Robust Euclidean Embedding

Lawrence Cayton, Sanjoy Dasgupta

University of California, San Diego

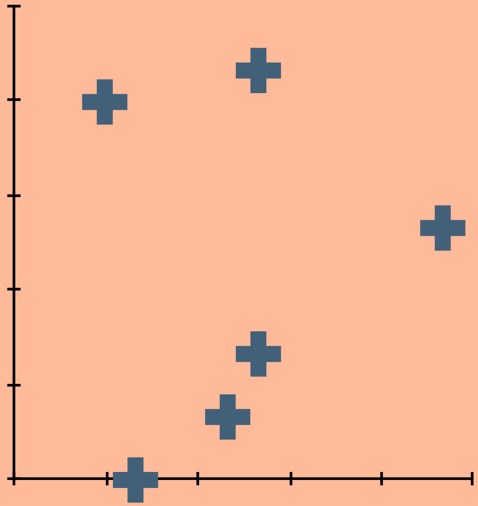
Embedding / multidimensional scaling

$D =$

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	0	3	2	9	10	7
x_2	3	0	14	1	2	1
x_3	2	14	0	2	7	9
x_4	9	1	2	0	2	2
x_5	10	2	7	2	0	13
x_6	7	1	9	2	13	0

$\longrightarrow$

$X =$



dissimilarity matrix

$D \in \mathbf{R}^{n \times n}$

vectors

$X \in \mathbf{R}^{n \times d} \quad : \quad \|X_i - X_j\|_2^2 \approx D_{ij}$

Another view:

Project D onto the Euclidean distance matrix $(\mathbb{EDM})$ cone.

Machine learning uses

- Visualization

- Dimensionality reduction:

$$\text{given } y_1, \ldots, y_n \in \mathbf{R}^D, \text{ find } x_1, \ldots, x_n \in R^d \text{ s.t.}$$

