. $\boxed{4}$

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 [KV05] for an interesting conjecture whose validity would imply that approximation with ratio better than in the Goemans–Williamson results is NP-complete).

15.6 A Tight Lower Bound by Fourier Transform

⇓ NEW ⇓

Here we present another class of n-point metric spaces requiring $\Omega(\log n)$ distortion for embedding into any Euclidean space, and more significantly, in the proof we will illustrate yet another powerful method of establishing lower bounds for distortion, based on harmonic analysis.

The basic scheme of the proof has much in common with the lower bounds for the Hamming cube and for expanders presented earlier. We construct an n-point metric space $(\bar{V}, \bar{\rho})$ (for reasons mentioned later, all things in this space will be denoted by letters with bars) and we define two sets of pairs $\bar{E} \subseteq \binom{\bar{V}}{2}$ and $\bar{F} = \binom{\bar{V}}{2}$. We bound from below the ratio

$$R_{\bar{E},\bar{F}}(\bar{\rho}) = \frac{\mathrm{ave}_2(\bar{\rho}, \bar{F})}{\mathrm{ave}_2(\bar{\rho}, \bar{E})}$$

of quadratic averages, thereby showing that a typical distance in $\bar{E}$ is *considerably smaller* than a typical distance in $\bar{F}$, i.e., $\bar{E}$ consists of "short edges." Finally, for any metric $\bar{\sigma}$ induced by an Euclidean embedding $\bar{f}\colon \bar{V} \to \ell_2$, we want to bound $R_{\bar{E},\bar{F}}(\bar{\sigma})$ from above (the "short edges" cannot be so short in the embedding), and this is where the new technique comes in: We will prove an inequality relating $\bar{\sigma}^2(\bar{E})$ and $\bar{\sigma}^2(\bar{F})$ using harmonic analysis. In the present proof we will use only very basic facts about Fourier coefficients, but the idea of employing harmonic analysis puts at one's disposal many tools from this well-developed field, which have already been useful in several problems concerning low-distortion embeddings. Let us note that, as we know from Section 15.5, estimating the maximum of $R_{\bar{E},\bar{F}}(\bar{\sigma})$ is equivalent to to estimating the second eigenvalue of a certain (Laplace) matrix, but here we will not use the language of eigenvalues.

Preliminaries on Fourier transform on the Hamming cube. Let V denote the vertex set of the m-dimensional Hamming cube C_m. We consider the elements of V as m-component vectors of 0's and 1's. For $u, v \in V$, $u+v$ denotes the vector in V whose ith component is $(u_i + v_i) \bmod 2$. We note that $u+v$ is the same as $u-v$.

For a function $f: V \to \mathbf{R}$ we define another function $\hat{f}: V \to \mathbf{R}$, the *Fourier transform* of f, by

$$\hat{f}(u) = 2^{-m} \sum_{v \in V} (-1)^{u \cdot v} f(v),$$

where the inner product $u \cdot v$ is defined as $(u_1 v_1 + u_2 v_2 + \cdots + u_m v_m) \bmod 2$.

Readers acquainted with Fourier series or with Fourier transforms in other settings will find the following fact familiar:

15.6.1 Fact (Parseval's equality). *For every function $f: V \to \mathbf{R}$ we have*

$$\sum_{v \in V} f(v)^2 = \sum_{u \in V} \hat{f}(u)^2.$$

In our setting, where we always deal with finite sums, Parseval's equality is quite simple to prove; see Exercise 1. Geometrically, the equality holds because $(f(v): v \in V)$ are the coordinates of a vector in an orthonormal basis, and $(\hat{f}(u): u \in V)$ are the coordinates of the same vector in another orthonormal basis. In textbooks on harmonic analysis one can find Parseval's equality in a general setting, encompassing our situation but also the classical case of Fourier series of periodic real functions on $\mathbf{R}$ and many other useful cases. Then the proofs are more demanding, since one has to deal with issues of convergence of infinite sums or integrals.

Good codes. With the componentwise addition modulo 2 introduced above, V is a vector space over the two-element field GF(2) (scalar multiplication is trivial; the scalars are only 0 and 1, and we have $0v = 0$ and $1v = v$). The construction of the badly embeddable space is based on a suitable subspace $C \subseteq V$ as in the next lemma.

15.6.2 Lemma. *For every sufficiently large m divisible by 4 there exists a subset $C \subseteq V$ with the following properties:*

(i) *C is a vector subspace of V of dimension $\frac{1}{4}m$.*
(ii) *Every two distinct $u, v \in C$ differ in at least δm components, where $\delta > 0$ is a suitable small constant ($\delta = 0.01$ will do, for example). In other words, u and v have Hamming distance at least δm.*

We leave a proof to Exercise 2. The lemma comes from the theory of *error-correcting codes*. Mathematically speaking, the main goal of this theory is to construct, for given integer parameters m and d, a subset $C \subseteq V = \{0,1\}^m$ (called a *code* in this context) that is as large as possible but, at the same time, has *minimum distance* at least d, meaning that every two distinct $u, v \in C$ have Hamming distance at least d. (More precisely, this concerns codes over the two-element alphabet $\{0,1\}$, while coding theory also investigates codes over larger alphabets.) In many of the known constructions, C is a vector subspace of V, and then it is called a *linear code*. A code $C \subseteq \{0,1\}^m$ whose minimum distance is at least a fixed fraction of m and such that $\log|C|$ is also at least a fixed fraction of m is often called a *good code*. So Lemma 15.6.2 simply claims the existence of a good linear code (the constant $\frac{1}{4}$ in part (i) is rather arbitrary and, for our purposes, it could be replaced by any smaller positive constant).

Given C as in Lemma 15.6.2, let us put

$$W = C^{\perp} = \{u \in V\colon u \cdot v = 0 \text{ for all } v \in C\}.$$

This definition resembles the usual orthogonal complement in real vector spaces, *except* that $C \cap C^{\perp}$ may contain nonzero vectors (consider $m = 2$ and $C = C^{\perp} = \{(0,0),(1,1)\}$)! However, the familiar properties $(C^{\perp})^{\perp} = C$ and $\dim C + \dim C^{\perp} = m$ hold; see Exercise 3. From now on, we will talk only about W (and forget about C), and we will use the following properties:

15.6.3 Lemma.

(W1) *W is a vector subspace of V of dimension $\frac{3}{4}m$.*

(W2) *For every nonzero $u \in V$ with less than δm ones there exists $v \in W$ with $u \cdot v = 1$.*

Proof. Part (W1) is clear from $\dim C = \frac{m}{4}$ and $\dim C + \dim C^{\perp} = m$. As for (W2), if $u \in V$ satisfies $u \cdot v = 0$ for all $v \in W$, then $u \in W^{\perp} = (C^{\perp})^{\perp} = C$. Every nonzero vector in C has at least δm ones, and (W2) follows. □

The badly embeddable space. Let m be divisible by 4 and let $W \subset V$ satisfy (W1) and (W2). We are going to construct the metric space $(\bar{V}, \bar{\rho})$. First we define an equivalence relation $\approx$ on V by $u \approx v$ if $u - v \in W$. The equivalence class containing u thus has thus the form $u + W$, and we will denote it by $\bar{u}$.

The points of $\bar{V}$ are the equivalence classes: $\bar{V} = \{\bar{u}\colon u \in V\}$. The number of equivalence classes is $\bar{n} = |\bar{V}| = |V|/|W| = 2^{m/4}$.

The metric $\bar{\rho}$ on $\bar{V}$ is defined as the shortest-path metric of a suitable graph $\bar{G} = (\bar{V}, \bar{E})$, where $\{\bar{u}, \bar{v}\} \in \bar{E}$ if there is at least one edge of the Hamming cube C_m connecting a vertex of the equivalence classes $\bar{u}$ to a vertex of the equivalence class $\bar{v}$. More formally, $\bar{E} = \{\{\bar{u}, \bar{v}\}\colon \{u, v\} \in E\}$, where E is the edge set of C_m.

It may be useful to think of $\bar{\rho}$ as follows. Let us suppose that it takes unit time to travel between the endpoints of an edge of the Hamming cube. Then the Hamming distance of two vertices $u, v \in V$ is the minimum time required to travel from u to v along edges. Now we imagine that every two vertices u, u' in the same equivalence class are connected by a “hyperspace link” that can be traveled instantaneously, in time 0. Then $\bar{\rho}(\bar{u}, \bar{v})$ is the shortest time needed to travel from a vertex of $\bar{u}$ to a vertex of $\bar{v}$ using ordinary edges and hyperspace links as convenient.

The following observation will be useful later, and it may help digest the definition of $\bar{\rho}$.

15.6.4 Observation. *For every two equivalence classes $\bar{u}, \bar{v} \in \bar{V}$, one can travel from any vertex of $\bar{u}$ to any vertex of $\bar{v}$ in time $\bar{\rho}(\bar{u}, \bar{v})$ by first using at most one hyperspace link and then traveling only through ordinary edges.*

Sketch of proof. It suffices to note that if $\bar{u}, \bar{v}$ are equivalence classes and there is at least one edge of the Hamming cube connecting them, then *every* vertex of $\bar{u}$ is connected to some vertex of $\bar{v}$. Indeed, if $\{u, v\} \in E$ is an edge, i.e., u and v differ in exactly one coordinate, and if $u' \in \bar{u} = u + W$, then u' and $v' = v + u' - u \in v + W$ differ in exactly the same coordinate, and thus $\{u', v'\} \in E$ as well. □

We can state the main result of the section:

15.6.5 Theorem. *Every embedding of the metric space $(\bar{V}, \bar{\rho})$ into a Euclidean space has distortion at least $\Omega(m) = \Omega(\log \bar{n})$.*

Preparatory steps. We begin with realizing the already announced scheme of the proof. We consider the two sets of pairs, $\bar{E}$ as above and $\bar{F} = \binom{\bar{V}}{2}$. First we estimate $R_{\bar{E},\bar{F}}(\bar{\rho})$. Since for $\{\bar{u}, \bar{v}\} \in \bar{E}$ we have $\bar{\rho}(\bar{u}, \bar{v}) = 1$, we get $\mathrm{ave}_2(\bar{\rho}, \bar{E}) = 1$.

