Convex Optimization Learning of Faithful Euclidean Distance Representations in Nonlinear Dimensionality Reduction

Chao Ding

Academy of Mathematics and Systems Science

Chinese Academy of Sciences

a joint work with Hou-Duo Qi @ University of Southampton, UK

ICCOPT, Tokyo
August 10, 2016

中國科學院
CHINESE ACADEMY OF SCIENCES

Outline

Distance geometry problem

- Given an integer $r > 0$ and a simple undirected graph $G = (V, E)$ whose edges are weighted by a nonnegative function $d : E \to \mathbb{R}_+$, find a function $x : V \to \mathbb{R}^r$ such that

$$\forall \{u, v\} \in E \quad \|x(u) - x(v)\| = d(u, v).$$

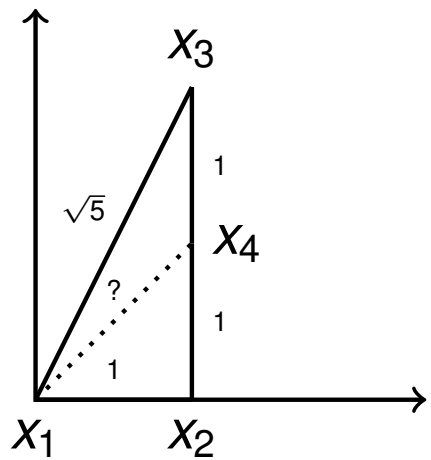
Distance geometry problem

- Given an integer $r > 0$ and a simple undirected graph $G = (V, E)$ whose edges are weighted by a nonnegative function $d : E \to \mathbb{R}_+$, find a function $x : V \to \mathbb{R}^r$ such that

$$\forall \{u, v\} \in E \quad \|x(u) - x(v)\| = d(u, v).$$

E.g., find $x_1, x_2, x_3, x_4 \in \mathbb{R}^2$ from

$$d = \begin{bmatrix} 0 & 1 & \sqrt{5} & ? \\ 1 & 0 & 2 & 1 \\ \sqrt{5} & 2 & 0 & 1 \\ ? & 1 & 1 & 0 \end{bmatrix} \quad \Longrightarrow$$



Solving DGPs

- **Euclidean Distance Matrix** (EDM): a $n \times n$ matrix D is a (squared) EDM if there exist points $x_1, \ldots, x_n$ in $\mathbb{R}^r$ such that

$$D_{ij} = \|x_i - x_j\|^2 \quad \forall\, i, j.$$

r is called the embedding dimension of D.

Solving DGPs

- **Euclidean Distance Matrix** (EDM): a $n \times n$ matrix D is a (squared) EDM if there exist points $x_1, \ldots, x_n$ in $\mathbb{R}^r$ such that

$$D_{ij} = \|x_i - x_j\|^2 \quad \forall\, i, j.$$

  r is called the embedding dimension of D.

- Given a true EDM D, the points $x_1, \ldots, x_n$ can be determined as follows (up to rotations, translations and reflections)

$$X = [x_1, \ldots, x_n] \in \mathbb{R}^{r \times n}, \quad K := X^T X = -\frac{1}{2} JDJ$$

  with $r = \operatorname{rank}(JDJ)$ and

$$J = I - \frac{1}{n} ee^T, \quad e = (1, \ldots, 1)^T$$

Solving DGPs

- **Euclidean Distance Matrix** (EDM): a $n \times n$ matrix D is a (squared) EDM if there exist points $x_1, \ldots, x_n$ in $\mathbb{R}^r$ such that

$$D_{ij} = \|x_i - x_j\|^2 \quad \forall\, i, j.$$

  r is called the embedding dimension of D.

