Embeddings of finite metric spaces in Euclidean space: a probabilistic view

Yuval Peres

May 11, 2006

Talk based on work joint with:
Assaf Naor, Oded Schramm and Scott Sheffield

Definition: An invertible mapping $f : X \to Y$, where (X, d_X) and (Y, d_Y) are metric spaces, is a C-embedding if there exists a number $r > 0$ such that for all $x, y \in X$

$$r \cdot d_X(x,y) \leq d_Y(f(x), f(y)) \leq Cr \cdot d_X(x,y).$$

The infimum of numbers C such that f is a C-embedding is called the distortion of f and is denoted by $dist(f)$. Equivalently, $dist(f) = ||f||_{\mathrm{Lip}}||f^{-1}||_{\mathrm{Lip}}$ where

$$||f||_{\mathrm{Lip}} = \sup \left\{ \frac{d_Y(f(x), f(y))}{d_X(x,y)} : x, y \in X, x \neq y \right\}.$$

Example

Theorem: (Enflo, 1969) Let $\Omega_k = \{0,1\}^k$ be the k-dimensional hypercube with ℓ_1 metric, then any $f : \Omega_k \to L^2$ has distortion at least $\sqrt{k}$.

Remark: This is tight, as can be easily seen by taking the identity function from Ω_k to ℓ_2^k. Enflo's proof is algebraic, and is hence fragile: if one edge of the cube is removed, the proof breaks down.

Theorem: (Bourgain, 1985) Every n-point metric space (X, d) can be embedded in an Euclidean space with an $O(\log n)$ distortion.

Remark: Any embedding of an expander graph family into Euclidean space has distortion at least $c \log n$ (Linial, London and Rabinovich, 1995).

Proof idea for Bourgain's theorem: For each cardinality $k < n$ which is a power of 2, randomly pick $\alpha \log n$ sets $A \subset V(G)$ independently by including each $x \in X$ with probability $1/k$. We have drawn $O(\log^2 n)$ sets $A_1, \ldots, A_{O(\log^2 n)}$. Map every vertex $x \in X$ to the vector

$$\frac{1}{\log n}(d(x, A_1), d(x, A_2), \ldots).$$

This mapping has distortion $O(\log n)$.

Theorem: (Bourgain, 1986) There is no bounded distortion embedding of the infinite binary tree into a Hilbert space.
More precisely, any embedding of a binary tree of depth M and $n = 2^{M+1} - 1$ vertices into a Hilbert space has distortion $\Omega(\sqrt{\log M}) = \Omega(\sqrt{\log \log n})$.

The Markov type of metric spaces

A Markov chain $\{Z_t\}_{t=0}^{\infty}$ with transition probabilities $p_{ij} := \Pr(Z_{t+1} = j \mid Z_t = i)$ on the state space $\{1, \dots, n\}$ is stationary if $\pi_i := \mathbf{P}(Z_t = i)$ does not depend on t and it is (time) reversible if $\pi_i p_{ij} = \pi_j p_{ji}$ for every $i, j \in \{1, \dots, n\}$.

Definition (Ball 1992): Given a metric space (X, d) we say that X has Markov type 2 if there exists a constant $M > 0$ such that for every stationary reversible Markov chain $\{Z_t\}_{t=0}^{\infty}$ on $\{1, \dots, n\}$, every mapping $f : \{1, \dots, n\} \to X$ and every time $t \in \mathbf{N}$,

$$\mathbf{E} d(f(Z_t), f(Z_0))^2 \leq M^2 t \mathbf{E} d(f(Z_1), f(Z_0))^2.$$

Theorem: (Ball 1992) **R** has Markov type 2 with constant $M = 1$.