It is easy to see that a "typical" pair of vertices in V has Hamming distance at least a constant fraction of m. The next lemma shows that although adding the hyperspace links in the construction of $\bar{V}$ shortens the distances, a typical distance under $\bar{\rho}$ is still a constant fraction of m.

15.6.6 Lemma. *For at least half of pairs $\{\bar{u}, \bar{v}\} \in \bar{F}$ we have $\bar{\rho}(\bar{u}, \bar{v}) \geq \alpha m$, where α is a suitable positive constant.*

Proof. Let us fix an arbitrary $\bar{u} \in \bar{V}$; it suffices to show that only $o(\bar{n})$ classes $\bar{v} \in \bar{V}$ satisfy $\bar{\rho}(\bar{u}, \bar{v}) \leq k = \lfloor \alpha m \rfloor$. Let $U \subseteq V$ denote the union of all these classes $\bar{v}$. By Observation 15.6.4, every vertex of U can be reached from some vertex of the class $\bar{u}$ by traveling a path of length at most k in the Hamming cube. In the Hamming cube, the number of vertices at Hamming distance at most k from a fixed vertex is exactly $\sum_{i=0}^{k} \binom{m}{i}$, and thus $|U| \leq |W| \cdot \sum_{i=0}^{k} \binom{m}{i}$. Therefore, the number of $\bar{v}$ with $\bar{\rho}(\bar{u}, \bar{v}) \leq k$ is no more than $|U|/|W| \leq \sum_{i=0}^{k} \binom{m}{i}$. A standard estimate for the last sum is $(em/k)^k$, and a simple calculation shows that this is $o(\bar{n})$ for α sufficiently small. □

By the lemma we have $\mathrm{ave}_2(\bar{\rho}, \bar{F}) = \Omega(m)$, and thus $R_{\bar{E},\bar{F}}(\bar{\rho}) = \Omega(m)$.

Next, we turn to bounding $R_{\bar{E},\bar{F}}(\bar{\sigma})$ from above, for any (pseudo)metric $\bar{\sigma}$ induced by an Euclidean embedding $\bar{f}\colon \bar{V} \to \ell_2$. We need to prove that $\mathrm{ave}_2(\bar{F}, \bar{\sigma}) = O(\mathrm{ave}_2(\bar{E}, \bar{\sigma}))$, and as we saw in the previous sections, it is sufficient to consider only one-dimensional embeddings $\bar{f}\colon \bar{V} \to \mathbf{R}$. Thus we want to prove

$$|\bar{F}|^{-1} \sum_{\{\bar{u},\bar{v}\}\in\bar{F}} \left(\bar{f}(\bar{u}) - \bar{f}(\bar{v})\right)^2 \le O(1) \cdot |\bar{E}|^{-1} \sum_{\{\bar{u},\bar{v}\}\in\bar{E}} \left(\bar{f}(\bar{u}) - \bar{f}(\bar{v})\right)^2. \quad (15.7)$$

To do so, we first "pull back" from $\bar{V}$ to the original vertex set V of the cube. Given $\bar{f}\colon \bar{V} \to \mathbf{R}$, we define $f\colon V \to \mathbf{R}$ by $f(v) = \bar{f}(\bar{v})$ (so f is constant on every equivalence class). Then, with $F = \binom{V}{2}$,

$$\sum_{\{\bar{u},\bar{v}\}\in\bar{F}} \left(\bar{f}(\bar{u}) - \bar{f}(\bar{v})\right)^2 = |W|^{-2} \cdot \sum_{\{u,v\}\in F} (f(u) - f(v))^2$$

(we note that if u and v are in the same equivalence class, they contribute 0 to the sum). Observing that for every $\{\bar{u},\bar{v}\} \in \bar{E}$ the equivalence classes $\bar{u}$ and $\bar{v}$ are connected by exactly $|W|$ edges of E (arguing as in the proof of Observation 15.6.4), we get similarly

$$\sum_{\{\bar{u},\bar{v}\}\in\bar{E}} \left(\bar{f}(\bar{u}) - \bar{f}(\bar{v})\right)^2 = |W|^{-1} \cdot \sum_{\{u,v\}\in E} (f(u) - f(v))^2.$$

Using $|\bar{F}| = \binom{\bar{n}}{2} = \Theta(2^{m/2})$ and $|\bar{E}| = |E|/|W| = \Theta(m2^{m/4})$, we can see that for proving (15.7) we need

$$\sum_{\{u,v\}\in F} (f(u) - f(v))^2 \le O(2^m/m) \cdot \sum_{\{u,v\}\in E} (f(u) - f(v))^2 \qquad (15.8)$$

for every $f\colon V \to \mathbf{R}$ satisfying $f(u) = f(v)$ whenever $u - v \in W$. Moreover, since increasing all values of f by the same number doesn't change either side of (15.8), we may assume that $\sum_{v\in V} f(v) = 0$, which will be useful (we already met this trick in Section 15.5).

A Fourier-analytic piece. We will express both sides in (15.8) using the Fourier coefficients of f, beginning with the right-hand side. The next two lemmas are standard tools in harmonic analysis.

15.6.7 Lemma. *For every function $f\colon V \to \mathbf{R}$ we have*

$$\sum_{\{u,v\}\in E} (f(u) - f(v))^2 = 2 \cdot \sum_{u\in V} \|u\|_1 \cdot \hat{f}(u)^2,$$

where $\|u\|_1$ denotes the number of 1's in the zero-one vector u.

We postpone the proof and continue with the proof of (15.8). The next lemma shows how the properties of W are reflected in the Fourier coefficients of f.

15.6.8 Lemma. *Let $f\colon V \to \mathbf{R}$ be such that $f(u) = f(v)$ whenever $u - v \in W$, where W satisfies (W1), (W2). Then $\hat{f}(u) = 0$ for every $u \in V$ with $0 < \|u\|_1 < \delta m$. If we also assume $\sum_{v\in V} f(v) = 0$, then $\hat{f}(0) = 0$ as well.*

We again postpone the proof.

By combining Lemma 15.6.8 and Lemma 15.6.7, we have

$$\begin{aligned}\sum_{\{u,v\}\in E}(f(u)-f(v))^2 &= 2\cdot\sum_{u\in V}\|u\|_1\cdot\hat f(u)^2\\ &= 2\cdot\sum_{u\in V,\|u\|_1\geq\delta m}\|u\|_1\hat f(u)^2\\ &\geq 2\delta m\sum_{u\in V}\hat f(u)^2.\end{aligned}$$

We now turn to the left-hand side of (15.8). By calculation we already did in Section 15.5 (see (15.5)), we obtain

$$\sum_{\{u,v\}\in F}(f(u)-f(v))^2=|V|\cdot\sum_{v\in V}f(v)^2,$$

and by Parseval's equality the right-hand side is $|V|\cdot\sum_{u\in V}\hat f(u)^2$. Altogether

$$\sum_{\{u,v\}\in F}(f(u)-f(v))^2=|V|\cdot\sum_{u\in V}\hat f(u)^2\leq\frac{|V|}{2\delta m}\sum_{\{u,v\}\in E}(f(u)-f(v))^2,$$

which provides the desired inequality (15.8). It remains to prove the lemmas, which is a relatively straightforward manipulation of identities.

Proof of Lemma 15.6.8. First we note that $\hat f(0)=2^{-m}\sum_{v\in V}f(v)$, so $\hat f(0)=0$ is immediate.

Next, we consider $u\in V$ with $0\neq\|u\|_1<\delta m$, and we fix $v_0\in W$ with $u\cdot v_0=1$, according to (W2). Since $v\mapsto v+v_0$ defines a bijection $V\to V$, we have

$$\hat f(u)=2^{-m}\sum_{v\in V}(-1)^{u\cdot v}f(v)=2^{-m}\sum_{v\in V}(-1)^{u\cdot(v+v_0)}f(v+v_0).$$

Since $v_0\in W$, $v+v_0$ is equivalent to v and so $f(v+v_0)=f(v)$. Further $u\cdot(v+v_0)=u\cdot v+u\cdot v_0=u\cdot v+1$, and so we arrive at

$$\hat f(u)=2^{-m}\sum_{v\in V}(-1)^{u\cdot v+1}f(v)=-\hat f(u).$$

So $\hat f(u)=0$ indeed. □

Proof of Lemma 15.6.7. Let $e_i\in V$ denote the ith vector of the standard basis, with 1 at position i and 0's elsewhere. We set $g_i(v)=f(v+e_i)-f(v)$ (this is something like a "derivative" of f according to the ith variable). Every edge in E has the form $\{v,v+e_i\}$ for some v and some i, and so

$$\sum_{\{u,v\}\in E}(f(u)-f(v))^2 = \frac{1}{2}\sum_{i=1}^{m}\sum_{v\in V} g_i(v)^2 = \frac{1}{2}\sum_{i=1}^{m}\sum_{u\in V}\hat{g}_i(u)^2,$$

the last equality being Parseval's. We compute

$$\begin{aligned}\hat{g}_i(u) &= 2^{-m}\sum_{v\in V}(-1)^{u\cdot v}(f(v+e_i)-f(v))\\ &= 2^{-m}\left(\sum_{v'\in V}(-1)^{u\cdot(v'-e_i)}f(v')\right)-\hat{f}(u)\\ &= (-1)^{u\cdot e_i}\hat{f}(u)-\hat{f}(u).\end{aligned}$$

Hence $\hat{g}_i(u)^2$ equals $4\hat{f}(u)^2$ if $u_i = 1$ and 0 otherwise, and

$$\frac{1}{2}\sum_{i=1}^{m}\sum_{u\in V}\hat{g}_i(u)^2 = 2\sum_{u\in V}\|u\|_1\cdot\hat{f}(u)^2.$$

Lemma 15.6.7, as well as Theorem 15.6.5, are proved. □

Bibliography and remarks. Among many textbooks of harmonic analysis we mention Körner [Kör89] or Dym and McKean [DM85], and a standard reference for error-correcting codes is Van Lint [vL99].

