we see that $\alpha_i(n_k) - \beta_i(n_k)$ is a simple random walk on $\mathbb{Z}$ with an absorbing state at 0. Recurrence of simple random walk on $\mathbb{Z}$ implies that $\lim_{n\to\infty} \alpha_i(n) - \beta_i(n) = 0$ a.s. Since this is true for all i, we conclude that $\mathbf{P}[X_n \neq Y_n] \to 0$ as $n \to \infty$. Finally, since h_1 and h_2 are arbitrary, f must be constant.

§7. Embeddings of Finite Metric Spaces.

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 $x, y \in X$

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

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, $dist(f) = \|f\|_{\text{Lip}} \|f^{-1}\|_{\text{Lip}}$, where

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

We will be interested in embeddings of finite metric spaces and in application of Markov type 2 results to prove lower bounds on distortions of embeddings of certain spaces. We will see that the any embedding of the hypercube $\{0,1\}^k$ in Hilbert space has distortion at least $c\sqrt{k}$, for some $c > 0$ (Enflo, 1969). In Exercise 2 we will show by Markov type arguments that any embedding of an expander family into Hilbert space has distortion at least $\Omega(\log n)$. This was originally shown by Linial, London and Rabinovich (1995) to prove that a theorem of Bourgain (1985), stating that any metric on n points can be embedded in $\ell_p^{\log n}$ with distortion $O(\log n)$, is tight.

We first prove a dimension reduction lemma due to Johnson and Lindenstrauss (1984).

LEMMA 7.1. *For any $0 < \epsilon < 1/2$ and $v_1, \ldots, v_n \in \mathbb{R}^n$ with Euclidean metric, there exists a linear map $A : \mathbb{R}^n \to \mathbb{R}^k$ where $k = O(\log n / \epsilon^2)$, with distortion at most $1 + \epsilon$ on the n point space $\{v_1, \ldots, v_n\}$.*

Proof. Let $A = \frac{1}{\sqrt{k}} (X_i^{(j)})_{1 \leq i \leq n, 1 \leq j \leq k}$ be an $n \times k$ matrix where the entries $X_i^{(j)}$ are independent standard normal $N(0,1)$ random variables. We prove that with positive probability this map has distortion at most $1 + \epsilon$. For any $i \neq j$, let $u = \frac{v_i - v_j}{\|v_i - v_j\|} \in S^{n-1}$, and denote $u = (u_1, \ldots, u_n)$. Clearly,

$$uA = \frac{1}{\sqrt{k}} \left(\sum_{i=1}^{n} u_i X_i^{(1)}, \cdots, \sum_{i=1}^{n} u_i X_i^{(k)} \right) .$$

So

$$\|uA\|^2 = \frac{1}{k}\sum_{j=1}^{k}\left(\sum_{i=1}^{n} u_i X_i^{(j)}\right)^2 .$$

Note that for any j the sum $\sum_{i=1}^{n} u_i X_i^{(j)}$ is distributed as a standard normal random variable with mean 0, and since $\sum_{i=1}^{n} u_i^2 = 1$, the variance is 1. So $\|uA\|^2$ is distributed as $\frac{1}{k}\sum_{j=1}^{k} Y_j^2$, where $Y_1, \ldots, Y_k$ are independent standard normal $N(0,1)$ random variables. We wish to show that uA is concentrated around its mean. To achieve that we compute the moment generating function of Y^2 where $Y \sim N(0,1)$. For any real $\lambda < 1/2$ we have

$$\mathbf{E}e^{\lambda Y^2} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\lambda y^2} e^{-y^2/2} dy = \frac{1}{\sqrt{1-2\lambda}},$$

and using Taylor expansion we get

$$\begin{aligned}\varphi(\lambda) &= |\log \mathbf{E}e^{\lambda(Y^2-1)}| = |-\frac{1}{2}\log(1-2\lambda) - \lambda| \\ &= \sum_{k=2}^{\infty}\frac{2^{k-1}\lambda^k}{k} \le 2\lambda^2(1 + 2\lambda + (2\lambda)^2 + \cdots) = \frac{2\lambda^2}{1-2\lambda}.\end{aligned}$$

