Chapter IV Low distortion Euclidean embeddings

Central to the "Ribe program" in functional analysis: Relate Banach space (linear, normed, complete) concepts to metric space concepts

(fascinating surveys: Ball, Naor)

→ Application to data science & algorithm design:

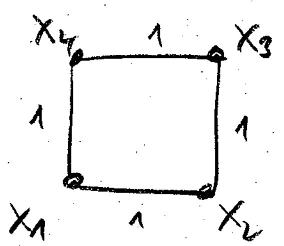
Data sets come equipped with a natural metric (for example by a similarity measure) but not with a linear one. Now embed the data (with as little distortion as possible) into a Banach space and use efficient algorithm which exploit the linear structure to work with the data.

§1 Embeddings of finite metric spaces

Def. A *finite metric space* is a pair (X,d) where X is a finite set and where $d : X \times X \to \mathbb{R}$ defines a metric on X: For all $x, y, z \in X$ we have

- (*non-negativity*) $d(x,x) \geq 0$ and $d(x,y) = 0 \iff x = y$
- (*symmetry*) $d(x,y) = d(y,x)$
- (*triangle inequality*) $d(x,z) \leq d(x,y) + d(y,z)$

Example *Shortest path metric* of a connected weighted graph $G = (V,E)$ with nonnegative weights on edges $w \in \mathbb{R}_+^E$.

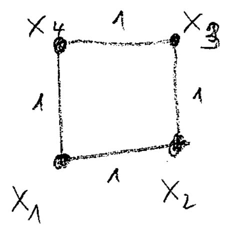


$$d(x_1, x_3) = 2$$

<u>Def.</u> Let (X, d) be a finite metric space. A <u>Euclidean embedding</u> of X is an injective map $f : X \to \mathbb{R}^n$. We say X embeds <u>isometrically</u> into $\mathbb{R}^n$ if there is a Euclidean embedding $f : X \to \mathbb{R}^n$ satisfying

$$d(x,y) = \|f(x) - f(y)\| \quad \forall x, y \in X.$$

<u>Example</u>



does not embed isometrically into $\mathbb{R}^n$.

<u>Goal</u> Approximate (X, d) by Euclidean metric. Measure quality of approximation by <u>distortion</u>.

Def. Let (X, d) be a finite metric space and let $f : X \to \mathbb{R}^n$ be a Euclidean embedding.

The distortion of f is defined by

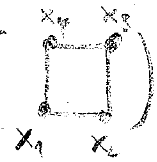
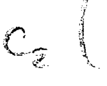
$$\text{distortion}(f) = \underbrace{\left(\max_{\substack{x,y \in X \\ x \neq y}} \frac{\|f(x) - f(y)\|}{d(x,y)} \right)}_{\text{expansion}(f)} \underbrace{\left(\max_{\substack{x,y \in X \\ x \neq y}} \frac{d(x,y)}{\|f(x) - f(y)\|} \right)}_{\text{contraction}(f)}.$$

Def. The minimal (Euclidean) distortion of (X, d) is defined as

$$c_2(X, d) = \min_{f : X \to \mathbb{R}^{|X|}} \text{distortion}(f)$$

Remarks

- To find the minimal distortion it suffice to consider embedding $f : X \to \mathbb{R}^n$ with $n = |X|$.
- The notation c_2 indicates that similar question make sense for other normed spaces.

Example $c_2($  $) \le \sqrt{2}$, but $c_2($  $) > 1$

Consider standard embedding f into $\mathbb{R}^2$.

Then expansion$(f) = 1$; contraction$(f) = \frac{2}{\sqrt{2}}$

Can compute $c_2(X,d)$ exactly and efficiently by solving a semidefinite program (SDP)

$$c_2(X,d)^2 = \min \gamma$$

(P) subject to $\gamma \in \mathbb{R}$, $f : X \to \mathbb{R}^{|X|}$

$$d(x,y)^2 \le \|f(x) - f(y)\|^2 \le \gamma\, d(x,y)^2$$

scale f s.t. contraction$(f) = 1$. for all $x, y \in X$, $x \neq y$.