Proof: Let $P = (p_{ij})$ be the transition matrix of the Markov chain. Time reversibility is equivalent to the assertion that P is a self-adjoint operator in $L^2(\pi)$, hence $L^2(\pi)$ has an orthogonal basis of eigenfunctions of P with real eigenvalues. Also, since P is a stochastic matrix $\| Pf \|_\infty \leq \| f \|_\infty$ and thus if λ is an eigenvalue of P then $|\lambda| \leq 1$.
We have

$$\mathbf{E} d(f(Z_t), f(Z_0))^2 = \sum_{i,j} \pi_i p_{ij}^{(t)} [f(i) - f(j)]^2 = 2\langle (I - P^t) f, f \rangle ,$$

and also

$$\mathbf{E} d(f(Z_1), f(Z_0))^2 = 2\langle (I - P) f, f \rangle .$$

So we are left to prove that,

$$\langle (I - P^t)f, f\rangle \leq t\langle (I-P)f, f\rangle\,.$$

Indeed, if f is an eigenfunction with eigenvalue λ this reduces to proving $(1-\lambda^t) \leq t(1-\lambda)$. Since $|\lambda| \leq 1$, this reduces to

$$1 + \lambda + \cdots + \lambda^{t-1} \leq t\,,$$

which is obviously true. For any other f take $f = \sum_{j=1}^n a_j f_j$ where $\{f_j\}$ is an orthonormal basis of eigenfunctions,

$$\begin{aligned}\langle (I-P^t)f, f\rangle &= \sum_{j=1}^n a_j^2 \langle (I-P^t)f_j, f_j\rangle \leq \sum_{j=1}^n a_j^2 t\langle (I-P)f_j, f_j\rangle \\ &= t\langle (I-P)f, f\rangle\,.\square\end{aligned}$$

Corollary: Any Hilbert space has Markov type 2 with constant $M = 1$.

Corollary: Any embedding of the hypercube $\{0,1\}^k$ into Hilbert space has distortion at least $\sqrt{k}/4$.

Proof: Let $\{X_j\}$ be the simple random walk on the hypercube. It is easy to see that

$$\mathbf{E}d(X_0, X_j) \geq \frac{j}{2} \qquad \forall j \leq k/4\,,$$

which by Jensen's inequality implies $\mathbf{E}d^2(X_0, X_j) \geq j^2/4$. Let $f : \{0,1\}^k \to L^2$ be a map. Assume without loss of generality that f is non-expanding mapping, i.e., $\| f \|_{\mathrm{Lip}} = 1$ (otherwise take $f/\| f \|_{\mathrm{Lip}}$). Since L^2 has Markov type 2 with constant $M = 1$, for any j, we have

$$\mathbf{E}d^2[f(X_0), f(X_j)] \leq j\,.$$

Take $j = k/4$, together this yields

$$\| f^{-1} \|_{\mathrm{Lip}} \geq \sqrt{k}/4\,.$$

□

Corollary: Any embedding of an (n, d, λ)-expander family has distortion at least $C_{d,\lambda} \log n$.

Proof: Let $\{X_j\}$ be the simple random walk on the expander with transition matrix P. Let $g = 1 - \lambda$, take $\alpha > 0$ such that $g\alpha < 1$ and take $t = \alpha \log n$, it can be shown that

$$P^t(x, y) \leq 2e^{-(1-\lambda)\alpha \log n} .$$

Fix $\gamma > 0$ small enough such that $d^\gamma e^{-(1-\lambda)\alpha} < 1$. We wish to show that up to time $t = \alpha \log n$, the random walk on the expander has positive speed. Indeed, for any $x \in V$, since the ball $B(x, \gamma \log n)$ of radius $\gamma \log n$ around x has at most $d^{\gamma \log n}$ vertices it follows that

$$\mathbf{P}_x[X_t \in B(x, \gamma \log n)] \leq d^{\gamma \log n} 2e^{-(1-\lambda)\alpha \log n} \to 0 .$$

This in turn implies that for large enough n

$$\mathbf{E} d^2(X_0, X_t) > \frac{\gamma^2 \log^2 n}{2} .$$

Let $f : V \to L^2$, and assume without loss of generality that $\| f \|_{\mathrm{Lip}} = 1$ (otherwise take $f / \| f \|_{\mathrm{Lip}}$). We have proved in the previous theorem that

$$\mathbf{E} d(f(X_t), f(X_0))^2 \leq (1 + \lambda + \lambda^2 + \cdots + \lambda^{t-1}) \mathbf{E} d(f(Z_1), f(Z_0))^2 \,.$$

This immediately implies that

$$\mathbf{E} d^2(f(X_0), f(X_t)) \leq \frac{1}{1-\lambda} \,.$$

Together this implies

$$\| f^{-1} \|_{\mathrm{Lip}} \geq \sqrt{1-\lambda} \gamma \log n \,.$$

$\square$

In similar ways one can prove that if a family of graphs have all degrees at least 3 and girth (size of smallest cycle) at least g then any embedding into Hilbert space has distortion at least $\Omega(\sqrt{g})$. So if the girth is at least $c \log n$ we cannot embed with distortion smaller than $c'\sqrt{\log n}$, but we can with distortion $O(\log n)$ (Bourgain's theorem).