Now,

$$\mathbf{P}[\|uA\|^2 > 1+\epsilon] = \mathbf{P}\left[e^{\lambda\sum_{i=1}^{k}(Y_j^2-1)} > e^{\lambda\epsilon k}\right] \le e^{-\lambda\epsilon k}e^{k\varphi(\lambda)} \le \exp\left(-\lambda\epsilon k + \frac{2\lambda^2 k}{1-2\lambda}\right).$$

Taking $\lambda = \epsilon/4$ and $k \ge 24\log n/\epsilon^2$, and recalling that $\epsilon < 1/2$, yields

$$\mathbf{P}[\|uA\|^2 > 1+\epsilon] \le \exp(-\epsilon^2 k/12) \le n^{-2}.$$

One can prove similarly that

$$\mathbf{P}[\|uA\|^2 < 1-\epsilon] \le n^{-2}.$$

Since we have $\binom{n}{2}$ pairs of vectors v_i, v_j we showed that with positive probability, for all $i \ne j$,

$$(1-\epsilon)\|v_i - v_j\| \le \|v_i A - v_j A\| \le (1+\epsilon)\|v_i - v_j\|,$$

which implies that the distortion of A is no more than $1+\epsilon$. ∎

REMARK 7.2. From algorithmic perspective it is important to achieve Lemma 7.1 using i.i.d., ± 1 with probability 1/2, random variables as our $X_i^{(j)}$. This is in fact possible for any random variable X for which there exists a constant $C > 0$ such that $\mathbf{E}e^{\lambda X} \le e^{C\lambda^2}$

(for $X = \pm 1$ with probability 1/2 we have $\mathbf{E}e^{\lambda X} = \cosh(\lambda) \leq e^{\lambda^2/2}$) by the following argument:

Let $Y = \sum_{i=1}^k u_i X_i$ with $\sum_{j=1}^k u_j^2 = 1$ and let Z be distributed $N(0,1)$ independently of $\{X_i\}$. Recall that for all real α we have $\mathbf{E}e^{\alpha Z} = e^{\alpha^2/2}$. Since Y and Z are independent, using Fubini's Theorem we get that for any $\lambda < \frac{C}{2}$

$$\mathbf{E}e^{\lambda Y^2} = \mathbf{E}e^{\frac{(\sqrt{2\lambda}Y)^2}{2}} = \mathbf{E}e^{\sqrt{2\lambda}YZ} = \mathbf{E}e^{\sum_{i=1}^k \sqrt{2\lambda}u_i X_i Z} = \mathbf{E}\,\mathbf{E}\Big[e^{\sum_{i=1}^k \sqrt{2\lambda}u_i X_i Z} \mid Z\Big]$$
$$\leq \mathbf{E}e^{C\sum_{i=1}^k \lambda u_i^2 Z^2} = \mathbf{E}e^{C\lambda Z^2} = \frac{1}{\sqrt{1-2C\lambda}},$$

and the rest of the argument is the same as Lemma 7.1.

THEOREM 7.3. *(Bourgain, 1985) Every n-point metric space (X,d) can be embedded in an $O(\log n)$ -dimensional Euclidean space with an $O(\log n)$ distortion.*

Proof. We follow Linial, London and Rabinovich (1995). Let $\alpha > 0$ be determined later. For each cardinality $k < n$ which is a power of 2, randomly pick $\alpha \log n$ sets $A \subset X$ 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

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

Denote this mapping by f. We will show this mapping to $\ell_2^{O(\log^2 n)}$ has almost surely $O(\log n)$ distortion, and using Lemma 7.1 this yields the required result.

It is easy to observe that this map is not expanding. By the triangle inequality, for any $x, y \in X$ and any $A_i \subset X$ we have $|d(x,A_i) - d(y,A_i)| \leq d(x,y)$, so

$$\|f(x) - f(y)\|_2^2 \leq \frac{1}{\log^2 n} \sum_{i=1}^{\alpha \log^2 n} |d(x,A_i) - d(y,A_i)|^2 \leq \alpha d(x,y)^2.$$

For the lower bound, let $B(x,\rho) = \{y \in X | d(x,y) \leq \rho\}$ and $B^o(x,\rho) = \{y \in X | d(x,y) < \rho\}$ denote the closed and open balls of radius ρ centered at x. Consider two points $x \neq y \in X$. Let $\rho_0 = 0$, and let ρ_t be the least radius ρ for which both $|B(x,\rho)| \geq 2^t$ and $|B(y,\rho)| \geq 2^t$. We define ρ_t as long as $\rho_t < \frac{1}{4}d(x,y)$, and let $\hat{t}$ be the largest such index. Also let $\rho_{\hat{t}+1} = \frac{d(x,y)}{4}$. Observe that $B(y,\rho_j)$ and $B(x,\rho_i)$ are always disjoint.