The construction of the metric space $(\bar{V},\bar{\rho})$ from the Hamming cube using the equivalence $\approx$ is an instance of a generally useful construction of a *quotient space* from a given metric space using an equivalence with finitely many classes; see, e.g., Gromov [Gro98] (we again add zero-length links between every two equivalent points and consider the resulting shortest-path metric). However, we note that for an arbitrary equivalence an analogue of Observation 15.6.4 no longer holds.

The material of this section is from Khot and Naor [KN05]. This paper also cites several earlier applications of harmonic analysis for distortion bounds. An immediate predecessor of it is a fundamental work of Khot and Vishnoi [KV05], which we will discuss in Section 15.9.

Khot and Naor solved several open problems in low-distortion embeddings. Perhaps most notably, they showed that the metric space $\{0,1\}^n$ with the *edit distance* needs $\Omega((\log n)^{1/2-o(1)})$ distortion for embedding into ℓ_1, where the edit distance of two n-bit strings v, w is the smallest number of edit operations (insertions or deletions of bits) required for converting v into w. Krauthgamer and Rabani [KR06] improved the lower bound to $\Omega(\log n)$, with a beautifully simple argument. Ostrovsky and Rabani [OR05] proved an upper bound of $2^{O(\sqrt{\log n \log\log n})}$. The need for such embeddings has strong algorithmic motivation, since a low-distortion embedding into ℓ_1 would allow for very fast approximate database queries ("Does the database contain a string very similar to a given query string?"; this is a basic problem in web searching, computational biology, etc.), or even a fast approximation of the edit distance of two given strings, whose exact computation is not easy.

In a way similar to the proof for expanders from the previous section, the proof in this section can easily be modified to show that there is no *squared* Euclidean metric that approximates the constructed space with distortion $o(\log n)$, and in particular, that any embedding into ℓ_1 requires distortion $\Omega(\log n)$.

Exercises

1. (a) Let $u_1 \neq u_2 \in V$. Prove that $\sum_{v\in V}(-1)^{u_1\cdot v}(-1)^{u_2\cdot v} = 0$. [2]
(b) Prove Fact 15.6.1 by direct calculation, substituting for $\hat{f}(u)$ on the right-hand side from the definition, expanding, and using (a). [2]
(c) Go through the following "more scientific" presentation of the proof. Let $\mathcal{F}_m$ denote the real vector space of all functions $f\colon V \to \mathbf{R}$, where addition and scalar multiplication are defined in the natural way, by $(f+g)(v) = f(v)+g(v)$ and $(\alpha f)(v) = \alpha \cdot f(v)$. Check that $\langle f, g\rangle = \sum_{v\in V} f(v)g(v)$ defines a real scalar product on $\mathcal{F}_m$, and that $(e_v\colon v \in V)$ form an orthonormal basis, where $e_v(v) = 1$ and $e_v(u) = 0$ for $u \neq v$. Verify that $(\varphi_u\colon u \in V)$ is another orthonormal basis, where $\varphi_u(v) = 2^{-m/2}(-1)^{u\cdot v}$, and that for every $f \in \mathcal{F}_m$, $f = \sum_{u\in V} \hat{f}(u)\cdot\varphi_u$ is an expression of f in this basis. [3]
2. (a) Let $C \subseteq V$ be a vector subspace. Check that if every nonzero $v \in C$ has at least d ones, then the Hamming distance of every two distinct vectors in C is at least d. [1]
(b) Prove Lemma 15.6.2 by induction as follows. Let m be given and let $d = \lfloor \delta m \rfloor$, where the constant $\delta > 0$ is chosen as small as convenient for the proof. Let us say that a k-tuple $(v_1, \ldots, v_k)$ of vectors of V is *good* if $v_1, \ldots, v_k$ are linearly independent and every nonzero vector in their linear span has at least d ones. Given any good k-tuple $(v_1, \ldots, v_k)$, $k < \frac{m}{4}$, estimate the number of $v \in V$ such that $(v_1, \ldots, v_k, v)$ is *not* a good $(k{+}1)$-tuple, and conclude that every good k-tuple can be extended to a good $(k{+}1)$-tuple. How does this imply the lemma? [4]
3. (a) Let A be a matrix with m columns and of rank r, over any field $\mathbf{K}$. Recall (or look up) a proof that the subspace of $\mathbf{K}^m$ consisting of all solutions to $Ax = 0$ has dimension $m - r$. [1]
(b) For a set $U \subseteq V = \{0,1\}^m$ we define $U^\perp = \{v \in V\colon u \cdot v = 0 \text{ for all } v \in U\}$, where $u \cdot v$ is the inner product as in the text. Show that if U is a vector subspace of V, then $\dim U + \dim U^\perp = m$. [1]
(c) Show that if U is a vector subspace of V, then $(U^\perp)^\perp = U$. [2]

⇑ NEW ⇑

15.7 Upper Bounds for ℓ_∞-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 ℓ_∞, which are somewhat simpler. We derive almost tight upper bounds.

Let (V,ρ) 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 that $|f_i(u)-f_i(v)| \le \rho(u,v)$ 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 $i=i(u,v)$ that "takes care" of the pair: $|f_i(u)-f_i(v)| \ge \frac{1}{D}\rho(u,v)$.

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, $f_i(u) = \rho(u,A_i) = \min_{a\in A_i} \rho(u,a)$. By the triangle inequality, we have $|\rho(u,A_i)-\rho(v,A_i)| \le \rho(u,v)$ 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 ℓ_∞.

15.7.1 Proposition (Fréchet's embedding). *Let (V,ρ) 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 ℓ_∞^n are indexed by the points of V, and the vth 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 $u \in V$:

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

□

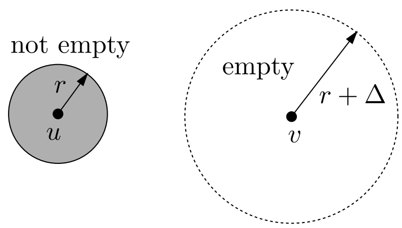
The dimension of the image in this embedding can be reduced a little; for example, we can choose some $v_0 \in V$ and remove the coordinate corresponding to v_0, and the above proof still works. To reduce the dimension significantly, though, we have to pay the price of distortion. For example, from Corollary 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:

15.7.2 Theorem. *Let $D = 2q-1 \geq 3$ be an odd integer and let (V, ρ) be an n-point metric space. Then there is a D-embedding of V into ℓ_∞^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.

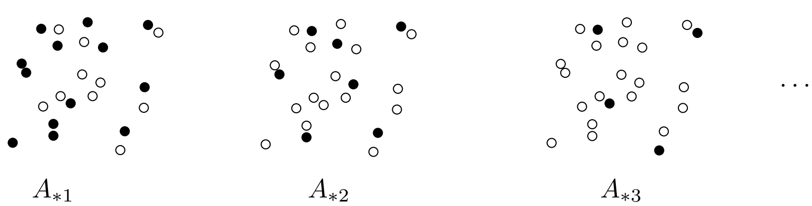
Let us consider two points $u, v \in V$. What are the sets A such that $|\rho(u, A) - \rho(v, A)| \geq \Delta$, for a given real $\Delta > 0$? For some $r \geq 0$, they must intersect the closed r-ball around u and avoid the open $(r+\Delta)$-ball around v; schematically,



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

In the favorable situation where the closed r-ball around u does not contain many fewer points of V than the open $(r+\Delta)$-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 ℓ_∞^{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 ℓ_∞^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.

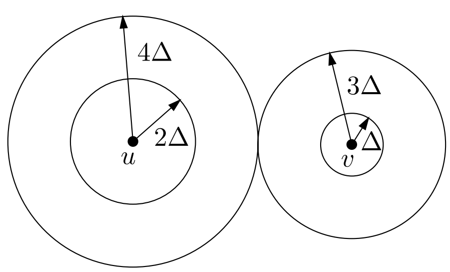
15.7.3 Lemma. *Let u, v be two distinct points of V. Then there exists an index $j \in \{1, 2, \ldots, q\}$ such that if the set A_{ij} is chosen randomly as above, then the probability of the event*

$$|\rho(u, A_{ij}) - \rho(v, A_{ij})| \geq \tfrac{1}{D}\, \rho(u, v) \tag{15.9}$$

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 (15.9) 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 (15.9) 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. □

Proof of Lemma 15.7.3. Set $\Delta = \frac{1}{D}\, \rho(u, v)$. Let $B_0 = \{u\}$, let B_1 be the (closed) Δ-ball around v, let B_2 be the (closed) 2Δ-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 $\rho(u, v)$; that is, 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 indices j and t such that

$$n_t \geq n^{(j-1)/q} \quad \text{and} \quad n_{t+1} \leq n^{j/q}. \tag{15.10}$$

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 (15.10) holds for the j such that I_j contains n_t. On the other hand, if $1 = n_0 \le n_1 \le \ldots \le n_q \le n$, 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 (15.10) holds for this j and t as well.

In this way, we have selected the index j whose existence is claimed in the lemma, and the corresponding index t. 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 (15.9). Since B_t and the interior of B_{t+1} are disjoint, the events E_1 and E_2 are independent.

We calculate

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

Using (15.10), we have $p_j n_t \ge p_j n^{(j-1)/q} = p_j p^{-j+1} = \min(\frac{1}{2}, p^j)p^{-j+1} \ge \min(\frac{1}{2}, p)$. For $p \ge \frac{1}{2}$, we get $\text{Prob}[E_1] \ge 1 - e^{-1/2} > \frac{1}{3} \ge \frac{p}{3}$, while for $p < \frac{1}{2}$, we have $\text{Prob}[E_1] \ge 1 - e^{-p}$, 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,

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

(since $p_j \le \frac{1}{2}$). Thus $\text{Prob}[E_1 \cap E_2] \ge \frac{p}{12}$, 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 15.8.1 explained in the subsequent section. Theorem 15.7.2 is from [Mat96b].

Exercises

1. (a) Find an isometric embedding of ℓ_1^d into $\ell_\infty^{2^d}$. [3]
   (b) Explain how an embedding as in (a) can be used to compute the diameter of an n-point set in ℓ_1^d in time $O(d2^d n)$. [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 ℓ_∞^m. [2]
   (Using results on approximation of convex bodies by polytopes, this yields useful approximate embeddings of arbitrary norms into ℓ_∞^k.)
3. Deduce from Theorem 15.7.2 that every n-point metric space can be D-embedded into ℓ_2^k with $D = O(\log^2 n)$ and $k = O(\log^2 n)$. [2]

15.8 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.