$$\|x_i - x_j\| \approx \|y_i - y_j\|$$

- Adapting non-Euclidean dissimilarity measures to Euclidean algorithms

To be discussed

I. Classical MDS and some problems with it.

II. An alternative: Robust Euclidean Embedding.

III. Hardness of embedding for dimensionality reduction.

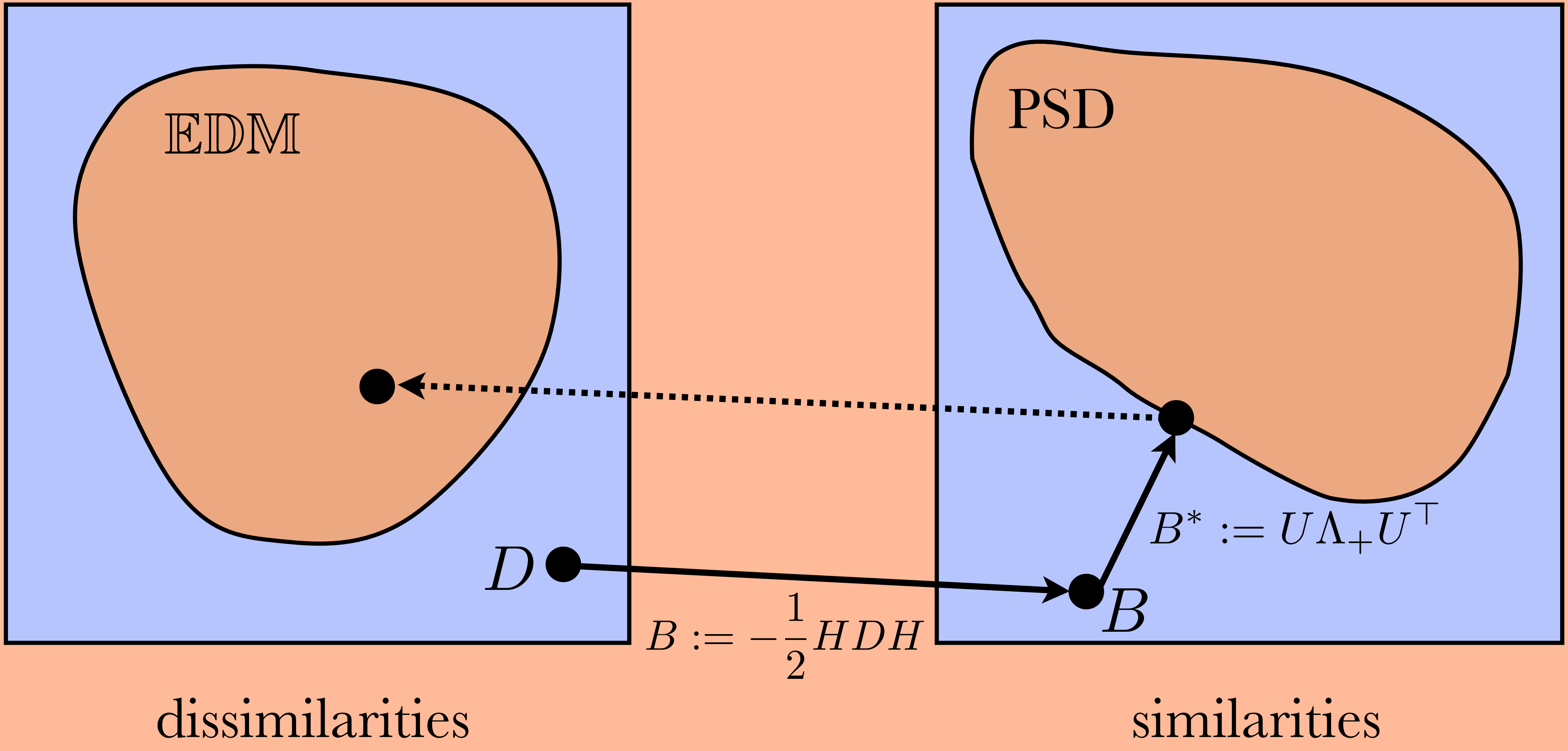
Schoenberg's EDM criterion

$$D \in \mathbb{EDM} \iff \begin{cases} -\frac{1}{2}HDH \succeq 0 & \\ D_{ij} \geq 0 & (i \neq j) \\ D_{ii} = 0 & \end{cases}$$

$$\text{where } H = I - \tfrac{1}{n}\mathbf{1}\mathbf{1}^\top$$

- When $D \in \mathbb{EDM}$, $B := -\dfrac{1}{2}HDH$ will be a Gram matrix for the underlying configuration.
- Otherwise?

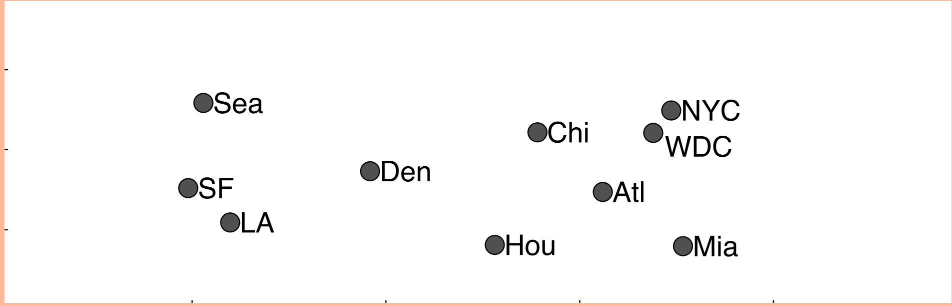
Classical multidimensional scaling



Map example

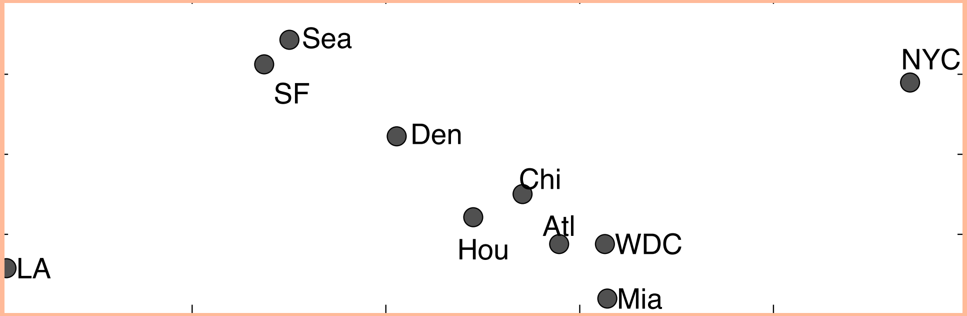
D = distances between 10 US cities

cMDS embedding:



Corrupt the distance between NYC and LA (double it).

cMDS embedding:



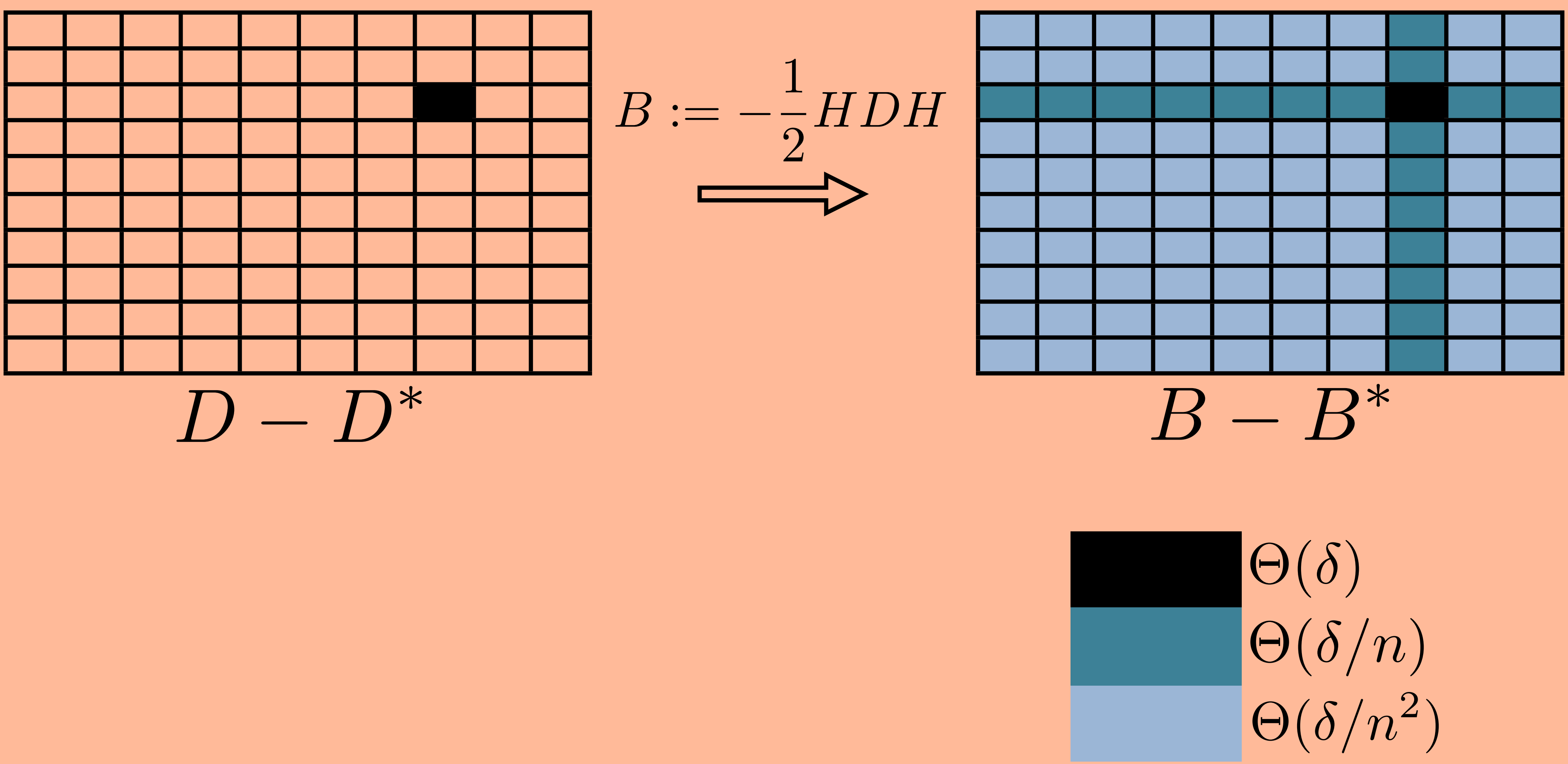
One corrupted distance (out of 45) ruined the embedding.

Kruskal & Wish 1978

Error dispersion

$D^* \in \mathbb{EDM}$

Form D by corrupting one entry of D^*



When is this effect felt?

- Dissimilarities where the noise scales with the magnitude

  e.g.: Isomap. Small local distances are accurate, large distances are rough estimates of geodesics.

- Fundamentally non-Euclidean dissimilarities at multiple scales

  e.g.: KL-divergence.