- Given a true EDM D, the points $x_1, \ldots, x_n$ can be determined as follows (up to rotations, translations and reflections)

$$X = [x_1, \ldots, x_n] \in \mathbb{R}^{r \times n}, \quad K := X^T X = -\frac{1}{2} JDJ$$

  with $r = \operatorname{rank}(JDJ)$ and

$$J = I - \frac{1}{n} ee^T, \quad e = (1, \ldots, 1)^T$$

- **Challenge**: D is not a true EDM (missing data + noise)

An application: social network analysis

One important topic: Structural patters of social network from the relationships that connect social actors

An application: social network analysis

One important topic: Structural patters of social network from the relationships that connect social actors

- relationships: kinship, affective, social role, transfer of informations, etc

An application: social network analysis

One important topic: Structural patters of social network from the relationships that connect social actors

- relationships: kinship, affective, social role, transfer of informations, etc
- social distances: presence/absence, similarity/dissimilarity

An application: social network analysis

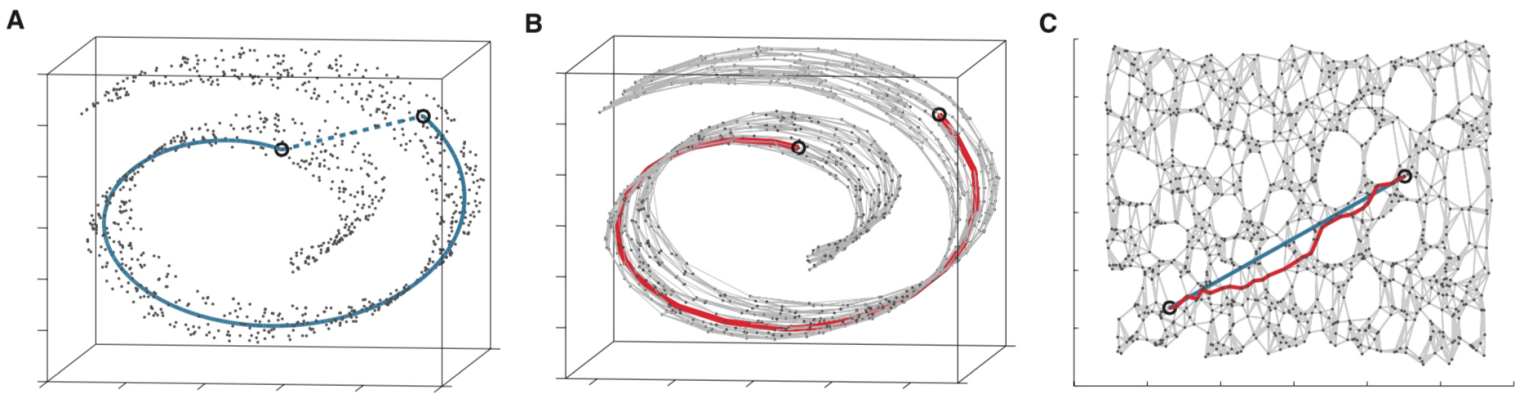
One important topic: Structural patters of social network from the relationships that connect social actors

- relationships: kinship, affective, social role, transfer of informations, etc
- social distances: presence/absence, similarity/dissimilarity

Objective: determine the faithful social positions of actors from the partial social distance measurements