15.8.1 Theorem (Bourgain's embedding into $\boldsymbol{\ell_2}$). *Every n-point metric space (V, ρ) 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 ℓ_∞^d in the proof of Theorem 15.7.2. The coordinates in ℓ_2^d are given by distances to suitable subsets. The situation is slightly more complicated than before: For embedding into ℓ_∞^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 15.7.3.

15.8.2 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, and we set $r_q = \frac{1}{4}\rho(u,v)$. 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, $B(x, r) = \{y \in V\colon \rho(x,y) \leq r\}$, and we set $r_j = \min(r_j^*, r_q)$. We are going to show that the claim of the lemma holds with $\Delta_j = r_j - r_{j-1}$.

We may assume that $r_{j-1} < r_q$, for otherwise, we have $\Delta_j = 0$ and the claim holds automatically. We note that $r_{j-1} < r_q$ implies that both of the balls $B(u, r_{j-1})$ and $B(v, r_{j-1})$ have at least 2^{j-1} points.

So let us fix $j \in \{1, 2, \ldots, q\}$ with $r_{j-1} < r_q$, 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 $B^\circ(x, r) = \{y \in V\colon \rho(x,y) < r\}$ denotes the open ball (this holds for $j = q$, too, because $|V| \leq 2^q$). We 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} (as was noted above) 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 15.7.3 with $p = \frac{1}{2}$. □

In the subsequent proof of Theorem 15.8.1 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 ν induced by a mapping $\varphi: V \to \mathbf{R}$, that is, given by $\nu(u,v) = |\varphi(u) - \varphi(v)|$. For each $A \subseteq V$, let ν_A be the line pseudometric corresponding to the mapping $v \mapsto \rho(v,A)$. As we have noted, each ν_A is *dominated* by ρ, 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 ρ on V can be approximated by a convex combination of line pseudometrics, each of them dominated by ρ, then a good embedding of (V,ρ) into ℓ_2 exists.

15.8.3 Lemma. *Let (V,ρ) 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,ρ) can be D-embedded into ℓ_2^N.

Proof. Let $\varphi_i: V \to \mathbf{R}$ be a mapping inducing the line pseudometric ν_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 ν_i are dominated by ρ and $\sum \alpha_i = 1$. On the other hand,

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

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

Proof of Theorem 15.8.1. As was remarked above, each of the line pseudometrics ν_A corresponding to the mapping $v \mapsto \rho(v,A)$ is dominated by ρ. It remains to observe that Lemma 15.8.2 provides a convex combination of these line pseudometrics that is bounded from below by $\frac{1}{48q}\cdot\rho$. The coefficient of each ν_A in this convex combination is given by the probability of A appearing as one of the sets A_j in Lemma 15.8.2. 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 15.8.2 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 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 15.8.3 now gives embeddability into ℓ_2 with distortion at most $48q$. Theorem 15.8.1 is proved. □

Remarks. Almost the same proof with a slight modification of Lemma 15.8.3 shows that for each $p \in [1, \infty)$, every n-point metric space can be embedded into ℓ_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 Johnson–Lindenstrauss flattening lemma, but for other ℓ_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 ℓ_1^d with $d = O(\log^2 n)$. By inspecting the proof, it is not difficult to give a randomized algorithm that computes such an embedding in polynomial expected time. We show a neat algorithmic application to a graph-theoretic 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 $n = |V|$ (the best known approximation guarantee for a polynomial-time algorithm is better, currently $O(\sqrt{\log n}\,)$). 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 reformulate the problem equivalently using cut pseudometrics. A *cut pseudometric* on V is a pseudometric τ 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 $\tau(E)/\tau(F)$, where τ is the corresponding cut pseudometric and $\tau(E) = \sum_{\{u,v\}\in E} \tau(u, v)$. Therefore, we would like to minimize the ratio $R_1(\tau) = \tau(E)/\tau(F)$ over all cut pseudometrics τ.

In the first step of the algorithm we relax the problem, and we find a pseudometric, not necessarily a cut one, minimizing the ratio $R_1(\rho) = \rho(E)/\rho(F)$. 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 ρ_0 minimizing $R_1(\rho)$, we find a D-embedding f of (V, ρ_0) into some ℓ_1^d with $D = O(\log n)$. If σ_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 σ_0 is an ℓ_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 τ_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 τ_0 is a cut pseudometric with the smallest possible $R_1(\tau_0)$. Therefore, the cut corresponding to this τ_i has density at most $O(\log n)$ times larger than the sparsest possible cut.

Bibliography and remarks. Theorem 15.8.1 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]. The improved $O(\sqrt{\log n})$ approximation is due to Arora, Rao, and Vazirani [ARV04].

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 $\mathcal{T}$ the class of all trees. Tree metrics generally behave much better than arbitrary metrics, but for embedding problems they are far from trivial.

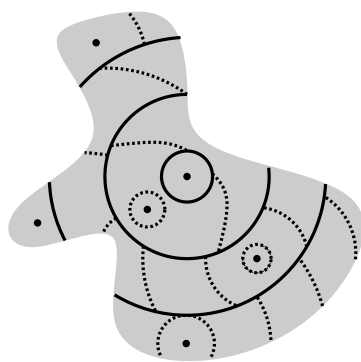
Bourgain [Bou86] proved, using martingales, a surprising lower bound for embedding tree metrics into ℓ_2: A tree metric on n points requires distortion $\Omega(\sqrt{\log\log n}\,)$ in the worst case. His example is the complete binary tree with unit edge lengths, and for that example, he also constructed an embedding with $O(\sqrt{\log\log n}\,)$ distortion. For embedding the complete binary tree into ℓ_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; that is, every n-point tree metric can be embedded into ℓ_p with distortion $O((\log\log n)^{\min(1/2,1/p)})$. 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 ℓ_2, was found by Linial and Saks [LS03] (Exercise 6).

In the notes to Section 15.3 we mentioned that general n-point metric spaces require worst-case distortion $\Omega(n^{1/\lfloor(d+1)/2\rfloor})$ for embedding into ℓ_2^d, $d \geq 2$ fixed. Gupta [Gup00] proved that for n-point tree metrics, $O(n^{1/(d-1)})$-embeddings into ℓ_2^d are possible. The best known lower bound is $\Omega(n^{1/d})$, from a straightforward volume argument. Babilon, Matoušek, Maxová, and Valtr [BMMV02] showed that every n-vertex tree with unit-length edges can be $O(\sqrt{n}\,)$-embedded into ℓ_2^2.

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 (the shorter but potentially confusing term *planar metric* is used in the literature). Rao [Rao99] proved that every n-point planar-graph metric can be embedded into ℓ_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 ℓ_2. For a matching lower bound, valid already for the class $Excl(K_4)$ (series-parallel graphs), and consequently for planar-graph metrics, see Exercise 15.4.2 or Exercise 15.9.1.

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 ρ be the corresponding graph metric (with all edges having unit length), and let Δ 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 $B_1 = \{v \in V(G) \colon \rho(v, v_0) \equiv r \,(\mathrm{mod}\, \Delta)\}$. 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; Δ 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 solid 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 $G \in \mathit{Excl}(H)$.

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

(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 Δ, 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 ℓ_2 in several steps. First we consider Δ 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 ρ.

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 Δ fixed and consider $\nu_{B,\sigma}$ for all possible B and σ. Letting $\alpha_{B,\sigma}$ be the probability that a particular pair (B, σ) 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 15.8.3, this yields a 1-Lipschitz embedding $f_\Delta: V(G) \to \ell_2^N$ (for some N) that 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, σ) in the embedding; it is easy to show that *const* $\cdot \log n$ independent random B and σ will do.)

To construct the final embedding $f: V(G) \to \ell_2$, we let $f(v)$ be the concatenation of the vectors f_Δ 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 ℓ_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 conclude only 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 Δ run through all the relevant powers of 2 (that is, such that $C\Delta < \rho(x, y) \leq 2C\Delta$ for some $x \neq y$), but for producing the decomposition for a particular Δ, we use a modified graph G_Δ obtained from G by contracting all edges shorter than $\frac{\Delta}{2n}$. In this way, we can have many more than $\log n$ values of Δ, 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] conjectured that for any fixed graph H, *Excl*(H)-metrics might be $O(1)$-embeddable into ℓ_1 (the constant depending on H). They proved the conjecture for $H = K_4$ and for $H = K_{2,3}$ (outerplanar graphs). Chekuri, Gupta, Newman, Rabinovich, and Sinclair [CGN$^+$03] established the conjecture for k-outerplanar graphs (for every fixed k), which are, roughly speaking, graphs admitting a planar drawing with no $k{+}1$ disjoint properly nested cycles, a canonical example being the $k{\times}n$ grid.

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 that 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 ℓ_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, ρ), 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 $D \cdot \text{Evol}(f(S))^{1/(k-1)} \geq \text{Vol}(S)^{1/(k-1)}$. For D small, this means that the image of any k-tuple spans nearly as large a volume as it possibly can for a 1-Lipschitz map. (Note, for example, that an isometric embedding of a path into ℓ_2 is *not* volume-respecting.)

Feige showed that $\text{Vol}(S)$ can be approximated quite well by an intrinsic parameter of the metric space (not referring to embeddings), namely, by the *tree volume* $\text{Tvol}(S)$, which equals the products of the edge lengths in a minimum spanning tree on S (with respect to the metric on S). Namely, $\text{Vol}(S) \leq \frac{1}{(k-1)!}\,\text{Tvol}(S) \leq 2^{(k-2)/2}\,\text{Vol}(S)$. He proved that for any n-point metric space and all $k \geq 2$, the embedding as in the proof of Theorem 15.8.1 is $(k, O(\log n + \sqrt{k \log n \log k}\,))$ volume-respecting. Later Krauthgamer, Lee, Mendel, and Naor [KLMN04] established the existence of embeddings that are $(k, O(\log n))$ volume-respecting for all $k = 1, 2, \ldots, n$, which is optimal.

