# Distance Shrinkage and Euclidean Embedding

Ming Yuan

Morgridge Institute for Research

and

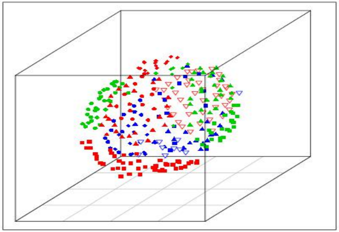
Department of Statistics
University of Wisconsin-Madison

myuan@stat.wisc.edu
http://www.stat.wisc.edu/~myuan

(Joint work with Luwan Zhang and Grace Wahba)

# VPU Sequence Variation

[Pickering et al., 2014]



Multidimensional Scaling

To what extent, does multidimensional scaling (MDS) faithfully reflect features in the original data?

# From Dissimilarity to Distance

- A set of objects $\{O_1, \ldots, O_n\}$ from an arbitrary domain $\mathcal{O}$
- Observe pairwise dissimilarity scores $x_{ij}$s

$$x_{ij} \approx \mathrm{dist}(O_i, O_j), \qquad (i,j) \in \Omega$$

- "Closest" Euclidean embedding – $p_1, \ldots, p_n \in \mathbb{R}^{n-1}$

$$\mathrm{dist}(O_i, O_j) = \|p_i - p_j\|^2, \qquad 1 \le i < j \le n$$

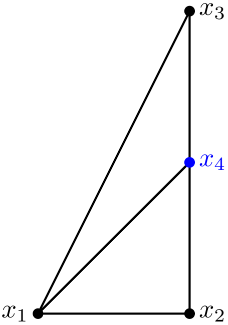
- Other applications – protein folding, chromosome conformation capture, graph drawing, ...

# What is an EDM

$$X = (x_{ij})_{1\le i,j\le n} \Longrightarrow D = (d_{ij} = \|p_i - p_j\|^2)_{1\le i,j\le n} \in \mathcal{D}_n$$

- Nonnegativity – $d_{ij} \ge 0$
- Identity – $d_{ii} = 0$
- Symmetry – $d_{ij} = d_{ji}$
- Triangle inequality – $\sqrt{d_{ij}} + \sqrt{d_{jk}} \ge \sqrt{d_{ik}}$
- More than a metric

$$\begin{pmatrix} 0 & 1 & 5 & d_{14} \\ 1 & 0 & 4 & 1 \\ 5 & 4 & 0 & 1 \\ d_{14} & 1 & 1 & 0 \end{pmatrix} \Longrightarrow (\sqrt{5} - 1)^2 \le d_{14} \le 4$$



# Geometry of EDM

A symmetric matrix $D \in \mathbb{R}^{n\times n}$ is an Euclidean distance matrix iff

- It is hollow – $d_{ii} = 0$
- It is conditionally negative semi-definite on

$$\mathcal{X}_n = \{x \in \mathbb{R}^n : x^\top \mathbb{1} = 0\}$$

Embedding can be identified with the eigenstructure of Schönberg transform of $D$

$$\mathscr{R}(D) = -\frac{1}{2} JDJ \qquad \text{where } J = I - \mathbb{1}\mathbb{1}^\top/n$$

Important consequence:

- The set of $n \times n$ EDMs is a convex cone without interior

[Schönberg (1935)]

# Estimating an EDM

$$x_{ij} = d_{ij} + \varepsilon_{ij} \Longrightarrow D$$

- Projection type of estimate – $\mathcal{D}_n$ is a closed convex hull

$$\mathscr{P}_{\mathcal{D}_n}(X) = \operatorname*{arg\,min}_{M \in \mathcal{D}_n} \|X - M\|_{\mathrm{F}}^2.$$

- Not working – at most $n(n-1)/2$ observations with $n(n-1)/2$ parameters
- Accounting for low embedding dimension – MDS

$$\mathscr{P}_{\mathcal{D}_n(r)}(X) = \operatorname*{arg\,min}_{M \in \mathcal{D}_n(r)} \|X - M\|_{\mathrm{F}}^2.$$

  - Computationally challenging
  - Statistically unstable

# Regularized Kernel Estimation

- From EDM to kernel $K = (k_{ij})$:

$$d_{ij} = \|p_i - p_j\|^2 = p_i^\top p_i + p_j^\top p_j - 2p_i^\top p_j =: k_{ii} + k_{jj} - 2k_{ij}$$

  - $D$ is an EDM iff $K \succeq 0$
- Regularized kernel estimate

$$\widehat{K} = \arg\min_{M \succeq 0} \left\{ \sum_{(i,j)\in\Omega} \left( x_{ij} - \underbrace{\left\langle M, (e_i - e_j)(e_i - e_j)^\top \right\rangle}_{m_{ii}+m_{jj}-2m_{ij}} \right)^2 + \lambda_n \operatorname{trace}(M) \right\}$$

