Workshop on Discrete Metric Spaces and their Algorithmic Applications

Haifa, March 3–7, 2002

Open Problems

edited by Jiří Matoušek

Preface

Finite metric spaces have emerged in recent years as a new and influential branch of discrete mathematics, with deep and surprising applications in Computer Science. The Workshop on Discrete Metric spaces and their Algorithmic Applications was organized at the Holiday Inn Hotel in Haifa, Israel, in March 2002, with the aim of summarizing the achievements in this area to date and clarifying the core questions of the emerging field.

The participation was very strong and included experts from quite diverse fields, ranging from Banach spaces to approximation algorithms. The scientific programme consisted of 23 talks, from one-hour tutorials to short presentations, and a problem session.

The open problems collected in this booklet have been presented during the talks at the workshop (in particular, an opening lecture by Nathan Linial was mostly devoted to open problems), during the problem session, or contributed later by the participants. The heading of each problem includes a short title (mostly assigned by the editor, in order to facilitate a quick orientation) and the name of the person presenting the problem. This is not necessarily the original author of the question; some of the problems seem to be folklore and it may be nontrivial to trace their origins. The problems were open, to the best knowledge of the people presenting them, at the time of editing this collection.

The financial support of the workshop by the US–Israel Binational Science Foundation (BSF), by the Caesaria Rothschild Foundation Institute, and by the Haifa University Research Counsel is gratefully acknowledged. Finally, we would like to thank all the participants for their contribution to the nice and productive atmosphere of the workshop.

Main Organizer:	Yuri Rabinovich (University of Haifa)
Co-organizers:	Uriel Feige (Weizmann Institute, Rehovot)
	Nathan Linial (Hebrew University of Jerusalem)
	Jiří Matoušek (Charles University, Prague)
	Ilan Newman (University of Haifa)
	Alistair Sinclair (University of California, Berkeley)

List of talks

SUNDAY, March 3

N. Linial	Theory of finite metric spaces: Introduction via a tour of open problems
P. Indyk	Computational aspects of embeddings
A. Magen	An extension of Johnson–Lindenstrauss Lemma for volumes

MONDAY, March 4

I. Newman	Outer-planar metrics and beyond
G. Cormode	Metrics on strings and permutations
U. Feige	Volumes in embeddings
C. Chekuri	Deterministic embeddings into dominating trees
M. Elkin	Semi-embeddings with additive distortion

TUESDAY, March 5

Y. Rabani	On geometric nearest neighbour
S. Rao	Partitioning for metric spaces and distributed shortest-paths algorithms
U. Zwick	Approximate distance oracles

WEDNESDAY, March 6

M. Katz	Dense packings, systolic geometry and period map
G. Schechtman	Tight linear embeddings of finite-dimensional normed spaces
Yu. Rabinovich	Graphs and their metrics
A. Naor	On Markov type of a metric space
C. Chekuri	Metric Labelling
S. Naor	Tree packing and approximate k-cuts
T. Tankel	Using particle mechanics for embeddings of distance maps

THURSDAY, March 7

S. Guha	Stream algorithms and embeddings
R. Ravi	Application of embedding technique in Approximation Algorithms
Y. Bartal	Embeddings into dominating trees
M. Deza	Computation of cones of metrics, and cones of quasi-, hemi-, and super-metrics
L. Roditty	Roundtrip spanners and roundtrip routing in directed graphs

General references

Some background material on the topics of the workshop can be found in the following references.

- P. Indyk: Algorithmic applications of low-distortion embeddings. In *Proc. 42nd IEEE Symposium on Foundations of Computer Science*, 2001.
  (A survey with emphasis on algorithmic applications.)
- J. Matoušek: *Lectures on Discrete Geometry*, Chapter 15, Springer, May 2002.
  (Proofs of several basic results and a survey.)
- M. M. Deza and M. Laurent: *Geometry of Cuts and Metrics*, Springer-Verlag, 1997.
  (Isometric embeddings, mainly into ℓ_1.)

Overview of the Problems

1 ℓ_1 and ℓ_2 embeddings

1.1 More examples badly embeddable into ℓ_2 (Nathan Linial)

Bourgain's basic theorem says that $c_2(X) \leq O(\log n)$ for every n-point metric space X. (Here $c_p(X)$ is the minimum distortion necessary for embedding X into ℓ_p.) This bound is known to be tight, but so far the only known extreme examples are (the metrics of) constant-degree expander graphs. Can you find other families of explicit examples?