Now express $\|f(x) - f(y)\|^2$ by the entries of the Gram matrix $Z \in S_+^X$ given by

$$Z_{xy} = f(x)^T f(y).$$

Then $\|f(x) - f(y)\|^2 = Z_{xx} - 2Z_{xy} + Z_{yy}$.

(P) is an instance of semidefinite programming: ($\leadsto$ Lecture "Convex Optimisation") Optimise a semidefinite matrix subject to linear constraints.

In general, one can solve SDPs efficiently, using the ellipsoid method or interior point method. Many practical solvers are available: MOSEK, cvx, cvxopt, csdp, ...

(Weak) SDP duality yields:

<u>Lemma</u>

$$c_2(X,d)^2 \geq \max \sum_{x \neq y} \alpha_{x,y}\, d(x,y)^2$$

$$\text{subject to } \alpha_{x,y}, \beta_{x,y} \geq 0$$

(D)

$$\sum \alpha_{xy}(-F_{xy}) + \sum \beta_{xy} F_{xy} \in S_+^X$$

$$\sum \beta_{xy}\, d(x,y)^2 \leq 1,$$

where $F_{xy} = e_x e_x^T - (e_x e_y^T + e_y e_x^T) + e_y e_y^T$.

<u>Proof</u> Rewrite (P) as

$$\min \gamma$$
$$\text{s.t. } \gamma \in \mathbb{R}_{+}, Z \in S_+^X$$

$$(\alpha_{xy} \geq 0) \quad \langle F_{xy}, Z \rangle \geq d(x,y)^2$$
$$(\beta_{xy} \geq 0) \quad \langle -F_{xy}, Z \rangle + \gamma d(x,y)^2 \geq 0$$

where $\langle A, B \rangle = \mathrm{Tr}(AB)$

Then we have $(P) \geq (D)$ because for feasible solutions of (P) and (D) we have

$$\gamma - \sum \alpha_{xy} d(x,y)^2 \geq \sum \beta_{xy} d(x,y)^2 \gamma - \sum \alpha_{xy} \langle F_{xy}, Z \rangle$$

$$\geq \sum \beta_{xy} \langle F_{xy}, Z \rangle - \sum \alpha_{xy} \langle F_{xy}, Z \rangle$$

$$= \langle \underbrace{\sum(\alpha_{xy}(-F_{xy}) + \beta_{xy} F_{xy})}_{\in S_+^X}, \underbrace{Z}_{\in S_+^X} \rangle \geq 0.$$

Here we need that $\langle A, B \rangle \geq 0$ if $A, B \in S_+^X$. This holds because for $A = \sum \lambda_i u_i u_i^T$, $\lambda_i \geq 0$, spectral decomposition, we have

$$\langle A, B \rangle = \langle \sum \lambda_i u_i u_i^T, B \rangle = \sum \lambda_i u_i^T B u_i \geq 0.$$

$\boxtimes$

Corollary

$$c(X,d)^2 \geq \max_{\substack{Y \in \mathcal{S}_+^X \\ Ye = 0}} \frac{\sum\limits_{\substack{x \neq y \\ Y_{xy} > 0}} Y_{xy}\, d(x,y)^2}{\sum\limits_{\substack{x \neq y \\ Y_{xy} < 0}} -Y_{xy}\, d(x,y)^2}\,,$$

where $e = (1, \dots, 1)^T$.

Proof → Exercise. $\boxtimes$

Corollary

$$c(X,d)^2 \geq \max_{\substack{Y \in S_+^X \\ Ye = 0}} \frac{\sum_{\substack{x \neq y \\ Y_{xy} > 0}} Y_{xy}\, d(x,y)^2}{\sum_{\substack{x \neq y \\ Y_{xy} < 0}} (-Y_{xy})\, d(x,y)^2} \qquad (*)$$

where $e = (1, \dots, 1)^T$.