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 $\varphi\colon V(G) \to \{1, 2, \ldots, n\}$, of $\max\{|\varphi(u) - \varphi(v)|\colon \{u, v\} \in E(G)\}$ (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 $O((\log n)^{\text{const}})$. The algorithm has two main steps: First, embed the graph (as a metric space) into ℓ_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 λ be a random line in ℓ_2^m and let $\psi(v)$ denote the orthogonal projection of $f(v)$ on λ. This $\psi\colon V(G) \to \lambda$ is almost surely injective, and so it provides a linear ordering of the vertices, that is, a bijective map $\varphi\colon V(G) \to \{1, 2, \ldots, n\}$, and this is used for estimating the bandwidth.

To indicate the analysis, we need the notion of *local density* of the graph G: $\text{ld}(G) = \max\{|B(v, r)|/r\colon v \in V(G),\ r = 1, 2, \ldots, n\}$, where $B(v, r)$ are all vertices at distance at most r from v. It is not hard to see that $\text{ld}(G)$ is a lower bound for the bandwidth, and Feige's analysis shows that $O(\text{ld}(G)(\log n)^{\text{const}})$ 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 λ are close; concretely, $|\psi(u) - \psi(v)| \leq \Delta = O(\sqrt{(\log n)/m}\,)$. 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 ψ, then some interval of length Δ on λ contains the images of

b vertices. Call a k-tuple $S \subset V(G)$ *squeezed* if $\psi(S)$ lies in an interval of length Δ. 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 a 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 ℓ_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. ⇓ NEW ⇓

Probabilistic approximation by dominating trees. As we have seen, in Bourgain's method, for a given metric ρ one constructs a convex combination $\sum \alpha_i \nu_i \geq \frac{1}{D}\,\rho$, where ν_i are line pseudometrics dominated by ρ. An interesting "dual" result is the possibility of approximating every ρ by a convex combination $\sum_{i=1}^{N} \alpha_i \tau_i$, where this time the inequalities go in the opposite direction: $\tau_i \geq \rho$ and $\sum \alpha_i \tau_i \leq D\rho$, with $D = O(\log n)$. The τ_i are not line metrics (and in general they cannot be), but they are tree metrics. (It is important to use many trees, since it is impossible to approximate a general metric by a *single* tree metric with any reasonable distortion—consider the shortest-path metric of a cycle of length n.) Since tree metrics embed isometrically into ℓ_1, the result yields an alternative proof of $O(\log n)$–embeddability of all n-point metric spaces into ℓ_1. This also implies that we must have $D = \Omega(\log n)$ in the above result.

The approximation of ρ by tree metrics as above is usually called a *probabilistic approximation by dominating trees*, which refers to the following alternative view. The coefficients $\alpha_1, \ldots, \alpha_N$ and the tree metrics $\tau_1, \ldots, \tau_N$ specify a probability distribution on tree metrics, where τ_i is chosen with probability α_i. Then the condition $\sum \alpha_i \tau_i \leq D\rho$ translates as follows: If we pick a tree metric τ at random according to this distribution, then $\mathbf{E}[\tau(x,y)] \leq D\rho(x,y)$ for every x, y (and $\tau(x,y) \geq \rho(x,y)$ always by the condition $\tau_i \geq \rho$). This is conveniently used in approximation algorithms: If we want to solve some optimization problem for a given metric ρ, we can compute the τ_i and α_i and solve the problem for the random tree metric τ, which is usually much easier than solving it for a general metric. Often it can be shown that the expected value of the optimal solution for τ is not very far from the optimal value for the original metric ρ.

The first result about probabilistic approximation by dominating trees, with $D = 2^{O(\sqrt{\log n \log\log n})}$, is due to Alon, Karp, Peleg, and West [AKPW95]. A fundamental progress was made by Bartal [Bar96], who obtained $D = O(\log^2 n)$ (improved to $O(\log n \log\log n)$ in a later paper), and documented the significance of the result by many impressive applications. This line of research was crowned by Fakcharoenphol, Talwar, and Rao [FRT03], who obtained the optimal bound $D = O(\log n)$, with a clean few-page proof.

The construction of Alon et al. [AKPW95] assumes that the input metric is given as the shortest-path metric of a graph G (with weighted edges), and the resulting tree metrics τ_i are induced by *spanning trees* of G. The subsequent constructions [Bar96], [FRT03] don't share this property, but Elkin, Emek, Spielman, and Teng [EEST05] later achieved $D = O((\log n \log\log n)^2)$ using spanning trees.

Bartal [Bar96] also introduced the useful notion of a *k-hierarchically well-separated tree metric*, where $k \geq 1$ is a real parameter. Here we present a newer, and apparently more convenient, definition by Bartal, Bollobás, and Mendel [BBM01]. A 1-hierarchically well-separated tree metric is the same as an *ultrametric*, that is, a metric that can be represented as the shortest-path metric on the leaves of a rooted tree T (with weighted edges) such that all leaves have the same distance from the root. For $k > 1$, we moreover require that $h(v) \leq h(u)/k$ whenever v is a child of u in T, where $h(v)$ denotes the distance of v to the nearest leaves. In the result of [FRT03] cited above, one can get the τ_i k-hierarchically well-separated for every prescribed $k \geq 1$, with the constant in the bound $D = O(\log n)$ depending on k (the same holds for Bartal's earlier results).

We sketch the construction of Fakcharoenphol et al. [FRT03]. We may suppose that the given n-point metric space (V, ρ) has all distances between 1 and 2^m. We describe a randomized algorithm for generating a random tree metric τ with $\tau \geq \rho$ always and $\mathbf{E}[\tau(x,y)] \leq O(\log n \cdot \rho(x,y))$ for every x, y. The α_i and τ_i specifying the probability distribution are thus given implicitly (and actually, the procedure as described may output infinitely many distinct τ_i, but easy modifications would lead to a finite collection). We first generate a sequence $(\mathcal{P}_m, \mathcal{P}_{m-1}, \ldots, \mathcal{P}_0)$, where each $\mathcal{P}_i$ is a partition of V, and $\mathcal{P}_i$ refines $\mathcal{P}_{i+1}$; that is, each set in $\mathcal{P}_{i+1}$ is a disjoint union of some sets of $\mathcal{P}_i$. We begin with $\mathcal{P}_m = \{V\}$, and proceed to $\mathcal{P}_{m-1}, \mathcal{P}_{m-2}, \ldots$, finishing with $\mathcal{P}_0$ that consists of n singleton sets. The diameter of each set in $\mathcal{P}_i$ is at most 2^{i+1}. Having constructed these $\mathcal{P}_i$, we arrange them into a rooted tree in a natural way. The vertices of the tree have the form (S, i), $S \in \mathcal{P}_i$, $i = 0, 1, \ldots, m$. The root is (V, m), and the children of a vertex $(S, i{+}1)$, $S \in \mathcal{P}_{i+1}$, are (T, i) with $T \in \mathcal{P}_i$ and $T \subseteq S$. The edge connecting $(S, i{+}1)$ to (T, i) has length 2^{i+1}, and finally, the metric τ

is the shortest-path metric induced by this tree on the set of its leaves (the leaves correspond to points of V). It is easily seen that $\tau \geq \rho$ and that τ is 2-hierarchically well-separated.

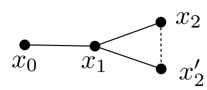
The construction of the $\mathcal{P}_i$ starts with arranging the points of V in a random order $(v_1, v_2, \ldots, v_n)$, and choosing a real number $\beta \in [1, 2]$ uniformly at random. Supposing that $\mathcal{P}_{i+1}$ has already been constructed, we set $\beta_i = \beta 2^{i-1}$, and for every $S \in \mathcal{P}_{i+1}$ we generate a partition of S into disjoint subsets $S_1, \ldots, S_n$ (don't get confused by the n; most of the S_ℓ are going to be empty!). Namely, for $\ell = 1, 2, \ldots, n$, we set $S_\ell = \{v \in S \setminus (S_1 \cup \cdots \cup S_{\ell-1})\colon \rho(v, v_i) \leq \beta_i\}$. That is, the first vertex v_1 in the random order, the "king," comes to S and he can take all vertices that are closer to him than β_i. Then, after the king has been satisfied, the "prime minister" v_2 comes, and he is allowed to grab from the leftovers all that is in his β_i-neighborhood. Then v_3 comes, and so on until v_n. The partition $\mathcal{P}_i$ consists of all nonempty S_ℓ thus generated for all $S \in \mathcal{P}_{i+1}$ and all ℓ. The proof of $\mathbf{E}[\tau(x, y)] \leq O(\log n \cdot \rho(x, y))$ is a simple but ingenious probabilistic analysis, for which we refer to [FRT03]. ⇑ NEW ⇑

Exercises

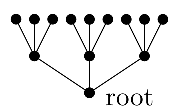
1. (Embedding into ℓ_p) Prove that under the assumptions of Lemma 15.8.3, the metric space (V, ρ) can be D-embedded into ℓ_p^N, $1 \leq p \leq \infty$, with distortion at most D. (You may want to start with the rather easy cases $p = 1$ and $p = \infty$, and use Hölder's inequality for an arbitrary p.) [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. [3]
   (b) Modify the proof of Theorem 15.8.1 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 15.8.2 and 15.8.3, prove that with a positive probability, the embedding $f\colon V \to \ell_2^{qm}$ given by $f(v)_{ij} = \rho(v, A_{ij})$ has distortion $O(\log n)$. [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
$$\frac{\alpha_1 a_1 + \alpha_2 a_2 + \cdots + \alpha_n a_n}{\alpha_1 b_1 + \alpha_2 b_2 + \cdots + \alpha_n b_n} \geq \min\{\frac{a_1}{b_1}, \frac{a_2}{b_2}, \ldots, \frac{a_n}{b_n}\}.$$
   [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. [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 ℓ_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 a dotted line).