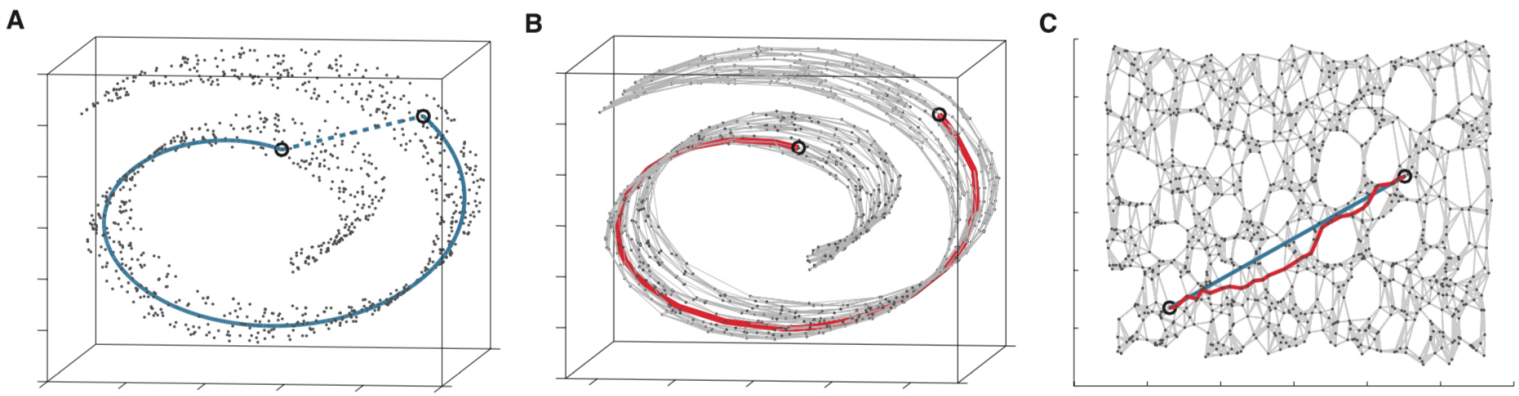
Existing methods

ISOMAP: *Tenenbaum et al. (2000)*



Existing methods

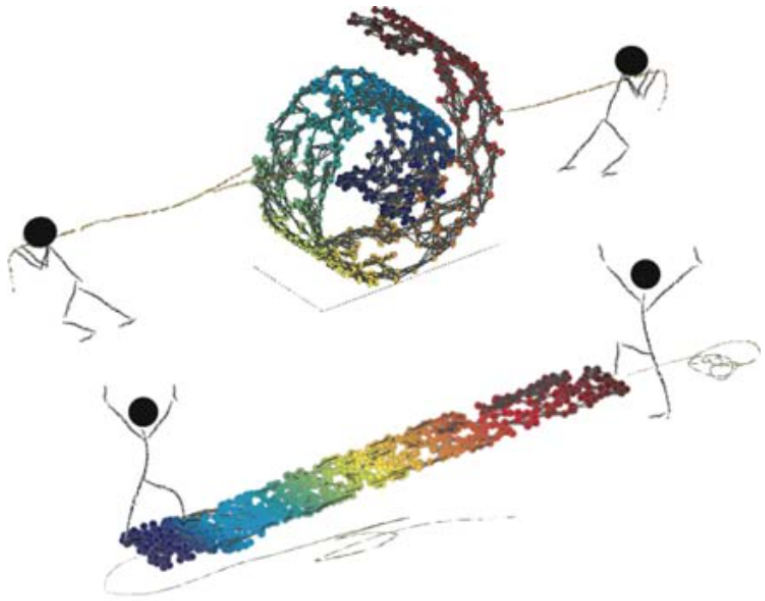
ISOMAP: *Tenenbaum et al. (2000)*



— the shortest path distances can accurately estimate the true geodesic distances with a high probability if the finite points are chosen randomly from a compact and convex submanifold

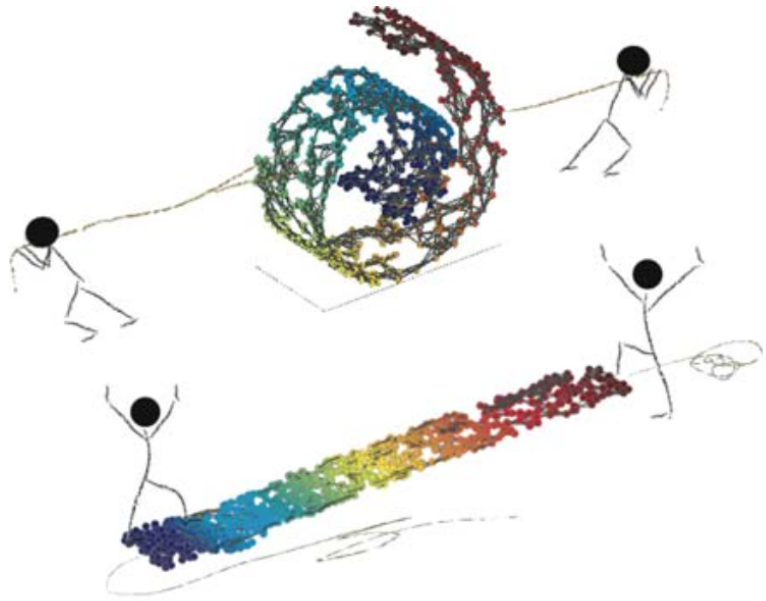
Existing methods (cont.)