Remark. • One can show that in $(*)$ one has even equality (by strong duality of SDPs).

• In case of equality we have

(a) $Y_{xy} > 0$ only for the most contracted pair (x,y), with $\frac{d(x,y)}{\|f(x) - f(y)\|}$ maximized.

(b) $Y_{xy} < 0$ only for the most expanded pair (x,y), with $\frac{\|f(x) - f(y)\|}{d(x,y)}$ maximized.

• If (X, d) is coming from the shortest path metric of a graph and if $f : X \to \mathbb{R}^n$ is a Euclidean embedding of X, then the most expanded pair (x,y) is attained at adjacent vertices x and y. (⇝ Exercise; follows from triangle inequality).

Theorem (Bourgain, 1985)

$$c_2(X,d) = \mathcal{O}(\log |X|).$$

Proof ⇝ see literature;
proof gives efficient randomised approximation algorithm.

Examples

(a) $c_2(\text{regular } r\text{-cube } C_r) = \sqrt{r} = \sqrt{\log |X|}$

(b) $c_2(\text{complete binary tree of depth } n) = \Theta(\sqrt{\log n}) = \Theta(\sqrt{\log \log |X|})$

(c) $c_2(\text{planar graph}) = \Theta(\sqrt{\log n})$

(d) $c_2(\text{strongly regular graphs})$ = explicit formula (→ later)

(e) $c_2(\text{graph with girth } g) = \Omega(\sqrt{g})$, where the girth g of a graph G is the length of a shortest cycle in G.

<u>Conjecture</u> (Linial, London, Rabinovich, 1995)

c_2(graph with girth g) $= \Omega(g)$.

(f) c_2(expander) $= \Theta(\log |X|)$ (LLR, 1995).

showing that Bourgain's theorem is tight.

One can prove (a), (d), (e), (f) by using the Corollary and exhibiting an appropriate matrix Y.

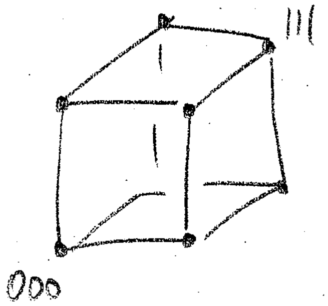
<u>Question</u> One should also be able to prove (b) using the Corollary. (no thesis)

Possible extension to q-ary tree.

(a) Least distortion embeddings of regular cubes

$$C_r = \left(\{0,1\}^r,\ d(i,j) = \left|\{k \in \{1,\dots,r\} : i_k \neq j_k\}\right| \right)$$

Hamming distance.

C_3  $d(000, 111) = 3$

<u>Clear</u>: $c_2(C_r) \le \sqrt{r}$ (standard embedding)

<u>Now</u>: $c_2(C_r) \ge \sqrt{r}$ by defining an appropriate Y.

Define $Y_{ij} = \begin{cases} -1, & \text{if } d(i,j) = 1 \\ r-1, & \text{if } i = j \\ 1, & \text{if } d(i,j) = r \\ 0, & \text{otherwise} \end{cases}$

Then

- $Ye = r(-1) + (r-1) + 1 = 0.$

- Y is positive semidefinite:

Of course, compute eigenvalues of Y using Fourier analysis.