1.2 ℓ_1 metrics into ℓ_2 (Nathan Linial)

Is it true that every n point metric X in ℓ_1 embeds into ℓ_2 with distortion $O(\sqrt{\log n})$? If not, how large can the required distortion be?

1.3 Subspaces of ℓ_1 into ℓ_1^n (Gideon Schechtman)

Given m points $y_i \neq 0$ in $\mathbb{R}^k$, let

$$K = \sum_{i=1}^{m} [-y_i, y_i] = \Big\{ a_1 + a_2 + \cdots + a_m;\ a_i \in [-y_i, y_i] \Big\}.$$

Problem: Is there a universal constant C such that given $x_1, \ldots, x_k \in K$, there exist signs $\varepsilon_1, \ldots, \varepsilon_k \in \{-1, +1\}$ with

$$\sum_{i=1}^{k} \varepsilon_i x_i \in C\sqrt{k} \cdot K?$$

Known:

1. There exist $\varepsilon_1, \ldots, \varepsilon_k$ with $\sum_{i=1}^{k} \varepsilon_i x_i \in C\sqrt{k \log\log k} \cdot K$.
2. A negative solution implies

$$\lim_{k \to \infty} \frac{N_1(k,2)}{k} = \infty,$$

where

$$N_1(k,2) \quad = \quad \sup_{\substack{X \text{ is a } k\text{-dim.} \\ \text{subspace of } L_1}} \inf \{n;\ X \text{ can be 2-embedded into } \ell_1^n\}.$$

1.4 Dimension reduction in ℓ_1 (Nathan Linial)

Given an integer n and $\gamma > 1$, what is the least $k = k(n, \gamma)$ such that every n-point metric in ℓ_1 can be embedded in ℓ_1^k with distortion at most γ? Specifically, is $k \leq O(\log n)$ for a bounded γ?

1.5 Squared ℓ_2 metrics into ℓ_1 (Nathan Linial)

Can the cut cone be approximated well by a cone for which we have efficient membership/separation oracles? Specifically, how far is it from the cone of squared ℓ_2-metrics? In particular, is there a constant C such that every squared ℓ_2-metric embeds into ℓ_1 with distortion at most C? A metric space (X, ρ) is a *squared ℓ_2-metric* (also called a *metric of negative type*) if there is a map $f\colon X \to \ell_2$ such that $\rho(x, y) = \|x - y\|^2$ for all $x, y \in X$.

1.6 Girth and ℓ_1/ℓ_2 embeddings (Nathan Linial)

In this question all vertices in the graph G have all degrees at least 3. Linial et al. [N. Linial, E. London, and Yu. Rabinovich: The geometry of graphs and some its algorithmic applications, *Combinatorica*, 15:215–245, 1995] conjectured that $c_2(G) \geq \Omega(\mathrm{girth}(G))$. Linial et al. [N. Linial, A. Magen, and N. Naor: Euclidean embeddings of regular graphs—the girth lower bound, *Geometric and Functional Analysis*, in press] proved $c_2(G) \geq \Omega(\sqrt{\mathrm{girth}(G)}\,)$ for regular graphs. (Here $c_p(X)$ is the minimum distortion to embed a metric space into ℓ_p.) What is the true bound? Also, does $c_1(G)$ tend to infinity with $\mathrm{girth}(G)$? The c_2 analogue of this weaker version is easy, since c_2 is unbounded for trees of large depth, but this is obviously not the case for c_1.

1.7 Pentagonal into ℓ_1 (Michel Deza)

Is any (finite) 5-gonal metric space embeddable into ℓ_1 with distortion bounded by a constant? Here we call a metric space $(2k+1)$-gonal if for every multiset A of k points and every multiset B of $k+1$ points in the space, the sum of all pairwise distances in A plus the sum of all pairwise distances in B does not exceed the sum of all distances between a point of A and a point of B. The case $k = 1$ corresponds to the triangle inequality, and a 5-gonal metric space corresponds to the case $k = 2$. We know, for any $k > k'$, examples of $(2k'+1)$-gonal, but not $(2k+1)$-gonal, metric spaces. Any metric subspace of ℓ_1 is hypermetric (i.e., $(2k+1)$-gonal for every natural k), but not vice versa (for example, the shortest path metric of $K_7 - C_5$ and $K_7 - P_3$).

1.8 Algorithmic difficulty of ℓ_1-embeddings (Chandra Chekuri; Anupam Gupta)

Can the minimum distortion for embedding of a given finite metric space into ℓ_1 be approximated within a constant factor in polynomial time? (Testing isometric ℓ_1-embeddability is known to be NP-hard.)