- Back to distance matrix

$$\widehat{d}_{ij} = \widehat{k}_{ii} + \widehat{k}_{jj} - 2\widehat{k}_{ij}$$

[Lu et al. (2005) and Weinberger et al. (2007)]

# Why Does it Work?

- Kernel is not estimable from distance data

$$\mathscr{T}:\mathcal{S}_n \to \mathcal{D}_n \quad M \mapsto (m_{ii}+m_{jj}-2m_{ij})_{1\le i,j\le n} \qquad \text{is not injective}$$

- What are we estimating – Minimum trace kernel

  - the preimage $\mathcal{M}(D)$ of any $D \in \mathcal{D}_n$ under $\mathscr{T}$ is convex
  - there is a unique minimum trace kernel in $\mathcal{M}(D)$

$$\mathscr{R}(D) = \underset{M\in\mathcal{M}(D)}{\arg\min}\ \mathrm{trace}(M)$$

  - $\mathscr{R}(\cdot)$ is the Schönberg transform

$$\mathscr{R}(D) = -\frac{1}{2}JDJ$$

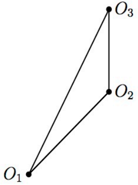
# Distance Shrinkage

$$\begin{tikzcd}
K \arrow[r, "\mathscr{T}"] & D \arrow[r, leftrightarrow] & \mathscr{R}(D) \\
& \widehat{D} \arrow[u, dashed] \arrow[r, leftrightarrow] & \widehat{K} \arrow[u, dashed]
\end{tikzcd}$$

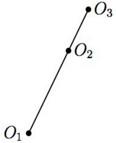
$\widehat{D} = \mathscr{T}(\widehat{K})$ and $\widehat{K} = \mathscr{R}(\widehat{D})$

$$\widehat{D} = \mathscr{T}(\widehat{K}) = \mathscr{P}_{\mathcal{D}_n}\left(X - \frac{\lambda_n}{2n}\mathbb{1}\mathbb{1}^\top\right)$$

# Effect of Distance Shrinkage




$$\mathscr{P}_{\mathcal{D}_3}(X) \Longrightarrow \mathscr{P}_{\mathcal{D}_3}(X - \eta D_0)$$




- Embedding dim is one if

$$\frac{1}{3}(x_{12} + x_{13} + x_{23}) - \frac{\Delta_x}{3} \leq \eta < \frac{1}{3}(x_{12} + x_{13} + x_{23}) + \frac{2\Delta_x}{3},$$

where

$$\Delta_x := \sqrt{2[(x_{12} - x_{13})^2 + (x_{12} - x_{23})^2 + (x_{13} - x_{23})^2]}$$

# How Well Does it work?

Let $\widehat{D} = \mathscr{P}_{\mathcal{D}_n}\left(X - \frac{\lambda}{2n}\mathbb{1}\mathbb{1}^\top\right)$. For any $\lambda \geq 2\|X - D\|$,

$$\|\widehat{D} - D\|_{\mathrm{F}}^2 \leq \inf_{M \in \mathcal{D}_n} \left\{ \|M - D\|_{\mathrm{F}}^2 + \frac{9}{4}\lambda^2(\dim(M) + 1) \right\}.$$

- If $\dim(D) = r$, then

$$\|\widehat{D} - D\|_{\mathrm{F}}^2 \lesssim r\|X - D\|^2$$

- For sub-Gaussian errors – $\lambda \sim \sqrt{n}$

$$\|\widehat{D} - D\|_{\mathrm{F}}^2 \lesssim_p rn \Longleftarrow \text{Minimax optimal}$$

# Fix Dimension Embedding

- Eigenvalue decomposition of $\mathscr{R}(\widehat{D})$
- Keep the leading $r$ eigenvalues – $W_r$
- Getting back to Euclidean distance matrix $\widehat{D}_r = \mathscr{T}(W_r)$

Let $D_r = \mathscr{P}_{\mathcal{D}_n(r)} D$. For any $\lambda \geq 2\|X - D\|$,

$$\|J(\widehat{D}_r - D_r)J\|_{\mathrm{F}}^2 \leq C\left(\min_{M\in\mathcal{D}_n(r)} \|J(D-M)J\|_{\mathrm{F}}^2 + \lambda^2 r\right)$$