Classical MDS issues

cost function: $f(D^*) = \|HDH - HD^*H\|_2$

I. Error dispersion

II. Frobenius norm

III. Can't handle missing entries

IV. Can't adjust weighting

Robust Euclidean Embedding

Classical MDS program:

$$\begin{aligned} \min \quad & \|HDH - HD^*H\|_2 \\ \text{subject to} \quad & D^* \in \mathbb{EDM} \end{aligned}$$

REE program:

$$\begin{aligned} \min \quad & \|D - D^*\|_1 \\ \text{subject to} \quad & D^* \in \mathbb{EDM} \end{aligned}$$

... as a SDP:

$$\begin{aligned} \min \quad & \sum_{i,j} \xi_{ij} \\ \text{subject to} \quad & -\xi_{ij} \le D_{ij} - B_{ii} - B_{jj} + 2B_{ij} \le \xi_{ij} \\ & \sum_{ij} B_{ij} = 0 \\ & B \succeq 0; \quad \xi_{ij} \ge 0 \end{aligned}$$

Linial, London, Rabinovich 1995; Dattorro 2005

Solving the REE program

For $n \approx 100$: general purpose SDP solver works [e.g. SDPT3].

For larger n: first-order descent method.

cost $$f(B) = \sum_{ij} W_{ij} |D_{ij} - [\mathrm{dist}(B)]_{ij}|$$

subgradient $$[G(B)]_{ij} = \begin{cases} W_{ij}\mathbf{I}\left([\mathrm{dist}(B)]_{ij} < D_{ij}\right) & \text{if } i \neq j; \\ \sum_k W_{ik}\mathbf{I}([\mathrm{dist}(B)]_{ik} > D_{ik}) & \text{if } i = j. \end{cases}$$

($\mathbf{I}$ denotes the indicator function returning +/- 1)