What about embedding into $\mathbb{R}^1$ (or $\mathbb{R}^d$), even for a tree metric?

2 Planar-graph metrics

Let $\mathcal{G}$ be a class of (finite) graphs. Each graph $G \in \mathcal{G}$ with nonnegative real weights on edges defines the shortest-path metric on its vertex set. A metric space is called a *$\mathcal{G}$-metric* if it is isometric to a subspace of such a metric space for some $G \in \mathcal{G}$. For $\mathcal{G}$ being the class of all planar graphs, we speak of *planar metrics*, or perhaps more descriptively, of *planar-graph metrics*.

2.1 Planar-graph metrics into ℓ_1 (Nathan Linial)

Is there a constant C such that the metric of every finite weighted planar graph embeds in ℓ_1 with distortion at most C? Even more generally, does the same hold for every nontrivial minor-closed family of graphs? Gupta et al. [A. Gupta, I. Newman, Yu. Rabinovich, and A. Sinclair: Cuts, trees and ℓ_1-embeddings of graphs, *Proc. 40th IEEE Symposium on Foundations of Computer Science,* 1999, pages 399–409] proved this for the class of graphs with no K_4-minor, as well as for the class of graphs with no $K_{2,3}$ minor.

2.2 Decision problem (Yuri Rabinovich)

Given a metric, can one decide in polynomial time whether it is planar metric?

2.3 Into ℓ_∞^d (Yuri Rabinovich)

Can every planar metric on n points be embedded into $\ell_\infty^{c_1 \log n}$ with distortion at most c_2, for some universal constants c_1, c_2?

2.4 How large graph? (Jiří Matoušek; probably folklore)

Let $\mathcal{G}$ be some class of graphs, say all planar graphs. What is the smallest $N = N(\mathcal{G}, n)$ such that every n-point $\mathcal{G}$-metric can be represented as a subspace of the metric of a graph $G \in \mathcal{G}$ with at most N vertices? (For $\mathcal{G}$ = all trees, $O(n)$ suffices.) What if we do not insist on an exact representation, but allow for some constant distortion, say?

2.5 Average distance for line embedding (Yuri Rabinovich)

Is there a constant $c > 0$ such that for any planar metric (X, ρ), a mapping $f: X \to \mathbb{R}^1$ exists with

- $|f(x) - f(y)| \leq \rho(x, y)$; that is, f is non-expanding, and
- $\sum_{x,y \in X} |f(x) - f(y)| \geq c \cdot \sum_{x,y \in X} \rho(x, y)$; that is, the average distance decreases at most by a constant?

This is known to hold for the metric of a graph of bounded treewidth (with c depending on the treewidth). On the other hand, for the metric of an n-vertex constant-degree expander, any non-expanding embedding into $\mathbb{R}^1$ decreases the average distance by at least $\Omega(\log n)$.

2.6 Into $\mathbb{R}^2$ (Jiří Matoušek)

Is there any planar-graph metric on n vertices requiring distortion more than $O(\sqrt{n})$ for embedding into $\mathbb{R}^2$? It is known, and not difficult to prove, that if every edge of a fixed non-planar graph is replaced by a path of length n, the resulting graph requires distortion $\Omega(n)$. On the other hand, every tree with unit-length edges can be embedded with $O(\sqrt{n})$ distortion [R. Babilon, J. Matoušek, J. Maxová, and P. Valtr: Low-distortion embeddings of trees, *Proc. Graph Drawing 2001*, Springer, 2002] and every n-point metric space can be $O(n)$-embedded even into $\mathbb{R}^1$ [J. Matoušek: Bi-Lipschitz embeddings into low-dimensional Euclidean spaces, *Comment. Math. Univ. Carolinae*, 31:589–600, 1990].

3 Probabilistic embeddings

3.1 Probabilistic into trees (Yair Bartal)

Prove or disprove the following conjecture: any metric space on n points can be $O(\log n)$-probabilistically embedded into hierarchically well-separated trees (equivalently, embedded into a convex combination of dominating HSTs). The known bounds are $\Omega(\log n)$ from below and $O(\log n \log\log n)$ from above.