Notice that $A \cap B^o(x,\rho_t) = \varnothing \iff d(x,A) \geq \rho_t$, and $A \cap B(y,\rho_{t-1}) \neq \varnothing \iff d(y,A) \leq \rho_{t-1}$. Therefore, if both conditions hold, then $|d(y,A) - d(x,A)| \geq \rho_t - \rho_{t-1}$.

Let us assume that $|B^o(x,\rho_t)| < 2^t$ (otherwise we argue for y). On the other hand, $|B(y,\rho_{t-1})| \geq 2^{t-1}$. Let $k = 2^t$ and let $A \subset X$ be chosen randomly by including each $x \in X$ with probability $1/k$. We have

$$\mathbf{P}[A \text{ misses } B^o(x,\rho_t)] \geq (1-2^{-t})^{2^t} \geq \frac{1}{4},$$

and

$$\mathbf{P}[A \text{ hits } B(y,\rho_{t-1})] \geq 1-(1-2^{-t})^{2^{t-1}} \geq 1-e^{-1/2} \geq \frac{1}{2}.$$

Since these events are independent, such an A has probability at least $\frac{1}{8}$ to both intersect $B(y,\rho_{t-1})$ and miss $B^o(x,\rho_t)$. Since for each k we choose $\alpha \log n$ such sets, by Theorem 1.1, the probability that less than $\frac{\alpha \log n}{16}$ of them have the previous property is less than

$$e^{-2(\alpha \log n/16)^2/(\alpha \log n)} \leq n^{-\alpha/128} \leq n^{-5},$$

by choosing α such that $\alpha/128 > 5$. So with probability tending to 1, for any $x, y \in X$ and k we have at least $\alpha \log n/16$ sets which satisfy the condition. Summing it up gives

$$\|f(x)-f(y)\|_2^2 \geq \frac{1}{\log^2 n} \sum_{i=1}^{\hat{t}+1} \frac{\alpha \log n}{16} (\rho_i - \rho_{i-1})^2.$$

Since $\sum_{i=1}^{\hat{t}+1}(\rho_i - \rho_{i-1}) = \rho_{\hat{t}+1} = \frac{d(x,y)}{4}$, we have

$$\|f(x)-f(y)\|_2^2 \geq \frac{\alpha}{16 \log n}\left(\frac{d(x,y)}{4(\hat{t}+1)}\right)^2 (\hat{t}+1) \geq \frac{\alpha d(x,y)^2}{256(\hat{t}+1)\log n} \geq \frac{\alpha d(x,y)^2}{256 \log^2 n},$$

hence the distortion of f is $O(\log n)$ with probability tending to 1. ∎

Proposition 7.4. *(Enflo, 1969) There exists $c > 0$ such that any embedding of the hypercube $\{0,1\}^k$ in Hilbert space has distortion at least $c\sqrt{k}$.*

Proof. Recall that in Exercise 1 of Chapter 3 we proved that if $\{X_j\}$ is a simple random walk in the hypercube, then

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

Take $j = \frac{k}{4}$. By Jensen's inequality, $\mathbf{E}d^2(X_0, X_{k/4}) \geq k^2/64$. Now let $f: \{0,1\}^k \to L^2$ be a map. Assume without loss of generality that f is a non-expanding, i.e., $\|f\|_{\text{Lip}} = 1$ (otherwise, take $f/\|f\|_{\text{Lip}}$). By Theorem 3.1 it follows that L^2 has Markov type 2 with constant $M = 1$, so,

$$\mathbf{E}d^2\Big(f(X_0), f(X_{k/4})\Big) \leq k.$$

We conclude

$$\|f^{-1}\|_{\text{Lip}}^2 k \geq \|f^{-1}\|_{\text{Lip}}^2 \mathbf{E}d^2(f(X_0), f(X_{k/4})) \geq \mathbf{E}d^2(X_0, X_{k/4}) \geq k^2/64,$$

hence $\|f^{-1}\|_{\text{Lip}} \geq \frac{\sqrt{k}}{8}$, which implies the result. ∎

REMARK 7.5. Enflo's original proof gives $c = 1$. See Exercise 1 for the proof of this fact.