Maximum variance unfolding (MVU): *Weiberger and Saul (2006)*



Existing methods (cont.)

Maximum variance unfolding (MVU): *Weiberger and Saul (2006)*



- Let $K = -\frac{1}{2}JDJ$. MVU solves the following SDP:

$$\begin{aligned} \max \quad & \langle I, K \rangle - \nu \sum_{(i,j)\in\Omega_0} \left(K_{ii} - 2K_{ij} + K_{jj} - d_{ij}^2 \right)^2 \\ \text{s.t.} \quad & \langle \mathbf{1}\mathbf{1}^T,\ K \rangle = 0 \quad \text{and} \quad K \succeq 0 \end{aligned}$$

Existing methods (cont.)

Minimum volume embedding (MVE): *Shaw and Jebara (2007)*

- Maximum the eigen gap between the leading r eigenvalues of K and the remaining eigenvalues.

$$\begin{aligned} \max \quad & \sum_{i=1}^{r} \lambda_i(K) - \sum_{i=r+1}^{n} \lambda_i(K) - \nu \sum_{(i,j)\in\Omega_0} \left(K_{ii} - 2K_{ij} + K_{jj} - d_{ij}^2\right)^2 \\ \text{s.t.} \quad & \langle \mathbf{1}\mathbf{1}^T,\ K \rangle = 0 \quad \text{and} \quad K \succeq 0 \end{aligned}$$

Existing methods (cont.)

Minimum volume embedding (MVE): *Shaw and Jebara (2007)*

- Maximum the eigen gap between the leading r eigenvalues of K and the remaining eigenvalues.

$$\begin{aligned} \max \quad & \sum_{i=1}^{r} \lambda_i(K) - \sum_{i=r+1}^{n} \lambda_i(K) - \nu \sum_{(i,j)\in\Omega_0} \left(K_{ii} - 2K_{ij} + K_{jj} - d_{ij}^2\right)^2 \\ \text{s.t.} \quad & \langle \mathbf{1}\mathbf{1}^T,\ K \rangle = 0 \quad \text{and} \quad K \succeq 0 \end{aligned}$$

- MVE is better than MVU for some applications.

Existing methods (cont.)

Minimum volume embedding (MVE): *Shaw and Jebara (2007)*

- Maximum the eigen gap between the leading r eigenvalues of K and the remaining eigenvalues.

$$\begin{aligned} \max \quad & \sum_{i=1}^{r} \lambda_i(K) - \sum_{i=r+1}^{n} \lambda_i(K) - \nu \sum_{(i,j)\in\Omega_0} \left(K_{ii} - 2K_{ij} + K_{jj} - d_{ij}^2\right)^2 \\ \text{s.t.} \quad & \langle \mathbf{1}\mathbf{1}^T,\ K\rangle = 0 \quad \text{and} \quad K \succeq 0 \end{aligned}$$

- MVE is better than MVU for some applications.
- However, MVE is nonconvex and solved by a sequential SDP method.

Existing methods (cont.)

Minimum volume embedding (MVE): *Shaw and Jebara (2007)*

- Maximum the eigen gap between the leading r eigenvalues of K and the remaining eigenvalues.

$$\begin{aligned} \max \quad & \sum_{i=1}^{r} \lambda_i(K) - \sum_{i=r+1}^{n} \lambda_i(K) - \nu \sum_{(i,j)\in\Omega_0} \left(K_{ii} - 2K_{ij} + K_{jj} - d_{ij}^2\right)^2 \\ \text{s.t.} \quad & \langle \mathbf{1}\mathbf{1}^T,\ K\rangle = 0 \quad \text{and} \quad K \succeq 0 \end{aligned}$$

- MVE is better than MVU for some applications.
- However, MVE is nonconvex and solved by a sequential SDP method.

Both MVU and MVE do not have theoretical guarantee as to how good the obtained Euclidean distances are.