The following definition of HST follows [Y. Bartal, B. Bollobás, and M. Mendel: A Ramsey-type theorem for metric spaces and its applications for metrical task systems and related problems, *Proc. 42nd IEEE Symposium on Foundat. Comput. Sci.*, 2001], and it slightly differs from the ones in earlier papers, although it is essentially equivalent for $k > 1$. A 1-HST is exactly an *ultrametric*; that is, the metric on the leaves of a rooted tree T (with weighted edges) such that all leaves have the same distance from the root. For a k-HST with $k > 1$ we require that, moreover, $\Delta(v) \leq \Delta(u)/k$ whenever v is a child of u in T, where $\Delta(v)$ denotes the diameter of the subtree rooted at v (w.l.o.g. we may assume that each non-leaf has degree at least 2, and so $\Delta(v)$ equals the distance of v to the nearest leaves).

3.2 Probabilistic into spanning trees (Yair Bartal; Nathan Linial)

What is the best statement of the form: Given any graph G with nonnegative edge-weights, there is a distribution π on G's spanning trees such that for every two vertices x and y,

$$\mathbf{E}_\pi d_T(x, y) \leq O(d_G(x, y) f(n)).$$

Here d_G and d_T are distances in G and in the tree T, and $\mathbf{E}_\pi$ stands for expectation w.r.t. the distribution π. Alon et al. [N. Alon, R. M. Karp, D. Peleg, and D. West: A graph-theoretic game and its application to the k-server problem, *SIAM J. Computing*, 24(1):78–100, 1995] showed this with $f(n) \leq \exp(O(\sqrt{\log n \log\log n}))$, but this might be true even with $f(n) \leq O(\log n)$. What is the truth?

3.3 Algorithmic complexity (Chandra Chekuri)

Given a weighted graph G (or equivalently, a metric on $V(G)$), what is the complexity of finding the optimal embedding of G into a convex combination of dominating trees? We know that an $O(\log n \log\log n)$ absolute distortion is achievable but this is true for all metrics on n vertices. We are interested in the optimal embedding for G. Can we obtain an $O(1)$ approximation for the optimal embedding? In the paper [M. Charikar, C. Chekuri, A. Goel, S. Guha, and S.A. Plotkin: Approximating a Finite Metric by a Small Number of Tree Metrics, Proc. 39th IEEE Sympos. Foundat. Comput. Sci. 1998, pages 379–388] it is shown that this is related to the following optimization problem. Given a graph G and a weight function $w\colon V \times V \to V$, find a dominating tree metric for G such that the quantity $\sum_{u,v \in V(G)} w(u,v) d_T(u,v)$ is minimized. An approximation of α for this later problem implies an α approximation for the former problem and vice versa.

Comment (Anupam Gupta): We do not even know how to find the best distortion (single) spanning tree for a metric. The minimum spanning tree gives a trivial $n-1$ upper bound, and if there is an isometric cycle of girth g, that implies an $\Omega(g)$ lower bound.

4 Ramsey-type theorems

4.1 HST in a metric space (Yair Bartal)

Let $R_{\mathrm{HST}}(c,n)$ be the largest number k so that any metric space on n points contains a subspace of size at least k that is c-embeddable into an HST (see problem 3.1). Prove or disprove the following conjecture: There exist constants c and $\varepsilon > 0$ such that $R_{\mathrm{HST}}(c,n) \geq n^{\varepsilon}$. In [Y. Bartal, B. Bollobás, and M. Mendel: A Ramsey-type theorem for metric spaces and its applications for metrical task systems and related problems, *Proc. 42nd IEEE Symposium on Foundat. Comput. Sci.*, 2001] it is shown that for every $c \geq 3$, $R_{\mathrm{HST}}(c,n) \geq 2^{\Omega(\log^{1-1/c} n)}$.

4.2 Euclidean subspace in a metric space (Yair Bartal; Nathan Linial)

Let $R_p(c,n)$ be the largest number k so that any metric space on n points contains a subspace of size at least k that is c-embeddable into ℓ_p. Since every HST is an ultrametric, and hence it embeds isometrically into ℓ_2, we have $R_p(c,n) \geq R_{\mathrm{HST}}(c,n)$. Can one get better lower bounds?

Specifically, prove or disprove the following conjecture: For every $\delta > 0$ there exists $c > 0$ such that $R_2(c,n) \geq n^{1-\delta}$. Or, a weaker one: There exist constants c and $\alpha > 0$ such that $R_2(c,n) \geq n^{\alpha}$.

For $c = 1+\varepsilon$ with with ε small enough, we know that $R_2(1+\varepsilon, n) = \Theta(\log n)$ [J. Bourgain, T. Figiel, and V. Milman: On Hilbertian subsets of finite metric spaces, Israel J. Math. 55:147–152, 1986]. Further, for every $c \geq 1$ there exists $\delta > 0$ with $R_2(c,n) \leq n^{1-\delta}$ (and even with $R_p(c,n) \leq n^{1-\delta}$ for every fixed $p \geq 1$) [Y. Bartal and M. Mendel, unpublished].