Eigenvectors of Y are characters of $\left(\mathbb{Z}/2\mathbb{Z}\right)^r$:

$$\chi_u(v) = \prod_{k=1}^{r} e^{\frac{2\pi i u_k v_k}{2}} = \prod_{k=1}^{r} (-1)^{u_k v_k}$$

$$= (-1)^{u^T v}$$

So

$$(Y\chi_u)_j = \sum_i Y_{ji}\chi_u(i) = \sum_i Y_{ji}(-1)^{u^T i}$$

$$= \underbrace{(-1)^{u^T j}}_{\chi_u(j)} \underbrace{\left(\sum_i Y_{ji}(-1)^{u^T(i-j)}\right)}$$

$$= \underbrace{(-1)\sum_{l=1}^{r}(-1)^{u^T e_l} + (r-1)(-1)^{u^T 0} + 1\,(-1)^{u^T\left(\sum_{l=1}^{r} e_l\right)}}_{\text{does not depend on } j. \text{ (no eigenvalue)}}$$

Possible values of the first summand are

$-r, -r+2, \ldots$

If the first summand equals $-\tau$, then $u = 0$, but then the last summand $1 (-1)^{0^T (\sum_{\ell=1}^{\tau} e_\ell)} = 1$.

Thus, all eigenvalues are nonnegative.

Hence,

$$c_2(C_\tau)^2 \geq \frac{\sum_{\substack{i \neq j \\ Y_{ij} > 0}} Y_{ij}\, d(i,j)^2}{\sum_{\substack{i \neq j \\ Y_{ij} < 0}} (-Y_{ij})\, d(i,j)^2} = \frac{1 \cdot \tau^2}{-(-1) \cdot \tau \cdot 1^2} = \tau.$$

$\boxtimes$

(d) Least distortion embeddings of strongly regular graphs

Def. A simple, loopless graph $G = (V, E)$ is called strongly regular with parameters (v, k, λ, μ) if

(a) G is not complete or edgeless

(b) $v = |V|$

(c) $\forall x \in V: \quad |N(x)| = k$

$$N(x) = \{y \in V : \{x,y\} \in E\} \quad \text{neighbors of } x$$

(d) $\forall x, y \in V,\ x \neq y$:

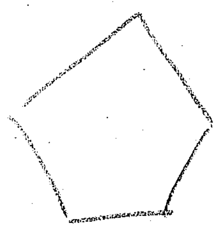
if $\{x,y\} \in E$: $\quad |N(x) \cap N(y)| = \lambda$

if $\{x,y\} \notin E$: $\quad |N(x) \cap N(y)| = \mu$

Examples

$(4,2,0,2)$

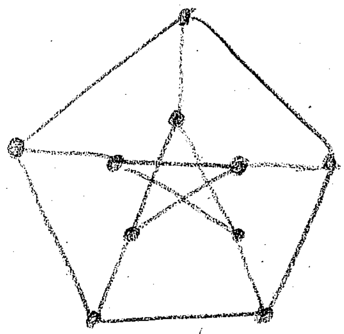
square



$(5,2,0,1)$

pentagon

Petersen graph



$(10, 3, 0, 1)$

Some strongly regular graphs are used to define sporadic simple groups, for example the Higman-Sims graph $(100, 22, 0, 6)$.

Consider adjacency matrix $A \in \{0,1\}^{V \times V}$ of a strongly regular graph, then

$$A^2 = (\lambda - \mu) A + (k - \mu) I + \mu J$$

since

$$A^2_{xy} = \# \text{ walks from } x \text{ to } y \text{ of length } 2,$$

and if $x = y$: $A^2_{xy} = k$

$\{x,y\} \in E$: $A^2_{xy} = \lambda$

$\{x,y\} \notin E$: $A^2_{xy} = \mu$

Lemma A has only 3 different eigenvalues:

$$k, \quad \frac{\lambda-\mu}{2} \pm \sqrt{\left(\frac{\lambda-\mu}{2}\right)^2 + k-\mu} = z_{1,2}$$

Proof We have $Ae = ke$. Let $u \perp e$ an eigenvector of A with eigenvalue z. Then

$$z^2 u = A^2 z = (\lambda-\mu) z u + (k-\mu) u$$

$$\Rightarrow \quad z^2 - (\lambda-\mu) z - (k-\mu) = 0. \qquad \boxtimes$$

Example Eigenvalues of Petersen graph: $3, \; -\frac{1}{2} \pm \sqrt{\frac{1}{4}+3-1} = -2, 1$

Ansatz Suppose G is connected, then setting

$$Y = -A + \alpha I + \beta J \quad \text{with } Ye = 0, \; Y \in S_+^V$$

gives optimization problem in one variable.