We now prove a theorem of Bourgain (1986).

THEOREM 7.6. *Any embedding of a binary tree of depth M and $n = 2^{M+1} - 1$ vertices into a Hilbert space has distortion $\Omega(\sqrt{\log M}) = \Omega(\sqrt{\log\log n})$.*

REMARK 7.7. See Exercise 3 for an embedding with distortion $O(\sqrt{\log M})$.

We first prove two lemmas.

LEMMA 7.8. *Let $M = 2^m$ for $m \geq 1$ and $y_0, \ldots, y_n \in \mathbb{R}$, then*

$$\sum_{i=1}^{M}(y_i - y_{i-1})^2 = \frac{(y_M - y_0)^2}{M} + \sum_{k=1}^{m}\frac{1}{2^k}\sum_{j=1}^{2^{m-k}}(y_{j2^k} - 2y_{(2j-1)2^{k-1}} + y_{(j-1)2^k})^2 .$$

Proof. This can be proved by induction on m, however, we will prove it using Parseval's identity. Consider the Haar orthonormal basis of $\mathbb{R}^M$ which is defined by the following vectors: for any $1 \leq k \leq m$ and any $1 \leq j \leq 2^{m-k}$ let $I(k;j)$ denote the set of indices $\{(j-1)2^k + 1, \ldots, j2^k\}$ and define

$$\psi_{I(k;j)}(i) = \begin{cases} \frac{1}{2^{k/2}}, & (j-1)2^k < i \leq (2j-1)2^{k-1}; \\ -\frac{1}{2^{k/2}}, & (2j-1)2^{k-1} < i \leq j2^k. \end{cases}$$

Together with the vector $\psi_1 = \frac{1}{\sqrt{M}}(1, \ldots, 1)$ this gives 2^m orthonormal vectors in $\mathbb{R}^M$. Now define $z \in \mathbb{R}^M$ by $z_i = y_i - y_{i-1}$, so the LHS of the lemma becomes $\sum_{i=1}^{M} z_i^2$, which, by Parseval's identity, is

$$\langle z, z\rangle = \langle z, \psi_1\rangle^2 + \sum_{k=1}^{m}\sum_{j=1}^{2^{m-k}}\langle z, \psi_{I(k;j)}\rangle^2 ,$$

which can easily be seen to be the RHS of the lemma. ∎

LEMMA 7.9. *Let $M = 2^m$, and suppose that $Y_0, Y_1, \ldots$ is a function of a Markov chain taking values in Hilbert space. For any $1 \leq k \leq m$ and $1 \leq j \leq 2^{m-k}$ let $r = (2j-1)2^{k-1}$ and let $\tilde{Y}(k;j)$ denote the random process which is equal to $\{Y_t\}$ for time $t \leq r$ and evolves independently for time $t > r$. Write $\mathcal{A}_{i=1}^{M}(\cdot) = \frac{1}{M}\sum_{i=1}^{M}(\cdot)$ for the averaging operator. Then*

$$\mathbf{E}\left[\mathcal{A}_{i=1}^{M}\|Y_i - Y_{i-1}\|^2\right] \geq \mathbf{E}\left[\frac{1}{2}\sum_{k=1}^{m}\mathcal{A}_{j=1}^{2^{m-k}}\frac{\|Y_{j2^k} - \tilde{Y}_{j2^k}(k;j)\|^2}{2^{2k}}\right] .$$

Proof. Since all the distances in the lemma are squared, we can assume without loss of generality that $\{Y_t\}$ is real valued. Let k, j be as in the lemma. Write $\mathbf{E}_r(\cdot) = \mathbf{E}[\cdot \mid Y_r]$

and $\tilde{Y} = \tilde{Y}(k;j)$. Let $t > r$ and denote $\mu_r = \mathbf{E}_r[Y_t]$. Note that by the definition of $\tilde{Y}$, we have that Y_t and $\tilde{Y}_t$ are independent given Y_r, and so $\mathbf{E}_r[Y_t\tilde{Y}_t] = \mathbf{E}_r[Y_t]\mathbf{E}_r[\tilde{Y}_t]$. Also, since Y_t has the same distribution as $\tilde{Y}_t$, we have