5 Low-dimensional embeddings

5.1 Assouad's problem (reminded by J. Matoušek)

Let X be a separable metric space such that any ball of radius $2r$ can be covered by at most L balls of radius r, for all $r > 0$. Do there exist numbers $k = k(L)$ and $C = C(L)$ such that any such X can be embedded into k-dimensional Euclidean space with distortion at most C? (It is known that for any fixed $p < 1$, the metric space arising from X by raising all distances to power p can be so embedded, with k and C depending on p as well; see [P. Assouad: Lipschitz embeddings into $\mathbb{R}^n$ (in French), *Bull. Soc. Math. Fr.* 111:429–448, 1983].)

5.2 A variation on a theme by L. Levin (Nathan Linial)

Given a graph G we seek an embedding $\varphi: V(G) \to \ell_2^d$ such that

(i) $\|\varphi(x) - \varphi(y)\| \geq 1$ for every two distinct vertices in G, and

(ii) $\|\varphi(x) - \varphi(y)\| \leq 2$ for every two adjacent vertices. (The constant 2 is arbitrary).

Our goal is to minimize the dimension d. By simple volume considerations $d \geq \Omega(\rho(G))$, where

$$\rho(G) := \max_{x,r} \frac{\log(1 + |B_r(x)|)}{\log(r+1)}$$

is G's growth rate. (Essentially, the smallest t such that every ball of radius r in G has no more than r^t vertices). Can the smallest d be bounded in terms of ρ? Perhaps even $d \leq O(\rho)$?

5.3 Bandwidth and ball growth (Nathan Linial)

An easy lower bound on the bandwidth of any graph G is

$$\mathrm{bw}(G) \geq \Omega(\beta(G))$$

where

$$\beta(G) := \max_{x,r} \frac{|B_r(x)|}{r}.$$

Feige [U. Feige: Approximating the bandwidth via volume respecting embeddings, *J. Comput. Syst. Sci*, 60:510–539, 2000] showed that these two parameters differ by a factor that is at most polylogarithmic in n. For expanders clearly a logarithmic gap exists. So we ask: Is it true that

$$\mathrm{bw}(G) \leq O(\beta(G) \log n)$$

always holds? This doesn't seem to be known even for trees.

5.4 Into $\mathbb{R}^3$ (Jiří Matoušek)

What is the worst-case distortion necessary to embed an n-point metric space into $\mathbb{R}^3$?

There is an $\Omega(\sqrt{n})$ lower bound and an $O(n^{2/3} \log^{3/2} n)$ upper bound; see [J. Matoušek: Bi-Lipschitz embeddings into low-dimensional Euclidean spaces, *Comment. Math. Univ. Carolinae*, 31:589–600, 1990; also available at `kam.mff.cuni.cz/~matousek`]. Similarly, there is a significant gap between known lower and upper bounds for the worst-case distortion into $\mathbb{R}^{2k+1}$, $k \geq 1$ (while for even dimensions the upper and lower bounds match up to a logarithmic factor).

5.5 Lipschitz mapping of n grid points onto a square (Uriel Feige)

Consider the infinite integer two-dimensional lattice $\mathbb{Z}^2$ and number n distinct points in it from 1 to n. This induces an n point finite metric space (S, d) with ℓ_1 distances defined in a natural way. For what value of $c > 1$ (possibly depending on n) is it always possible to number the $\sqrt{n}$ by $\sqrt{n}$ grid G from 1 to n while ensuring that for every i and j, their ℓ_1 distance in G is at most c times their distance in S? (We allow here distances to shrink arbitrarily.)

This problem was motivated by generalizations of the bandwidth problem to two dimensions, with (pseudo) applications in VLSI layout.

Comment (J. Matoušek): This reminds of a beautiful (but not necessarily closely related) problem of Laczkovich in geometric measure theory: Does there exist, for every set $E \subseteq \mathbb{R}^d$ of positive Lebesgue measure, a Lipschitz map $f: \mathbb{R}^d \to \mathbb{R}^d$ with $[0,1]^d \subseteq f(E)$? A positive answer is known for $d = 2$, due to Preiss, while the question is wide open for all $d \geq 3$. A "discrete" version might perhaps go as follows: Does there exist a constant $C = C(d)$ such that every Cn^d-point set in $\mathbb{Z}^d$ can be mapped *onto* the "cube" $\{1, 2, \ldots, n\}^d$ by a C-Lipschitz map? For $d = 2$, this is a weakening of Feige's problem above, and it is known to hold, but nothing is known for $d = 3$.