Then

$$0 = Ye = -ke + \alpha e + \beta v e \quad \Rightarrow \quad \alpha = k - \beta v$$

and

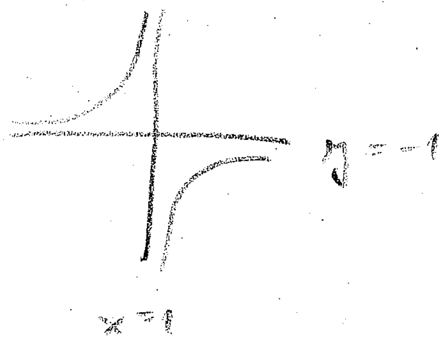
$$\sum_{\substack{x \neq y \\ Y_{xy} > 0}} Y_{xy} \overbrace{d(xy)^2}^{=4} = v(v-k-1)\beta \cdot 4,$$

$$\sum_{\substack{x\neq y\\ Y_{xy}<0}} (-Y_{xy}) \overbrace{d(x,y)^2}^{=1} = v\cdot k\cdot(1-\beta).$$

Together,

$$\frac{4(v-k-1)}{k}\,\frac{\beta}{1-\beta} \longrightarrow \max$$

Plot $\frac{\beta}{1-\beta} = -1 - \frac{1}{\beta-1}$



$$Y = -A + (k-\beta v)I + \beta J \in S_+^v$$

$$\iff -z_{1,2} + k - \beta v \geq 0$$

$$\iff \beta \leq \frac{-z_{1,2}+k}{v}$$

$$\iff \beta \leq \underbrace{\frac{-\left(\frac{\lambda-\mu}{2}+\sqrt{\left(\frac{\lambda-\mu}{2}\right)^2+k-\mu}\right)+k}{v}} \leq 1.$$

Choose $\beta =$.

This yields for the square $(4,2,0,2)$:

$$c_2(\square)^2 \geq \underbrace{\frac{4(4-2-1)}{2}}_{2} \frac{\beta}{1-\beta} = 2$$

since

$$\beta = \frac{1}{4}\left(-\left(\frac{-2}{2} + \sqrt{\left(\frac{-2}{2}\right)^2 + 2 - 2}\right) + 2\right) = \frac{1}{2},$$

for the pentagon $(5,2,0,1)$

$$c_2(\pentagon)^2 \geq \underbrace{\frac{4\ (5-2-1)}{2}}_{4} \frac{\beta}{1-\beta} \approx 1{,}527864$$

$$= \left(\frac{2}{\frac{1}{2}+\frac{1}{2}\sqrt{5}}\right)^2$$

since

$$\beta = \frac{1}{5}\left(-\left(\frac{-1}{2} + \sqrt{\frac{1}{4} + 2 - 1}\right) + 2\right)$$

$$= \frac{1}{5}\left(-\left(-\frac{1}{2} + \sqrt{\frac{5}{4}}\right) + 2\right)$$

$$= \frac{1}{2} - \frac{1}{2\sqrt{5}} \approx 0{,}27639\ldots$$

for the Petersen graph $(10,3,0,1)$

$$c_2(\text{Petersen})^2 \geq \underbrace{\frac{4\ (10-3-1)}{3}}_{8} \frac{\beta}{1-\beta} = \frac{16}{3}$$

Since

$$\beta = \frac{1}{10}\left(-\left(\frac{-1}{2} + \sqrt{\frac{1}{4} + 3 - 1}\right) + 3\right) = 2$$