$$\mathbf{E}_r|Y_t - \tilde{Y}_t|^2 = \mathbf{E}_r|(Y_t - \mu_r) - (\tilde{Y}_t - \mu_r)|^2 = 2\mathbf{E}_r(Y_t - \mu_r)^2 \le 2\mathbf{E}_r(Y_t - \lambda_r)^2 ,$$

for any λ_r which is Y_r-measurable. The last inequality follows from the fact that $\mathbf{E}_r(Y_t - \mu_r)^2$ is the squared length of the projection of Y_t on the space of Y_r-measurable functions. Taking expectation w.r.t to Y_r on the last inequality gives

$$\mathbf{E}(Y_t - \lambda_r)^2 \ge \frac{1}{2}\mathbf{E}(Y_t - \tilde{Y}_t)^2 . \tag{7.1}$$

Now apply Lemma 7.8 with $y_i = Y_i$:

$$\mathcal{A}_{i=1}^{M}(Y_i - Y_{i-1})^2 \ge \sum_{k=1}^{m} \mathcal{A}_{j=1}^{2^{m-k}} \frac{(Y_{j2^k} - 2Y_{(2j-1)2^{k-1}} + Y_{(j-1)2^k})^2}{2^{2k}} .$$

Take expectations and apply (7.1) with $\lambda_r = -2Y_{(2j-1)2^{k-1}} + Y_{(j-1)2^k}$ to get

$$\mathbf{E}\mathcal{A}_{i=1}^{M}(Y_i - Y_{i-1})^2 \ge \frac{1}{2}\sum_{k=1}^{m} \mathbf{E}\mathcal{A}_{j=1}^{2^{m-k}} \frac{(Y_{j2^k} - \tilde{Y}_{j2^k}(k;j))^2}{2^{2k}} ,$$

as required. ∎

Proof of Theorem 7.6. Let T denote the full binary tree with depth $M = 2^m$ (for general depths, consider the tree up to depth which a power of 2) . Let $\{Z_i\}$ be the forward random walk on it starting from the root (i.e., at each vertex it goes right/left with probability 1/2). Clearly $d(Z_i, Z_{i+1})^2 = 1$ a.s., so $\mathbf{E}\mathcal{A}_{i=1}^{M} d(Z_i, Z_{i-1})^2 = 1$. Also, in the forward random walk, after the splitting at time r, with probability 1/2 the two independent walks will accumulate distance which is twice the number of steps. Thus, $\mathbf{E}d^2(Z_{j2^k}, \tilde{Z}_{j2^k}(k;j)) \ge 2^{2k-1}$, and we get that

$$\mathbf{E}\mathcal{A}_{i=1}^{M} d^2(Z_i, Z_{i-1}) = 1 \le \frac{2}{m}\sum_{k=1}^{m} \mathbf{E}\mathcal{A}_{j=1}^{2^{m-k}} \frac{d^2(Z_{j2^k}, \tilde{Z}_{j2^k}(k;j))}{2^{2k}} .$$

Now let $F : T \to H$ be an embedding with $\|F\|_{\text{Lip}} = 1$, then the previous inequality holds for $F(Z_i)$ up to a factor of $\|F^{-1}\|_{\text{Lip}}$, i.e.

$$\mathbf{E}\mathcal{A}_{i=1}^{M} d^2(F(Z_i), F(Z_{i-1})) \le \frac{2\|F^{-1}\|_{\text{Lip}}^2}{m}\sum_{k=1}^{m} \mathbf{E}\mathcal{A}_{j=1}^{2^{m-k}} \frac{d^2(F(Z_{j2^k}), F(\tilde{Z}_{j2^k}(k;j)))}{2^{2k}} .$$

By Lemma 7.9 we have

$$\mathbf{E}\mathcal{A}_{i=1}^{M} d^2(F(Z_i), F(Z_{i-1})) \ge \mathbf{E}\left[\frac{1}{2}\sum_{k=1}^{m} \mathcal{A}_{j=1}^{2^{m-k}} \frac{d^2(F(Z_{j2^k}) - F(\tilde{Y}_{j2^k}(k;j)))}{2^{2k}}\right] ,$$

which, combined with previous inequality, yields $\|F^{-1}\|_{\text{Lip}}^2 \ge \frac{m}{4}$, as required. ∎