6 Volume-respecting embeddings

6.1 Volume-respecting for large k (Uriel Feige)

The following question was asked in my talk on volume respecting embeddings. For the relevant definitions see [U. Feige: Approximating the bandwidth via volume respecting embeddings, *J. Comput. Syst. Sci*, 60:510–539, 2000]. The question is whether every finite metric space has a volume respecting embedding with distortion $O(\log n)$. Specifically, the problem is that when we consider sets of k vertices and k is large, then the proof no longer shows a distortion of $O(\log n)$. The current proof gives upper bounds on the distortion that grow at a rate of $\sqrt{k}$. Another related result is that of Satish Rao that can be used to give a different embedding that has $O(\log k)$ dependence on k, but a worse dependence on n.

Comment (A. Gupta): By a technique similar to Rao's, one can also get a bound of $O((\log n)^{3/2})$ for all k (no dependence on k); a proof is in Gupta's thesis.

6.2 Path into $\mathbb{R}^3$, volume-respecting (Uriel Feige)

Can the path P_n on n vertices be embedded via a contraction in $\mathbb{R}^3$ with only polylogarithmic distortion in the area of triangles? That is, for points $i < j < k$ we want the area of the triangle formed by the images of these points to be at least $(j-i)(k-j)/(\text{polylog } n)$. For embeddings in $\mathbb{R}^2$ the answer is negative, and for embeddings in $\mathbb{R}^{O(\log n)}$ the answer is positive.

6.3 Path into ℓ_2 (Nathan Linial)

Feige [U. Feige: Approximating the bandwidth via volume respecting embeddings, *J. Comput. Syst. Sci*, 60:510–539, 2000] introduced an interesting notion of volume to the field of discrete metric spaces, but other reasonable definition can be considered as well. Here is a problem illustrating one such alternative definition: Is there a map $\varphi : \{1, \ldots, n\} \to \ell_2$ such that

$$\text{area}(\varphi(i), \varphi(j), \varphi(k)) = \Theta((j-i)(k-j))$$

for every i, j, k, $n \geq k > j > i \geq 1$? (Unlike in Feige's definition, φ is *not* required to be non-expanding.)

7 Embeddings of special metrics

7.1 Levenstein metric into ℓ_1 (Piotr Indyk)

Let Σ be an "alphabet" set, and let Σ^* denote all strings with symbols from Σ. Let $G = (\Sigma^*, E)$ be an infinite undirected graph, such that $\{s, s'\} \in E$ iff s' can be obtained from s by inserting one symbol into s, deleting one symbol from s, or substituting one symbol in s by another symbol. Let M_Σ (*the Levenstein metric*) be the shortest path metric of G, and let $M_\Sigma^{\leq d}$ denote M_Σ restricted to the set of strings of length at most d.

Problem: Find the smallest value of $c = c(d)$ such that $M_\Sigma^{\leq d}$ can be embedded into ℓ_1 with distortion c.

Comments: The edit distance metric restricted to the set

$$\{010, 1010, 0110, 0101, 01010\}$$

is isometric to the shortest path metric over $K_{2,3}$ [Sofya Raskhodnikova, personal communication], and it follows that the Levenstein metric cannot be embedded into ℓ_1 with distortion better than $4/3$. This approach has been suggested by Michel Deza.

The ℓ_1 norm seems to be the best candidate for the "host" norm. This is due to the fact that $M_\Sigma^{\leq d}$ "contains" the Hamming metric $M_H^d = (\{0,1\}^d, D_H)$. Specifically, there exist an isometry from M_H into $M_\Sigma^{\leq d}$; e.g., for $\Sigma = \{1 \ldots 2d\}$ we can use the mapping f that for any argument $x \in \{0,1\}^d$ substitutes x_i by $2i + x_i$, for all $i = 1 \ldots d$. Unbounded Σ is not crucial—one can easily obtain a constant distortion embedding of M_H into $M_{\{0,1\}}^{\leq d}$ as well.

However, a low-distortion embedding of $D_\Sigma^{\leq d}$ into $\ell_p^{d'}$ for $p \neq 1$, with $d' = 2^{o(d)}$ (ideally, $d' = d^{O(1)}$) would also be interesting.