(f) $c_2(\text{expander}) = \Omega(\log n)$

Theorem Let $G = (V, E)$ be a k-regular graph with $n = |V|$. Suppose $k - \lambda_2 \geq \varepsilon > 0$. Then,

$$c_2(G) \geq \sqrt{\frac{\varepsilon}{2k}} \left(\left\lfloor \log_k \frac{|V|}{2} \right\rfloor - 1 \right) = \Omega(\log n).$$

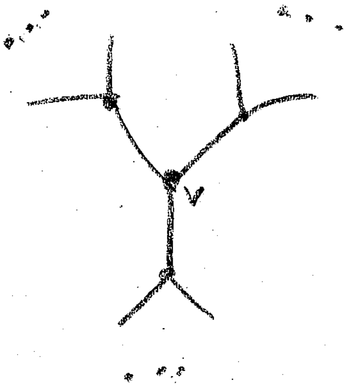
Proof Construct appropriate matrix Y.

Set $r = \left\lfloor \log_k \frac{|V|}{2} \right\rfloor - 1$. and define auxilliary graph $H = (V, E_H)$ where

$$\{u, v\} \in E_H \iff d_G(u, v) \geq r.$$

Claim Every vertex in H has degree $\geq \frac{|V|}{2}$.

Because Since G is k-regular every vertex $v \in V$ has $\leq k^{r+1}$ vertices at distance $\leq r$:



distance

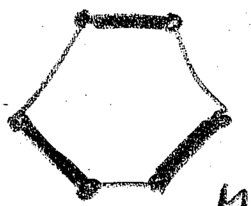
$$\underbrace{1}_{0} + \underbrace{k}_{1} + \underbrace{(k-1)\cdot k}_{2} + \dots + \underbrace{k(k-1)^{r-1}}_{r}$$

$$\leq \sum_{i=0}^{r} k^i \leq k^{r+1}$$

Now r is chosen so that $k^{r+1} \leq \frac{|V|}{2}$. $\boxtimes$

By a classical theorem of Dirac: H has an Hamiltonian cycle (a cycle which contains all vertices).

For simplicity we assume that n is even. Then H contains a perfect matching $M \subseteq E_H$:



M

Let $B \in S^V$ be the adjacency matrix of the perfect matching M, let $A \in S^V$ by the adjacency matrix of G.

Define $\boxed{Y = kI - A + \frac{\varepsilon}{2}(B - I)}$

Then

- $Ye = ke - ke + \frac{\varepsilon}{2}(\underbrace{Be}_{=e} - e)$ because B is a permutation matrix (M perfect matching).

$= 0$

- $Y \in S_+^V$:

For $x \in \mathbb{R}^V$, $x \perp e$, we have

$$x^T Y x = x^T(kI - A)x + \frac{\varepsilon}{2} x^T(B - I)x$$

$$\underbrace{\geq \underbrace{(k - \lambda_2)}_{\geq \varepsilon} \underbrace{x^T x}_{=\|x\|^2} + \underbrace{\frac{\varepsilon}{2} \sum_{\{u,v\} \in M} \underbrace{(2x_u x_v - x_u^2 - x_v^2)}_{\geq -2(x_u^2 + x_v^2)}}_{\geq \frac{\varepsilon}{2}(-2 \underbrace{\sum_{\{u,v\} \in M} x_u^2 + x_v^2}_{= -2\|x\|^2})}}_{\geq 0.}$$

• Objective value:

$$\sum_{\substack{u \neq v \\ Y_{uv} > 0}} Y_{uv}\, d(u,v)^2 \geq \frac{\varepsilon}{2} |V| \, r^2$$

$$\sum_{\substack{u \neq v \\ Y_{uv} < 0}} (-Y_{uv})\, d(u,v)^2 = 1 \cdot k \cdot |V|$$

$$\Rightarrow \quad c_2(G)^2 \geq \frac{\varepsilon}{2k} r^2 .$$ ☒

* * *