Then $\|x_2 - x_2'\| \le \varepsilon$; that is, the remaining distance must be very short. [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. [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 ℓ_2. [4]

6. (Another lower bound for embedding trees into ℓ_2)
(a) Let $x_0, x_1, \ldots, x_n$ be arbitrary points in a Euclidean space (we think of them as images of the vertices of a path of length n under some embedding). Let $\Gamma = \{(a, a+2^k, a+2^{k+1})\colon a = 0, 1, 2, \ldots, a+2^{k+1} \le n, k = 0, 1, 2 \ldots\}$. Prove that

$$\sum_{(a,b,c)\in\Gamma} \frac{\|x_a - 2x_b + x_c\|^2}{(c-a)^2} \le \sum_{a=0}^{n-1} \|x_a - x_{a+1}\|^2;$$

this shows that an average triple (x_a, x_b, x_c) is "straight" (and provides an alternative solution to Exercise 4 for $Z = \ell_2$). [3]

(b) Prove that the complete binary tree $T_{2,m}$ requires $\Omega(\sqrt{\log m})$ distortion for embedding into ℓ_2. Consider a nonexpanding embedding $f\colon V(T_{2,m}) \to \ell_2$ and sum the inequalities as in (a) over all images of the root-to-leaf paths. [4]

7. (Bourgain's embedding of complete binary trees into ℓ_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 $O(\sqrt{\log m})$. [4]
8. Prove that any finite tree metric can be isometrically embedded into ℓ_1. [3]
9. (Low-dimensional embedding of trees)
(a) Let T be a tree (in the graph-theoretic sense) on $n \geq 3$ vertices. Prove that there exist subtrees T_1 and T_2 of T that share a single vertex and no edge and together cover T, such that $\max(|V(T_1)|, |V(T_2)|) \leq 1+\frac{2}{3}n$. [3]
(b) Using (a), prove that every tree metric space with n points can be isometrically embedded into ℓ_∞^d with $d = O(\log n)$. [4]
This result is from [LLR95].

15.9 Approximate Embeddings 2002–2005

⇓ NEW ⇓

The progress in the field considered in the present chapter has been amazing since the first edition of the book. Mainly by the efforts of several young researchers, many problem that were open and considered difficult in 2002 have been solved in the next few years. The development is vividly documented in the problem collection [Mat05], which started at a workshop in 2002 and has been continuously updated with new problems and solutions.

Few of the new results are mentioned in updated remarks to the previous sections or in the newly written Section 15.6. Here we outline some more, in a format similar to the bibliographic parts of the other sections.

Starring: $\boldsymbol{\ell_1}$ metrics. Most of the main results from the period 2002–2005 have a unifying theme: a key role of ℓ_1. There are several sources of its prominence. First, from a mathematical point of view, ℓ_1 can be thought of as a "frontier of wilderness." The Euclidean metrics have all kinds of nice properties, and ℓ_p with $1 < p < \infty$ share some of them, such as uniform convexity. On the other side, ℓ_∞ metrics are the same as all metrics, and so nobody expects anything nice from them. The space ℓ_1 is somewhere in between, with a (deceptively) simple definition of the norm, some good properties, but many surprises and mysteries. For example, the following famous innocent-looking problem is still open: Does the 3-dimensional space ℓ_1^3 have a finite order of congruence? A metric space M is said to have *order of congruence* at most m if every finite metric space that is not isometrically embeddable in M has a subspace with at most m points that is not embeddable in M.

A second source of importance of ℓ_1 is algorithmic. As we saw at the end of Section 15.8, low-distortion embeddings into ℓ_1 lead to efficient approximation algorithms for partitioning problems such as the sparsest cut. Moreover, various efficient data structures are known for ℓ_1-metrics, for example, for nearest neighbor queries (see, e.g., Indyk's survey [Ind04]). So ℓ_1 is much richer than ℓ_2 but still manageable in many respects.

No flattening in $\boldsymbol{\ell_1}$. Given the algorithmic importance of ℓ_1 and the usefulness of the Jonhson–Lindenstrauss flattening lemma (Theorem 15.2.1), it was natural to ask for an analogue of the flattening lemma in ℓ_1. In a breakthrough paper, Brinkman and Charikar [BC03] showed that flattening is almost impossible. Namely, for all $D \geq 1$, there exist n-point subspaces of ℓ_1 that cannot be D-embedded into ℓ_1^d unless $d = \Omega(n^{c/D^2})$, with a positive absolute constant c. A different and simpler proof was found by Lee and Naor [LN04]. Both proofs use the same example, the shortest-path metrics of the diamond graphs defined in Exercise 15.4.2. We outline the latter proof, which is a beautiful application of the geometry of ℓ_p spaces with $p > 1$ in a question seemingly concerning only $p = 1$. First, one can check that all diamond graphs embed into ℓ_1 (unlimited dimension) with constant distortion. Second, Lee and Naor proved that for every $p > 1$ any embedding of a diamond graph into ℓ_p (unlimited dimension) requires distortion at least $\Omega(\sqrt{(p-1)\log n}\,)$, with the implicit constant independent of p. The proof is based on the following classical inequality for ℓ_p-spaces: $\|x+y\|_p^2 + (p-1)\|x-y\|_p^2 \leq 2(\|x\|_p^2 + \|y\|_p^2)$ for all $x, y \in \ell_p$ (this inequality can be traced back to Figiel [Fig76], although it is not stated explicitly there). Given the inequality, the proof is not hard, and it proceeds analogously to the proof for $p = 2$ in Exercise 15.4.2. Third, a simple calculation shows that for $p(d) := 1 + \frac{1}{\log d}$, the identity map $\mathbf{R}^d \to \mathbf{R}^d$ is an $O(1)$-embedding of ℓ_1^d into $\ell_{p(d)}^d$, and consequently, a D-embedding of any space X into ℓ_1^d yields an $O(D)$-embedding of X into $\ell_{p(d)}$. Hence for the diamond graphs we obtain $D = \Omega(\sqrt{1 + (p(d)-1)\log n}\,)$, and calculation leads to the claimed lower bound $d \geq n^{c/D^2}$.

This bound is known to be almost tight for D of order $\sqrt{\log n}\log\log n$, since by a result of Arora, Lee, and Naor [ALN05] discussed later, every n-point subspace of ℓ_1 embeds into ℓ_2 with distortion $O(\sqrt{\log n}\log\log n)$, the image can further be flattened to $\ell_2^{O(\log n)}$ with distortion 2, say, and the latter space 2-embeds into $\ell_1^{O(\log n)}$.

In spite of the almost complete answer for ℓ_1, the situation concerning flattening in ℓ_p for $p \notin \{1, 2, \infty\}$ is unclear, and in particular, no nontrivial lower bound is known for $p \in (1, 2)$.

Metrics of negative type and $\boldsymbol{\ell_1}$. Let V be a finite set. A metric ρ on V is called *of negative type* if there is a mapping $f\colon V \to \ell_2$ such that $\rho(u,v) = \|f(u) - f(v)\|^2$ for all $u, v \in V$. That is, the distances in ρ can be represented as *squared* Euclidean distances of points in a Euclidean space, or in other words, V with the metric $\sqrt{\rho}$ embeds isometrically into ℓ_2. There is a subtlety in this definition: Not every embedding $f\colon V \to \ell_2$ induces a metric of negative type, since the distance function given by $\|f(u) - f(v)\|^2$ generally need not obey the triangle inequality! Indeed, if $f(u)$, $f(v)$, $f(w)$ are three distinct collinear points, for instance, then the triangle inequality is violated. In the terminology introduced after Proposition 15.5.2, metrics of negative type exactly correspond to points in the intersection of the metric cone with the cone of squared Euclidean metrics.

As was mentioned in the notes to Section 15.5, every ℓ_1 metric is also a metric of negative type. On the other hand, not every metric of negative type embeds isometrically into ℓ_1, but it was conjectured by Goemans and independently by Linial that metrics of negative type are not very far from being ℓ_1; namely, that all metrics of negative type embed into ℓ_1 with distortion bounded by a universal constant.

To appreciate the importance of this conjecture, let us return to the problem of finding a sparsest cut in a given graph $G = (V, E)$. At the end of Section 15.8 we have seen that the problem is equivalent to computing an ℓ_1 metric τ on V such that $\sum_{u,v \in V} \tau(u,v) = n^2$ and $\tau(E) = \sum_{\{u,v\} \in E} \tau(u,v)$ is minimum (we have changed the scaling compared to Section 15.8, in order to have simpler expressions later on). The minimum is attained by a cut pseudometric that corresponds to a sparsest cut. This latter problem, which is an instance of linear optimization over the cone of all ℓ_1-metrics (usually called the *cut cone*), is NP-hard, and we want to solve it approximately. The algorithm outlined in Section 15.8 minimizes over *all* metrics, instead of just all ℓ_1 metrics, and then it approximates the optimum metric by an ℓ_1 metric. This is based on two key facts: First, linear optimization over the metric cone can be done efficiently, and second, a general metric can be $O(\log n)$-embedded in ℓ_1. So a natural way to improve on the approximation factor is to take some subclass of all metrics such that, first, linear optimization over it is still "easy," and second, every metric in this class can be embedded in ℓ_1 with distortion $o(\log n)$, or even a constant.

Linear optimization over all metrics of negative type can be solved efficiently by semidefinite programming, which was briefly discussed at the end of Section 15.5. Indeed, as we have shown there, a function $\tau\colon V \times V \to \mathbf{R}$ is a squared Euclidean (pseudo)metric on $V = \{1, 2, \ldots, n\}$ if and only if there exists an $n \times n$ symmetric positive semidefinite matrix Q such that $\tau(u,v) = q_{uu} + q_{vv} - 2q_{uv}$ for all u, v. In order to pass from squared Euclidean metrics to metrics of negative type, it suffices to enforce the triangle inequality $\tau(u, w) \leq \tau(u, v) + \tau(v, w)$ for all $u, v, w \in V$, which means imposing *linear* inequalities for the entries of Q. Thus, the problem of minimizing a given linear function, such as $\tau(E)$, over all metrics of negative type on V, possibly satisfying additional linear constraints such as $\sum_{u,v \in V} \tau(u, v) = n^2$, is an instance of semidefinite programming, and it can be solved (reasonably) efficiently. Hence, if we could embed every metric of negative type into ℓ_1 with distortion at most D, and do so efficiently, then we could approximate the sparsest cut problem with ratio at most D.