It is known that if, in addition to insertions, deletions and substitutions, one allows certain additional operations, then the resulting metric can be embedded into ℓ_1 with distortion $O(\log d \log^* d)$. See [G. Cormode and S. Muthukrishnan: The string edit distance matching problem with moves, *Proc. ACM–SIAM Sympos. Discr. Algo.* 2002] for the exact statement of the result (in particular, for the exact definition of the metric).

A major implication of a low-distortion embedding would be an approximate nearest neighbor data structure for the Levenstein metric using small (polynomial) space. For this application, it is *not* crucial for the embedding to be deterministic. In particular, a randomized embedding with very low probability of contraction but only constant probability of expansion is sufficient (see the talk by Piotr Indyk). The only non-trivial result for this nearest neighbor problem is an $O(1)$-approximate data structure, using polylogarithmic query time and n^{d^ε} space, for any constant $\varepsilon > 0$ [Indyk'02, unpublished]. The latter result does not use embeddings.

7.2 Fréchet metric into ℓ_∞ (Piotr Indyk)

Let C^d be the set of all polygonal chains in $\mathbb{R}^2$ consisting of d pieces. We represent each curve in C^d as a function $f\colon [0,1] \to \mathbb{R}^2$. For any two curves f, g, define

$$D_F(f,g) = \inf_{\pi:[0,1]\to[0,1]} \sup_{t\in[0,1]} \|f(\pi(t)) - g(t)\|_2$$

where π is continuous, monotone increasing, $\pi(0) = 0$, and $\pi(1) = 1$. (This is called the *Fréchet metric* or *dogkeeper's metric*.)

Problem I: Find the smallest $c = c(d)$ such that (C^d, D_F) can be embedded into $\ell_\infty^{d'}$ for a finite d' with distortion c.

Comments: Note that the set C^d is infinite. Thus the universality of ℓ_∞ results in an embedding into an infinite-dimensional space. The ℓ_∞ norm seems to be the best candidate for the host norm. This is because for any bounded set $S \subset \ell_\infty^d$, the metric (S, ℓ_∞) can be isometrically embedded into (C^{3d}, D_F) (easy to see).

For the practical purposes, we would like d' to be small (ideally, polynomial in d). Possibly, this could be helped by assuming that the curves are discrete. That is, one can consider the space C_I^d containing all polygonal chains with d segments, such that the endpoints of all segments are from $\{0, \ldots, I\}^2$. This leads to a discrete version of Problem I.

Discrete Problem I: Find the smallest $c = c(d, I)$ such that (C_I^d, D_F) can be embedded into $\ell_\infty^{d'}$ (for $d' = (d + \log I)^c$) with distortion c. (Of course, different (but small) c's bounding the distortion and the dimension are also interesting).

Comments: As in the case of the Levenstein metric, the approximate nearest neighbor problem is a major application of any embedding result for D_F. The only algorithm known for the latter problem has bounds similar to those for the Levenstein distance [P. Indyk: Approximate nearest neighbor algorithms for Fréchet distance via product metrics, Proc. 18th ACM Symposium on Computational Geometry, 2002].

The only previously known result with a similar flavor is the embedding of the *Hausdorff* metric over bounded subsets of ℓ_p^2, with constant distortion, into low-dimensional ℓ_∞ norm [M. Farach-Colton and P. Indyk: Approximate nearest neighbor algorithms for Hausdorff metrics via embeddings, Proc. 40th IEEE Symposium on Foundations of Computer Science, 1999].

7.3 Earth-mover distance (Piotr Indyk)

Consider the metric M_E^d defined over d-subsets of $\mathbb{R}^2$. For any two sets A, B, the distance $D_E(A, B)$ is defined as the minimum-weight bipartite matching between A and B; i.e.,

$$\min_{\pi: A \to B, \pi \text{ is one-to-one}} \sum_{a \in A} \|a - \pi(a)\|_2$$

Problem: Find the smallest $c = c(d)$ such that M_E^d can be embedded into ℓ_1 with distortion c.

Comments: If the metric is defined over subsets of $\mathbb{R}$, then it can be isometrically embedded into ℓ_1 (and vice versa). This uses the fact that the optimal matching is non-crossing (see the talk by Chandra Chekuri). Moreover, the EMD over s-element subsets of $(\{1 \ldots \Delta\}^k, \ell_p)$ can be embedded into ℓ_1 with distortion $O(k \log \Delta)$; this follows from [M. Charikar: Similarity estimation techniques from rounding, *Proc. 34th Annu. ACM Sympos. on Theory of Computing*, 2002].

A variation of the above metric (with sum replaced by max) is also of interest; no non-trivial result for this case is known.