Our model

MVU and MVE:

- both are faithful to the observed distances
- both encourage high EDM scores from the first few leading eigenvalues.

Our model

MVU and MVE:

- both are faithful to the observed distances
- both encourage high EDM scores from the first few leading eigenvalues.

$$\text{EDM scores}(r) := \sum_{i=1}^{r} \lambda_i(K) / \sum_{i=1}^{n} \lambda_i(K), \qquad r = 1, 2, \ldots, n.$$

Our model

MVU and MVE:

- both are faithful to the observed distances
- both encourage high EDM scores from the first few leading eigenvalues.

$$\text{EDM scores}(r) := \sum_{i=1}^{r} \lambda_i(K) / \sum_{i=1}^{n} \lambda_i(K), \qquad r = 1, 2, \ldots, n.$$

Our new convex model:

Our model

MVU and MVE:

- both are faithful to the observed distances
- both encourage high EDM scores from the first few leading eigenvalues.

$$\text{EDM scores}(r) := \sum_{i=1}^{r} \lambda_i(K) / \sum_{i=1}^{n} \lambda_i(K), \qquad r = 1, 2, \ldots, n.$$

Our new convex model:

- inherit the good properties of MVU and MVE

Our model

MVU and MVE:

- both are faithful to the observed distances
- both encourage high EDM scores from the first few leading eigenvalues.

$$\text{EDM scores}(r) := \sum_{i=1}^{r} \lambda_i(K) / \sum_{i=1}^{n} \lambda_i(K), \qquad r = 1, 2, \ldots, n.$$

Our new convex model:

- inherit the good properties of MVU and MVE
- derive the statistical error bounds for estimators

Observation models

We assume that the observed distances d_{ij} are sampled in the following fashion.

$$d_{ij} = \overline{d}_{ij} + \eta\xi_{ij}, \qquad (i,j) \in \Omega_0$$

Observation models

We assume that the observed distances d_{ij} are sampled in the following fashion.

$$d_{ij} = \overline{d}_{ij} + \eta\xi_{ij}, \qquad (i,j) \in \Omega_0$$

- Ω_0: the sampled index set;
- $\overline{d}_{ij}$: the unknown true Euclidean distances;
- ξ_{ij} are i.i.d. noise variables with $\mathbb{E}(\xi) = 0$, $\mathbb{E}(\xi^2) = 1$ and $\eta > 0$ is a noise magnitude control factor.

Observation models

We assume that the observed distances d_{ij} are sampled in the following fashion.

$$d_{ij} = \overline{d}_{ij} + \eta\xi_{ij}, \qquad (i,j) \in \Omega_0$$

- Ω_0: the sampled index set;
- $\overline{d}_{ij}$: the unknown true Euclidean distances;
- ξ_{ij} are i.i.d. noise variables with $\mathbb{E}(\xi) = 0$, $\mathbb{E}(\xi^2) = 1$ and $\eta > 0$ is a noise magnitude control factor.

There are three commonly used models to select Ω_0.

- Uniform sampling rule

Observation models

We assume that the observed distances d_{ij} are sampled in the following fashion.

$$d_{ij} = \overline{d}_{ij} + \eta\xi_{ij}, \qquad (i,j) \in \Omega_0$$

- Ω_0: the sampled index set;
- $\overline{d}_{ij}$: the unknown true Euclidean distances;
- ξ_{ij} are i.i.d. noise variables with $\mathbb{E}(\xi) = 0$, $\mathbb{E}(\xi^2) = 1$ and $\eta > 0$ is a noise magnitude control factor.

There are three commonly used models to select Ω_0.