Open question: What is the minimal possible distortion for an embedding of an n vertex graph of girth $g = c\log(n)$ in Euclidean space?

Theorem: (Naor, P., Sheffield, Schramm, 2004) L_p for $p > 2$, trees and hyperbolic groups have Markov type 2.

Open question: Do planar graphs (with graph distance) have Markov type 2?

Key ideas in proofs

Stationary reversible Markov chains in $\mathbf{R}$ (and more generally, in any normed space) are difference of two martingales, a forward martingale, and a backward martingale; the squared norms of the martingale increments can be bounded using the original increments. For martingales, powerful inequalities due to Doob and Pisier are available.

On a tree, the length of a path can be bounded by twice the difference between the maximum and the minimum distance to the root along the path.

We will use a decomposition of a stationary reversible Markov chain into forward and backward martingales (inspired by a decomposition due to Lyons and Zhang for stochastic integrals.)

A central Lemma

Lemma: Let $\{Z_t\}_{t=0}^{\infty}$ be a stationary time reversible Markov chain on $\{1, \ldots, n\}$ and $f : \{1, \ldots, n\} \to \mathbf{R}$. Then, for every time $t > 0$,

$$\mathbf{E} \max_{0 \le s \le t} [f(Z_s) - f(Z_0)]^2 \le 15t\mathbf{E}[f(Z_1) - f(Z_0)]^2 \,.$$

Proof: Let $P : L^2(\pi) \to L^2(\pi)$ be the Markov operator, i.e. $(Pf)(i) = \mathbf{E}[f(Z_{s+1}) | Z_s = i]$. For any $s \in \{0, \ldots, t-1\}$ let

$$D_s = f(Z_{s+1}) - (Pf)(Z_s) \,,$$

and

$$\widetilde{D}_s = f(Z_{s-1}) - (Pf)(Z_s) \,.$$

The first are martingale differences with respect to the natural filtration of $Z_1, \ldots, Z_t$, and the second, because of reversibility are martingale differences with respect to the natural filtration on $Z_t, \ldots, Z_1$.

Subtracting,

$$f(Z_{s+1}) - f(Z_{s-1}) = D_s - \widetilde{D}_s \, .$$

Thus for any m

$$f(Z_{2m}) - f(Z_0) = \sum_{k=1}^{m} D_{2k-1} - \sum_{k=1}^{m} \widetilde{D}_{2k-1} \, .$$

So,

$$\begin{aligned} \max_{0 \le s \le t} f(Z_s) - f(Z_0) \quad &\le \quad \max_{m \le t/2} \sum_{k=1}^{m} D_{2k-1} + \max_{m \le t/2} \sum_{k=1}^{m} -\widetilde{D}_{2k-1} \\ &+ \quad \max_{\ell \le t/2} \left| f(Z_{2\ell+1}) - f(Z_{2\ell}) \right| . \end{aligned}$$

Take squares and use the fact $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$, which is implied by the Cauchy-Schwarz inequality, to get

$$\begin{aligned}
\max_{0\leq s\leq t} |f(Z_s) - f(Z_0)|^2 \leq \quad & 3 \quad \max_{m\leq t/2} \Big| \sum_{k=1}^{m} D_{2k-1} \Big|^2 \\
+ \quad & 3 \max_{m\leq t/2} \Big| \sum_{k=1}^{m} \widetilde{D}_{2k-1} \Big|^2 \\
+ \quad & 3 \sum_{\ell\leq t/2} \Big| f(Z_{2\ell+1}) - f(Z_{2\ell}) \Big|^2 .
\end{aligned}$$