The bad news is that the Goemans–Linial conjecture is *false*: Khot and Vishnoi [KV05] proved that there are n-point metrics of negative type requiring distortion $\Omega((\log \log n)^{1/4-\varepsilon})$ for embedding into ℓ_1, for every fixed $\varepsilon > 0$. Their proof of this (purely geometric) fact is based on algorithmic thinking and, in particular, on *probabilistically checkable proofs*, the current most powerful device for establishing hardness of approximation. Technically, it relies on advanced results in harmonic analysis. We will not attempt to sketch the proof, or even the construction of their badly embeddable example. The lower bound was improved to $\Omega(\log \log n)$, with the same example, by Krauthgamer and Rabani [KR06].

Improved approximation to sparsest cut. The good news is that one *can* improve on the $O(\log n)$ approximation factor for the sparsest cut problem by the above approach, although not all the way to a constant. However, the historical development went differently: First came a breakthrough by Arora, Rao, and Vazirani [ARV04], who discovered an $O(\sqrt{\log n}\,)$-approximation algorithm for the sparsest cut, as well as for several other graph-partitioning problems, and only later and with considerable additional effort it was understood that the geometric part of their argument also leads to low-distortion embeddings.

Here is a "hyperplane-partitioning" algorithm for the sparsest cut. Given a graph $G = (V, E)$, by semidefinite programming we find a metric τ of negative type minimizing $\tau(E)$ subject to $\sum_{u,v \in V} \tau(u, v) = n^2$, as was outlined above. We fix $v_0 \in V$ such that the ball $B(v_0, 4)$ in the τ-metric contains the largest number of points of V; since the sum of all distances is n^2, it is easily calculated that $|B(v_0, 4)| \geq \frac{3}{4}n$. Next, we let x_v, $v \in V$, be points in ℓ_2^n representing τ, i.e. with $\|x_u - x_v\|^2 = \tau(u, v)$. This geometric representation of τ is used to find two large and well-separated subsets $L, R \subset V$. Namely, we choose a random hyperplane h through x_{v_0}, and we let S be the slab consisting of points with distance at most $\varepsilon/\sqrt{n}$ from h, where $\varepsilon > 0$ is a suitable small constant. We let H_L and H_R denote the two open half-spaces whose union is the complement of the slab S, and we define disjoint subsets $L, R \subseteq V$ by $L = \{v \in V \colon x_v \in H_L\}$ and $R = \{v \in V \colon x_v \in H_R\}$. These are not yet the final L and R; we have to prune them as follows: As long as there is a pair $(u, v) \in L \times R$ with $\tau(u, v) \leq C/\sqrt{\log n}$, for a suitable constant C, we remove u from L and v from R. The resulting L and R clearly have

τ-distance at least $C/\sqrt{\log n}$, and they are used as a "core" of the desired sparse cut. Namely, we sort the vertices of V as $(v_1, v_2, \ldots, v_n)$ according to their τ-distance to L, and we output the sparsest among the $n-1$ "candidate" cuts $(\{v_1, \ldots, v_i\}, \{v_{i+1}, \ldots, v_n\})$, $i = 1, 2, \ldots, n-1$. Actually, one more adjustment of the algorithm is needed: At the beginning, if there is an $u \in V$ with $|B(u, \frac{1}{4})| \geq \frac{n}{2}$, then we skip the choice of the random hyperplane and use the ball $B(u, \frac{1}{4})$ as the set L that defines the candidate cuts.

The described algorithm comes essentially from [ARV04], but the authors proved only an approximation bound of $O((\log n)^{2/3})$ for it, and they used a more complicated algorithm to get $O(\sqrt{\log n}\,)$. Lee [Lee05] showed that the simpler algorithm above has approximation ratio $O(\sqrt{\log n}\,)$ as well.

The most time-consuming part of the algorithm is the semidefinite programming step. Arora, Hazan, and Kahale [AHK04] found a way of replacing it by an iterative method of Freund and Schapire for solving zero-sum games, and they achieved a near-quadratic running time.

Improved embedding of metrics of negative type. The heart of the ingenious and somewhat complicated analysis of algorithm described above is in showing that with probability bounded below by a constant, the resulting sets L and R are large; $|L|, |R| = \Omega(n)$. In particular, it follows that for any n-point V with a metric τ of negative type normalized so that $\sum_{u,v\in V} \tau(u,v) = n^2$, and with no very large and tight cluster (a ball of radius $\frac{1}{4}$ containing at least half of the set), there are subsets $L, R \subset V$ of size $\Omega(n)$ with $\tau(L,R) = \Omega((\log n)^{-1/2})$.

To understand the statement better, let us consider the m-dimensional Hamming cube (with $n = 2^m$) as an example; since the average distance in it is of order m, we have to scale it by roughly $\frac{1}{m}$ to meet the condition $\sum_{u,v\in V} \tau(u,v) = n^2$. So in the original Hamming cube we look for linear-size subsets separated by a gap of order $\sqrt{m}$. On the one hand, we can take all vertices with at most $\frac{m}{2}$ ones for L and all vertices with at least $\frac{m}{2} + \sqrt{m}$ ones for R. On the other hand, showing that the gap cannot be of order larger than $\sqrt{m}$ is a nice exercise using measure concentration (Theorem 14.2.3 slightly generalized to the case of $\mathrm{P}[A]$ bounded below by an arbitrary positive constant, rather than by $\frac{1}{2}$). So the result of Arora et al. can be thought of as showing that, in a sense, no metric space of negative type can have measure concentration stronger than the Hamming cube.

The result was further refined, analyzed, and applied in subsequent papers. Lee [Lee05] simplified the proof and provided a stronger version (roughly speaking, the sets R and L generated by the algorithm are shown to be "more random" than in [ARV04], which is crucial for low-distortion embeddings; see below). Naor, Rabani, and Sinclair [NRS04] derived a graph-theoretic consequence, and observed that the proof of [ARV04] uses only little of the properties of metrics of negative type: They proved analogous results for metrics *uniformly embeddable* into ℓ_2 and for metrics *quasisymmetrically embeddable* into ℓ_2. (The latter, less well-known notion, due to Beurling and Ahlfors, is defined as follows: An embedding $f\colon (X,\rho) \to (Y,\sigma)$ is *quasisymmetric with modulus* η, where $\eta\colon [0,\infty] \to [0,\infty)$ is a strictly increasing function, if $\sigma(f(x),f(y))/\sigma(f(x),f(z)) \le \eta(\rho(x,y)/\rho(x,z))$ for every $x,y,z \in X$.)

Chawla, Gupta, and Räcke [CGR05] used the geometric results of [ARV04] for proving that every metric of negative type on n points embeds into ℓ_2 with distortion $O((\log n)^{3/4})$, and Arora, Lee, and Naor [ALN05] obtained an improved and nearly tight bound of $O(\sqrt{\log n}\cdot\log\log n)$. This , of course, also provides embeddings of negative type metrics into ℓ_1, and thus an alternative algorithm for the sparsest cut problem, whose approximation guarantee is little worse than that of Arora, Rao, and Vazirani [ARV04], but which is more general: It can also handle a weighted version of the problem (with general capacities and demands, for those familiar with multicommodity flows).

Since ℓ_1 metrics are a subclass of metrics of negative type, Arora, Lee, and Naor [ALN05] also almost solved another famous open problem: What is the largest distortion needed for embedding an n-point ℓ_1 metric into ℓ_2? By their result we now know that the m-dimensional Hamming cube, which needs distortion $\Omega(\sqrt{m}\,)$, is nearly the worst example.

Refined embedding methods. The proof of Arora et al. [ALN05] combines the main geometric result of [ARV04] mentioned above with a general-purpose embedding technique, called *measured descent*. This technique, developed by Krauthgamer, Lee, Mendel, and Naor [KLMN04], is a common generalization and significant refinement of the embedding methods of Bourgain and of Rao, and it has been applied for solving several other problems. We outline some of the key ingredients.

We want to embed an n-point metric space (V,ρ) into ℓ_2 (variations for other target spaces are obviously possible but here we stick to the Euclidean case). We may assume, after re-scaling ρ, that the distances in (V,ρ) are between 1 and 2^m. The first idea for constructing a low-distortion embedding, present explicitly in Rao's embedding (see the notes to Section 15.8), and less explicitly in Bourgain's embedding, is to divide the task conceptually into two steps:

(i) For each *scale* $\Delta = 2^0, 2^1, \ldots, 2^m$, construct a mapping $\varphi_\Delta\colon V \to \ell_2$ that "takes care" of all pairs $(u,v) \in V\times V$ with $\Delta \le \rho(u,v) < 2\Delta$.
(ii) Combine the φ_Δ over all Δ into a single D-embedding $f\colon V \to \ell_2$.

First we need to clarify what do we mean by "taking care" in (i). Following Lee [Lee05], we say that a mapping $\varphi\colon V \to \ell_2$ is an *embedding with deficiency* D_0 *at scale* Δ if φ is 1-Lipschitz (nonexpanding) and we have $\|\varphi(u)-\varphi(v)\|_2 \ge \rho(u,v)/D_0$ for all $u,v \in V$ with $\Delta \le \rho(u,v) < 2\Delta$.

We have already seen an example in the description of Rao's embedding: If the reader has the energy to look it up, she can see that the mapping f_Δ constructed there was an embedding with deficiency $O(1)$ at scale $C\Delta$. A useful abstraction of the approach used there is the notion of *padded decomposition*. A padded decomposition should be imagined as a randomized algorithm that receives a finite metric space (V, ρ) as input and outputs a random partition $\mathcal{P}$ of V. The definition is as follows: A *D_0-padded decomposition at scale* Δ for (V, ρ) is a probability distribution on partitions of V, such that

- In each partition $\mathcal{P}$ occurring in this distribution, all classes have diameter smaller than Δ.
- For every $v \in V$, the ball $B(v, \Delta/D_0)$ is fully contained in a single class of $\mathcal{P}$ with probability bounded below by a positive constant.