# Projection to EDM

Recall from Schönberg (1935),

$$\mathcal{D}_n = S_1 \cap S_2$$

where

$$S_1 = \{M \in \mathbb{R}^{n\times n} : JMJ \preceq 0\},$$

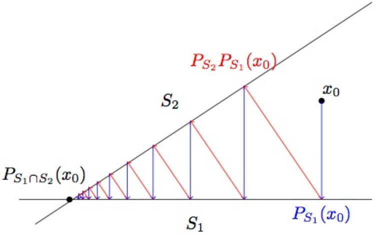
and

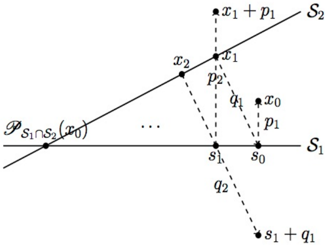
$$S_2 = \{M \in \mathbb{R}^{n\times n} : \mathrm{diag}(M) = \mathbf{0}\}.$$

Both $P_{S_1}$ and $P_{S_2}$ have analytic forms amenable to computation.

[Glunt et al. (1990)]

# Alternating Projection Methods




- von Neumann's method of alternating projections (von Neumann, 1933)

$$S_1, S_2 \text{ closed} \Longrightarrow \lim_{n\to\infty} (P_{S_1}P_{S_2})^n x_0 = P_{S_1 \cap S_2} x_0$$

- Dykstra's algorithm (1986)

$$x_n^0 := x_{n-1}^2, \quad x_n^i := P_{S_i}(x_n^{i-1} - y_{n-1}^i), \quad y_n^i = x_n^i - (x_n^{i-1} - y_{n-1}^i)$$

# Dealing with Missing Observations

- Observe

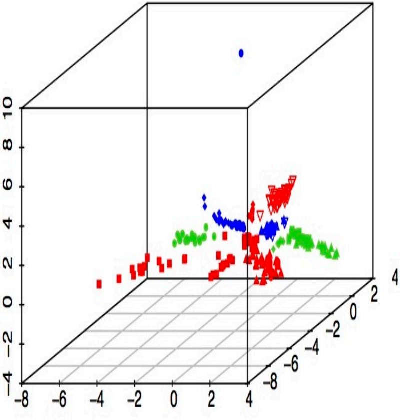
$$x_{ij} = d_{ij} + \varepsilon_{ij}, \qquad (i,j) \in \Omega \subset \{(i,j) : 1 \leq i < j \leq n\}$$

- Goal

$$\min_{M \in \mathcal{D}_n} \sum_{(i,j)\in\Omega} \left[(x_{ij} - \eta) - m_{ij}\right]^2$$

- EM Algorithm
  - Initialization $x_{ij}$ for $(i,j) \in \Omega^c$
  - M Step – $D^{(t+1)} \approx \mathscr{P}_{\mathcal{D}_n}(X^{(t)} - \eta D_0)$
  - E Step – $x_{ij}^{(t+1)} = d_{ij}^{(t+1)}$ for $(i,j) \in \Omega^c$
- Cross-validation to choose $\eta$

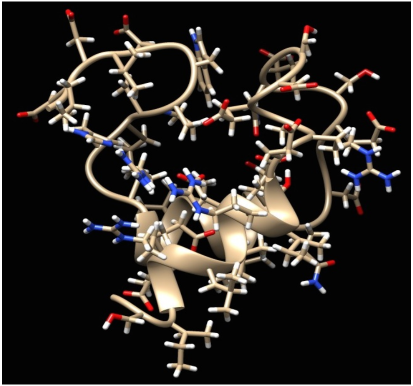
# VPU Sequence Variation

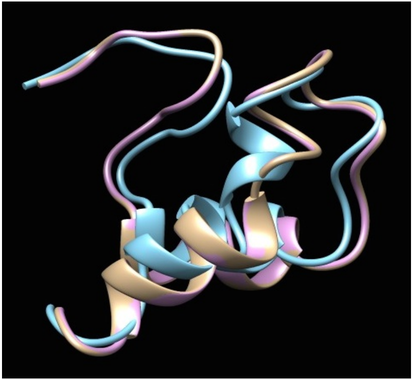



- 304 Vpu Sequences from 14 HIV-1 infected individuals
  - 5 long term non-progressors
  - 4 normal progressors
  - 5 Rapid progressors

# Protein Secondary Structure

- Coordinates of 671 atoms taken from PDB (ID: 2K7Y)
- Simulate pairwise distances with noise





# Summary

- The problem of reconstructing EDM from pairwise dissimilarity scores arises naturally in many applications
- Motivated in particular by biological problems:
  - Notion of distance between genomic sequences
  - Molecular structure determination
  - Chromosome conformation
- Distance shrinkage
  - Encourages low dimensional embedding
  - Leads to improved estimation risk
  - Efficient to compute
- Looking ahead – clustering, tree, . . .