We will use Doob's L^2 maximum inequality for martingales (see, e.g., Durrett 1996)

$$\mathbf{E} \max_{0 \le s \le t} M_s^2 \le 4\, \mathbf{E}|M_t|^2 .$$

Consider

$$M_{s+1} = \sum_{j \le s, j \text{ odd}} D_j .$$

Since M_s is still a martingale, we have

$$\begin{aligned} \mathbf{E} \max_{0 \le s \le t} |f(Z_s) - f(Z_0)|^2 \le \quad & 12 \quad \mathbf{E}\Big| \sum_{k=1}^{\lfloor t/2 \rfloor} D_{2k-1} \Big|^2 \\ + \quad & 12\mathbf{E}\Big| \sum_{k=1}^{\lfloor t/2 \rfloor} \widetilde{D}_{2k-1} \Big|^2 \\ + \quad & 3 \sum_{0 \le \ell \le t/2} \mathbf{E}\Big| f(Z_{2\ell+1}) - f(Z_{2\ell}) \Big|^2 . \end{aligned}$$

Denote $V = \mathbf{E}[|f(Z_1) - f(Z_0)|^2]$, and notice that

$$D_0 = f(Z_1) - f(Z_0) - \mathbf{E}[f(Z_1) - f(Z_0) \mid Z_0]\,,$$

which implies that D_0 is orthogonal to $\mathbf{E}[f(Z_1) - f(Z_0) \mid Z_0]$ in $L^2(\pi)$. So, by the Pythagorian law, for any s we have $\mathbf{E}[D_s^2] = \mathbf{E}[D_0^2] \le V$. Summing everything up gives

$$\mathbf{E} \max_{0 \le s \le t} |f(Z_s) - f(Z_0)|^2 \le 6tV + 6tV + 3(t/2+1)V \le 15tV\,,$$

which concludes the proof of the Lemma. $\square$

Trees have Markov Type 2

Theorem: (Naor, P., Sheffield, Schramm, 2004) Trees have Markov Type 2.

Proof: Let T be a weighted tree, $\{Z_j\}$ be a reversible Markov chain on $\{1, \ldots, n\}$ and $F : \{1, \ldots, n\} \to T$. Choose an arbitrary root and set for any vertex v, $\psi(v) = d(root, v)$. If $v_0, \ldots, v_t$ is a path in the tree, then

$$d(v_0, v_t) \leq \max_{0 \leq j \leq t} \Big(|\psi(v_0) - \psi(v_j)| + |\psi(v_t) - \psi(v_j)| \Big),$$

since choosing the closest vertex to the root on the path yields equality.

Let $X_j = F(Z_j)$. Connect X_i to X_{i+1} by the shortest path for any $0 \leq i \leq t-1$ to get a path between X_0 and X_t. Since now the closest vertex to the root can be on any of the shortest paths between X_j and X_{j+1}, we get

$$d(X_0, X_t) \leq \max_{0 \leq j < t} \Big(|\psi(X_0) - \psi(X_j)| + |\psi(X_t) - \psi(X_j)| + 2d(X_j, X_{j+1}) \Big) .$$

Square, and use Cauchy-Schwarz again,

$$\begin{aligned} d(X_0, X_t)^2 \leq & \quad 3 \max_{0 \leq j \leq t} \left(|\psi(X_0) - \psi(X_j)|^2 + |\psi(X_t) - \psi(X_j)|^2 \right) \\ & + \quad 12 \sum_{0 \leq j < t} d^2(X_j, X_{j+1}) . \end{aligned}$$

By our central Lemma with $f = \psi F$ we get,

$$\mathbf{E}d(X_0, X_t)^2 \le 90t\mathbf{E}|\psi(X_0) - \psi(X_1)|^2 + 6 \sum_{0 \le j \le t} \mathbf{E}d^2(X_j, X_{j+1})\,.$$

Since in any metric space $|\psi(X_1) - \psi(X_0)| \le d(X_0, X_1)$ and since the Markov chain is stationary we have $\mathbf{E}d(X_0, X_1) = \mathbf{E}d(X_j, X_{j+1})$ for any j. So

$$\mathbf{E}d(X_0, X_t)^2 \le 96t\mathbf{E}d(X_0, X_1)^2\,,$$

which concludes our proof. □