Thus, we have partitions into pieces of diameter at most Δ, but each point has a good chance to be at least Δ/D_0 away from the boundary of its piece. By repeating the argument in the description of Rao's embedding almost verbatim, we get that a D_0-padded decomposition at scale Δ yields an embedding $V \to \ell_2$ with deficiency $O(D_0)$ at scale Δ.

In the embedding of a metric space (V, τ) of negative type into ℓ_2 by [ALN05], the way of producing an embedding for a given scale Δ is not based on a padded decomposition. In the first auxiliary step, (V, τ) is mapped to (V, τ') by a nonexpanding map, where τ' is also of negative type but has diameter $O(\Delta)$, and moreover, $\tau'(u, v) \geq \frac{1}{2}\tau(u, v)$ for all u, v with $\tau(u, v) \leq 2\Delta$. Next, the randomized algorithm of Arora, Rao, and Vazirani is applied to (V, τ'), it produces the large and well-separated subsets L and R, and these are used for defining one coordinate in the embedding for scale Δ. The hope is that a fixed pair (u, v) with $\Delta \leq \tau(u, v) \leq 2\Delta$ has a good chance of having $u \in L$ and $v \in R$, and then (u, v) is taken care of. In reality, things are more complicated, since not every pair (u, v) has a sufficiently large probability of $u \in L$, $v \in R$, and a reweighting strategy is used: The algorithm is called repeatedly, and the pairs (u, v) that were unfortunate so far are assigned more weight, so that their chances at being separated increase. This reweighting strategy, well known in other areas, was first used in this problem by Chawla et al. [CGR05].

We now come to the second issue: Assuming that we can produce embeddings with small deficiency for every scale, how do we glue them together to make a low-distortion embedding? In both Bourgain's and Rao's method, the coordinates of all the scale embeddings are simply put side-by-side. This is sufficient in the cases dealt with by Bourgain and by Rao, but it is not enough for more ambitious applications. The main contribution of Krauthgamer et al. [KLMN04] is in a more sophisticated way of putting the scale embeddings together. A neat way of encapsulating the results with a user-friendly interface is due to Lee [Lee05], who proved the following *gluing lemma*: *Let (V, ρ) be an n-point metric space, and suppose that for every scale Δ there is an embedding $\varphi_\Delta \colon V \to \ell_2$ with deficiency at most D_0 at scale Δ. Then (V, ρ) embeds in ℓ_2 with distortion $O(\sqrt{D_0 \log n})$.* A more refined version by Arora et al. [ALN05] can take advantage of better scale embeddings of smaller subspaces: *Let (V, ρ) be an n-point metric space, and suppose that for every scale Δ and every m-point subspace $S \subseteq V$, $2 \leq m \leq n$, there exists* *an embedding* $\varphi_{S,\Delta}\colon S \to \ell_2$ *with deficiency at most* $C(\log m)^{1/2+\alpha}$*, where* C *and* $\alpha \in [0, \frac{1}{2})$ *are constants. Then* (V, ρ) *embeds in* ℓ_2 *with distortion* $O((\log n)^{1/2+\alpha} \log\log n)$.

In the proof of the gluing lemma and similar results, the scale embeddings are combined using suitable "partitions of unity" (a well-known tool for constructing maps in analysis), which assign more importance to regions of the considered space with larger *local dimension*. The local dimension at scale Δ at a point $v \in V$ is defined as $\log_2(|B(v, 2\Delta)|/|B(v, \Delta)|)$. For a version of the gluing lemma that refers to the local dimension explicitly see [ALN05].

A result of Krauthgamer et al. [KLMN04] proved using measured descent states that an n-point metric space with doubling constant λ (that is, for every $r > 0$, every ball of radius $2r$ can be covered by at most λ balls of radius r) can be embedded in ℓ_2 with distortion $O(\sqrt{\log \lambda \log n}\,)$, strengthening previous results by Gupta et al. [GKL03]. In this case, the assumption allows one to produce suitable padded decompositions (this was done in [GKL03]), and the proof is finished by a direct application of the gluing lemma cited above.

Embeddings of doubling metric spaces were investigated by Assouad in [Ass83b], which is one of the pioneering works on low-distortion embeddings. He asked whether every finite metric space with doubling constant λ can be embedded in $\mathbf{R}^d$ with distortion at most D, with d and D depending only on λ. This was answered negatively by Semmes (we refer, e.g., to [GKL03] for references not given here), who noted that a solution is a direct consequence of a theorem of Pansu. Subsequently other counterexamples were found, such as the Laakso graphs defined in Exercise 1 below, which show that the $O(\sqrt{\log n}\,)$ of [KLMN04] is asymptotically tight for constant λ.

Yet another theorem from Krauthgamer et al. [KLMN04] guarantees an $O(1)$-embedding of every n-point planar-graph metric in $\ell_\infty^{O(\log n)}$.

On the impossibility of local characterizations of ℓ_1. Arora, Lovász, Newman, Rabani, Rabinovich, and Vempala [ALN$^+$06] studied the following question: Let (V, ρ) be an n-point metric space such that every k-point subspace can be $D_{\rm loc}$-embedded in ℓ_1. What can be said about the minimum distortion D required for embedding all of (V, ρ) in ℓ_1? This is a question about a "local characterization" of metrics that are approximately ℓ_1. For every fixed $\varepsilon > 0$ and for all n they constructed n-point (V, ρ) such that every subspace on $n^{1-\varepsilon}$ points embeds in ℓ_1 with distortion $O(\varepsilon^{-2})$, while (V, ρ) itself requires the (worst possible) distortion $\Omega(\log n)$. The example is the shortest-path metric of a random 3-regular graph with all edges in cycles of length below $c \log n$ deleted, for a suitable constant $c > 0$. Such a graph is known to be a good expander, and hence the $\Omega(\log n)$ lower bound follows, while the proof of embeddability of $n^{1-\varepsilon}$-point subspaces is a gem using some polyhedral combinatorics. For the case of $D_{\rm loc} = 1$, i.e., k-point subspaces embedding isometrically, Arora et al. construct an example where all k-point subspaces embed isometrically even to ℓ_2, but embedding the whole of the space in ℓ_1, or even in a metric space of negative type, requires distortion $\Omega((\log n)^{\frac{1}{2}\log(1+\frac{1}{k-1})})$. This shows that approximate embeddability in ℓ_1 cannot be characterized by conditions imposed on subspaces with constantly many points. Arora et al. also deal with classes more general than ℓ_1. One of their main open questions was answered by Mendel and Naor [MN06].

Ramsey-type results. Let $R_p(n, \alpha)$ denote the largest m such that for every n-point metric space Y there exists an m-point subspace of ℓ_p that α-embeds in Y. Let $R_{\mathrm{UM}}(n, \alpha)$ be defined similarly with "subspace of ℓ_p" replaced with "ultrametric space" (see the notes to Section 15.8 for the definition). It is not hard to show that every ultrametric embeds isometrically in ℓ_2, and so $R_2(n, \alpha) \geq R_{\mathrm{UM}}(n, \alpha)$.

Bourgain, Figiel, and Milman [BFM86] were the first to ask the Ramsey-type question, how large almost-Euclidean subspaces must be present in any n-point metric space? In other words, what is the order of magnitude of $R_2(n, \alpha)$? (This was motivated as a metric analog of Dvoretzky's theorem about almost Euclidean subspaces of normed spaces; see Chapter 14.) They showed that $R_2(n, 1+\varepsilon)$ is of order $\log n$ for every fixed $\varepsilon \in (0, \varepsilon_0)$, where where ε_0 is a sufficiently small constant (see Exercise 15.4.3). This seemed to have settled the question, in a not too interesting way, until Bartal, Bollobás, and Mendel [BBM01] discovered that if considerably larger distortions are allowed, then quite large almost-Euclidean, and even almost-ultrametric, subspaces always exist. Their results were greatly improved and extended by Bartal, Linial, Mendel, and Naor [BLMN03], who found out that, unexpectedly, distortion 2 represents a sharp threshold in the Ramsey behavior. On the one hand, for distortions below 2 both R_2 and R_{UM} are logarithmic: For every $\alpha \in (1, 2)$, we have

$$c(\alpha)\log n \leq R_{\mathrm{UM}}(n, \alpha) \leq R_2(n, \alpha) \leq 2\log n + C(\alpha)$$

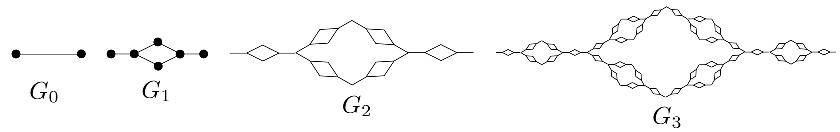
for all n, with suitable positive $c(\alpha)$ and $C(\alpha)$. On the one hand, for distortions $\alpha > 2$, both R_2 and R_{UM} behave as a power of n:

$$R_2(n, \alpha) \geq R_{\mathrm{UM}}(n, \alpha) \geq n^{c_1(\alpha)}$$

for all n, with a suitable $c_1(\alpha) > 0$. Moreover, the order of magnitude of the best possible $c_1(\alpha)$ for sufficiently large α has been determined quite well: $1 - C_2 \cdot \frac{\log \alpha}{\alpha} \leq c_1(\alpha) \leq 1 - \frac{c_2}{\alpha}$ with positive constants C_2 and c_2. Similar bounds are also known for R_p with $1 \leq p < \infty$; the case $1 \leq p \leq 2$ is from Bartal et al. [BLMN03], while the final touch for $p > 2$ was made by Charikar and Karagiozova [CK05]. Besides the results, the methods of Bartal et al. [BLMN03] are also of interest, since the embedding of an ultrametric space is constructed in a way quite different from the "usual" Bourgain's technique and related approaches.

Exercises

1. (Laakso graphs) Let $G_0, G_1, \ldots$ be the following sequence of graphs:



(a) Similar to Exercise 15.4.2, prove that any Euclidean embedding of the shortest-path metric of G_k incurs distortion $\Omega(\sqrt{k})$. [3]
(b) Prove that there is a constant λ such that for every $r > 0$, every ball of radius $2r$ in G_k can be covered by at most λ balls of radius r. [4]

In this form, the result is from Lang and Plaut [LP01]; Laakso's formulation and proof were somewhat different.

⇑ NEW ⇑