7.4 Convex extensions (Yuval Rabani)

Let X be a set, $|X| = n$, and let $Y \subset X$, $|Y| = k$. Let d be a metric on Y.

A semi-metric space (X, δ) is a *metric extension* of (Y, d) if $\delta(x, y) = d(x, y)$ for all $x, y \in Y$. It is a 0-*extension* if for every $x \in X \setminus Y$ there is a $y \in Y$ with $\delta(x, y) = 0$. It is a *convex extension* if there is a mapping $\varphi: X \to \Delta^{k-1}$ (where Δ^{k-1} denotes a $(k-1)$-dimensional simplex) such that Y is mapped bijectively onto the vertices of Δ^{k-1} and $d(x, y)$ is the earth-mover distance between $\varphi(x)$ and $\varphi(y)$. Here each point $p \in \Delta^{k-1}$ can be uniquely written as a convex combination of the vertices of Δ^{k-1}, and we interpret it as a probability distribution $(p_i : i \in Y)$ over Y, where p_i is the coefficient of the vertex $\varphi(i)$ in that convex combination. In the particular case when $d(i, j) = 1$ for all $i, j \in Y$, we simply require $\delta(x, y) = \frac{1}{2}\|\varphi(x) - \varphi(y)\|_1$ for all $x, y \in X$. For (Y, d) arbitrary, $\delta(x, y)$ should be the minimum cost of converting the distribution $\varphi(x)$ to the distribution $\varphi(y)$ by moving mass between points, where a move between i and j costs $d(i, j)$ per unit of moved mass; that is,

$$\delta(x,y) = \min\Big\{ \sum_{i,j\in Y} \varepsilon_{ij} d(i,j) : \varphi(y)_i = \varphi(x)_i - \sum_j \varepsilon_{ij} + \sum_j \varepsilon_{ji} \text{ for all } i,\ \varepsilon_{ij} \geq 0 \Big\}.$$

Questions: Embed metric or convex extensions into convex combinations of 0-extensions. What is the best distortion? The minimum expansion in the worst case? For:

- *d uniform (all distances are 1), δ a convex extension.*
  Known: There is a tight bound of 12/11 on the minimum expansion for $k = 3$. No other tight bounds are known. Asymptotically in k, the bound is between 8/7 and 1.3438. There are better bounds for small values of k.

- *d arbitrary, δ a convex extension.*
  Known: There is a lower bound of $\Omega(\sqrt{\log k})$ and an upper bound of $O(\log k)$ on the minimum expansion. The upper bound is also the best known for convex extensions. (Notice that in the case of metric extensions, the question of distortion does not make sense, because it is easy to construct extensions that require arbitrarily large contraction.)

Conjecture: For convex extensions, the minimum expansion is a constant. (The best known lower bound is that for uniform d.)

- *d arbitrary, δ a metric extension.*
  Known: There is an upper bound of $O(\log k \log\log k)$. The best lower bound is that for the uniform case.

Embeddings of convex extensions where $x \in X \setminus Y$ can be mapped only to $y \in Y$ with $\varphi(x)_y > 0$ are also interesting.

Comment (C. Chekuri): Yuval Rabani's problem is essentially asking for the integrality gap of an LP relaxation in the paper [C. Chekuri, S. Khanna, J. Naor, and L. Zosin: Approximation Algorithms for the Metric Labeling Problem via a New Linear Programming Formulation, Proc. ACM–SIAM Sympos. Discr. Algo. 2001, pages 109–118] for the 0-extension problem. There are several open questions related to the more general metric labeling problem in the slides of my talk on this at the workshop.

8 Graph representations

8.1 Explicit graphs with high sphericity (Nathan Linial)

An embedding $\varphi: V(G) \to \ell_2^d$ is called a *proximity map* for G if

$$\|\varphi(x) - \varphi(y)\| \leq 1 \text{ if and only if } x \text{ and } y \text{ are adjacent.}$$

(The smallest d such that G has a proximity map into ℓ_2^d is also called the *sphericity* of G in the literature.) It is known that every graph has a proximity map in $d = n-1$ dimensions, but most graphs do not have such a map unless $d \geq \Omega(n)$. In particular, the complete bipartite graph $K_{n,n}$ requires $d \geq n$; see [J. Reiterman, V. Rödl, and E. Šiňajová: Embeddings of graphs in Euclidean spaces, *Discrete Comput. Geom.* 4: 349–364, 1989] for references and related results. Can you find other explicit families of graphs that require $d = \Omega(n)$?