procedure

```
loop
```

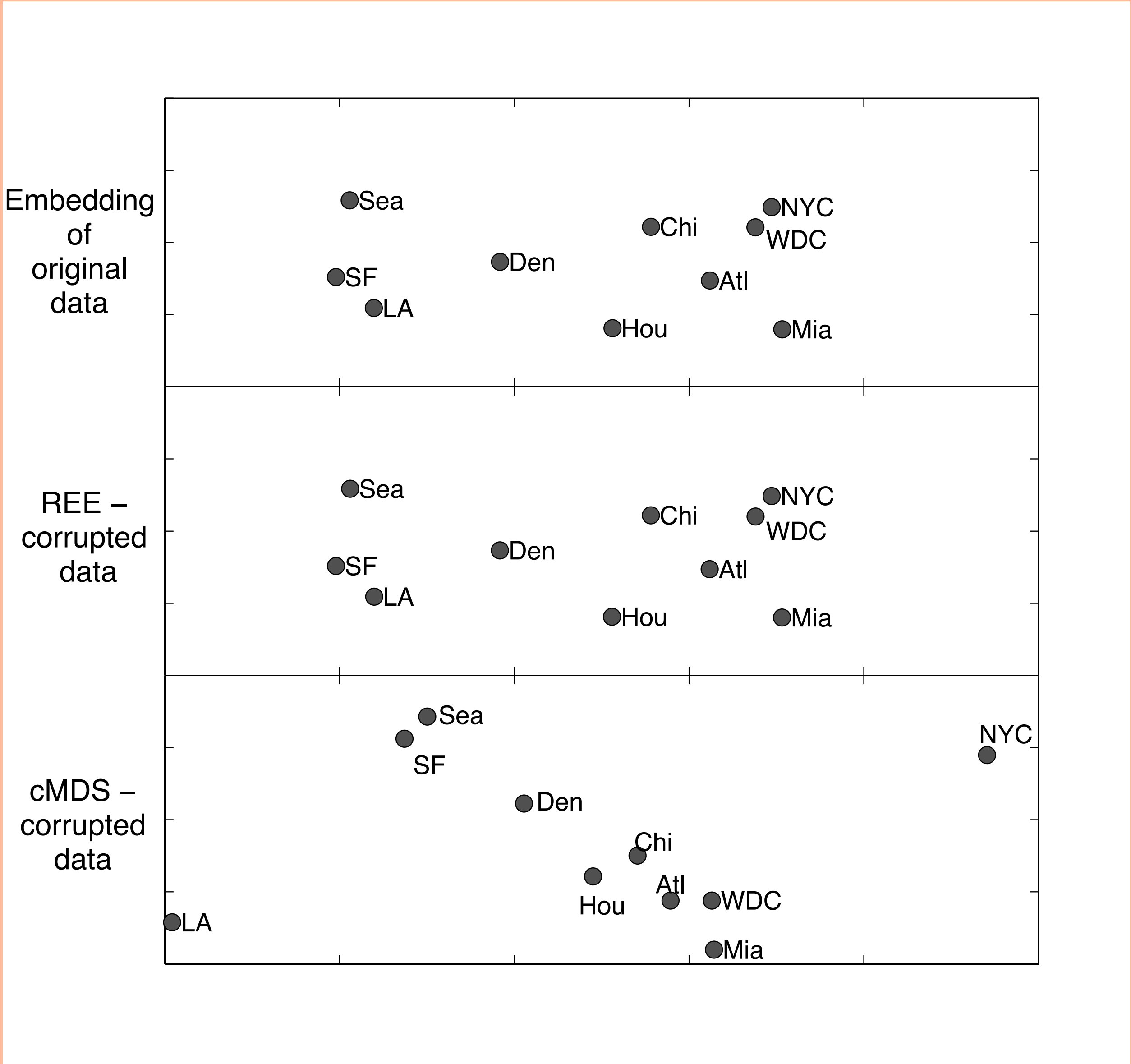
I. Move along subgradient

II. Project back onto PSD cone

Robust Euclidean Embedding
(subgradient implementation)
input: D, $W \in \mathbf{R}^{n\times n}$

1. Set $B^0 \in \mathbf{R}^{n\times n}$ randomly.
2. for $k = 1, 2, \ldots$
   - Set $B := B^{k-1} - \alpha_k G(B^{k-1})$.
   - Spectrally decompose B: $B = U\Lambda U^\top$.
   - Set $[\Lambda_+]_{ij} := \max\{\Lambda_{ij}, 0\}$.
   - $B^k := U\Lambda_+ U^\top$.
3. Pick k minimizing $\left(\sum_{ij} W_{ij}|D_{ij} - \mathrm{dist}(B^k)|\right)$.
4. Return $X := U\Lambda^{1/2}$, where $U\Lambda U^\top$ is the spectral decomposition of B^k.

Map example


Embedding of original data
Sea
Chi
NYC
WDC
Den
SF
Atl
LA
Hou
Mia
REE – corrupted data
Sea
Chi
NYC
WDC
Den
SF
Atl
LA
Hou
Mia
cMDS – corrupted data
Sea
SF
NYC
Den
Chi
Hou
Atl
WDC
LA
Mia

Shape distance

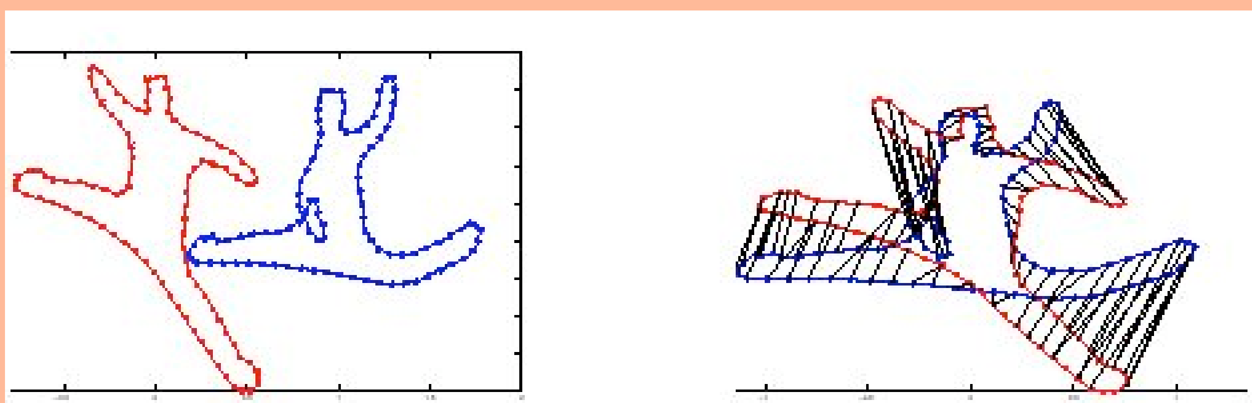
Measures the similarity of two images *as shapes.*

Procedure:

I. Sample points from each image

II. Match up the points (bi-partite matching)

III. Compute the energy necessary to morph one image into the other.



http://www.seas.upenn.edu/~cse399b/

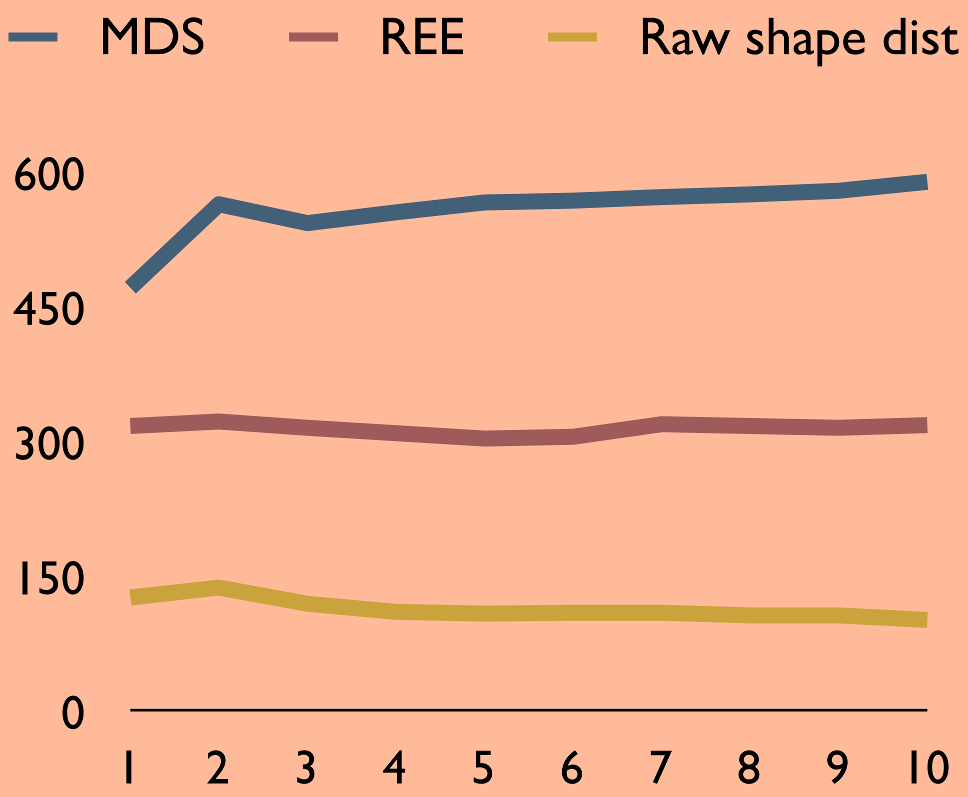
Belongie, Malik, Puzicha 2002

Shape distance* experiment

1.Computed the shape distance for 1000 MNIST digit images.

2. Embedded the distances using both cMDS and REE.

3. Classified each image using its nearest Euclidean neighbor among the remaining 999.



Low dimensional embedding

Input: D, k

Find $x_1, \ldots x_n \in \mathbf{R}^k$ such that

$$\|x_i - x_j\|_2^2 \approx D_{ij}$$

Can find an optimal solution under the cMDS cost function

$$f(D^*) = \|HDH - HD^*H\|_2$$

by embedding and then running principal components analysis

Dhamdhere, Gupta, Ravi 2004; Hastad, Ivansson, Lagergren 1998

Low dimensional embedding

What about under the REE cost function?

$$f(D^*) = \|D - D^*\|_1$$

Running PCA afterwards is no longer optimal.

Hardness result

Input: D

Problem: find an embedding minimizing the average distortion:

$$f(D^*) = \sum_{i,j} |D_{ij} - |x_i - x_j||$$

NP-hard.

More generally, for

$$f(D^*) = \sum_{i,j} h\Big(g(D_{ij}) - g(|x_i - x_j|)\Big)$$

the embedding problem is NP-hard.

[h, g are symmetric, bi-lipshitz, & monotonic.]

Trace heuristic

Common rank-reduction heuristic

$$\begin{aligned} \min \quad & \sum_{i,j} \xi_{ij} + \boxed{\gamma \cdot \mathrm{trace}(B)} \\ \text{subject to} \quad & -\xi_{ij} \leq D_{ij} - B_{ii} - B_{jj} + 2B_{ij} \leq \xi_{ij} \\ & \sum_{ij} B_{ij} = 0 \\ & B \succeq 0; \quad \xi_{ij} \geq 0 \end{aligned}$$

conflicts with the dual

$$\begin{aligned} \max \quad & \sum_{ij} D_{ij} S_{ij} \\ \text{subject to} \quad & S_{ij} \in [-1, +1] \text{ for } i \neq j \\ & S\mathbf{1} = \gamma \mathbf{1} \\ & S \succeq 0 \end{aligned}$$

Summary

I. Robustness of Classical MDS

II. Robust Euclidean Embedding

III. Hardness of low-dimensional embedding