- Uniform sampling rule
- k nearest neighbors (k-NN) rule
- Unit ball rule

Our model

$$\begin{aligned} \min_{D \in S^n} \quad & \frac{1}{2m}\|y \circ y - \mathcal{O}(D)\|^2 + \rho_1 \left(\langle I, -JDJ\rangle - \rho_2 \langle \Theta, -JDJ\rangle\right) \\ \text{s.t.} \quad & D \in \text{EDM}, \quad \|D\|_\infty \leq c. \end{aligned}$$

Our model

$$\begin{aligned} \min_{D \in S^n} \quad & \frac{1}{2m}\|y \circ y - \mathcal{O}(D)\|^2 + \rho_1 \left(\langle I, -JDJ\rangle - \rho_2 \langle \Theta, -JDJ\rangle\right) \\ \text{s.t.} \quad & D \in \text{EDM}, \quad \|D\|_\infty \leq c. \end{aligned}$$

- $\Theta = \widetilde{P}_1 \widetilde{P}_1^T$ and $\widetilde{P}_1$ is the leading eigenvectors of $\widetilde{D}$.

Our model

$$\begin{aligned} \min_{D \in S^n} \quad & \frac{1}{2m}\|y \circ y - \mathcal{O}(D)\|^2 + \rho_1 \left(\langle I, -JDJ\rangle - \rho_2 \langle \Theta, -JDJ\rangle\right) \\ \text{s.t.} \quad & D \in \text{EDM}, \quad \|D\|_\infty \le c. \end{aligned}$$

- $\Theta = \widetilde{P}_1 \widetilde{P}_1^T$ and $\widetilde{P}_1$ is the leading eigenvectors of $\widetilde{D}$.
- $\widetilde{D}$: a given initial estimator

Our model

$$\begin{aligned} \min_{D \in S^n} \quad & \frac{1}{2m}\|y \circ y - \mathcal{O}(D)\|^2 + \rho_1 \left(\langle I, -JDJ\rangle - \rho_2 \langle \Theta, -JDJ\rangle\right) \\ \text{s.t.} \quad & D \in \text{EDM}, \quad \|D\|_\infty \le c. \end{aligned}$$

- $\Theta = \widetilde{P}_1 \widetilde{P}_1^T$ and $\widetilde{P}_1$ is the leading eigenvectors of $\widetilde{D}$.
- $\widetilde{D}$: a given initial estimator
- D is a EDM $\iff$ $\begin{cases} \text{diag}(D) = 0 \\ -D \in \mathcal{K}_+^n \end{cases}$

$$\mathcal{K}_+^n := \left\{ A \in \mathcal{S}^n \mid y^T A y \ge 0 \; \forall \, y \in e^\perp \right\}$$

cf. Schoenberg (1935); Young and Householder (1937)

Model interpretation

If the leading eigenspace $\mathrm{span}\{\widetilde{P}_1\}$ is close to $\mathrm{span}\{P_1\}$, then

Model interpretation

If the leading eigenspace $\mathrm{span}\{\widetilde{P}_1\}$ is close to $\mathrm{span}\{P_1\}$, then

- $-\langle \widetilde{P}_1 \widetilde{P}_1^T,\ -JDJ \rangle$: maximize

$$t := \sum_{i=1}^{r} \lambda_i$$

Model interpretation

If the leading eigenspace $\mathrm{span}\{\widetilde{P}_1\}$ is close to $\mathrm{span}\{P_1\}$, then

- $-\langle \widetilde{P}_1 \widetilde{P}_1^T,\ -JDJ \rangle$: maximize

$$t := \sum_{i=1}^{r} \lambda_i$$

- $\langle I,\ -JDJ \rangle = \| - JDJ \|_*$: find the smallest embedding dimension, i.e., minimize $t + s$ where

$$s := \sum_{i=r+1}^{n} \lambda_i$$

Model interpretation

If the leading eigenspace $\mathrm{span}\{\widetilde{P}_1\}$ is close to $\mathrm{span}\{P_1\}$, then

- $-\langle \widetilde{P}_1 \widetilde{P}_1^T,\ -JDJ \rangle$: maximize

$$t := \sum_{i=1}^{r} \lambda_i$$

- $\langle I,\ -JDJ \rangle = \| -JDJ \|_*$: find the smallest embedding dimension, i.e., minimize $t + s$ where

$$s := \sum_{i=r+1}^{n} \lambda_i$$

Consequently, the EDM score $f(t, s) := \frac{t}{t+s}$ gets higher, since

$$f(t_2, s_2) > f(t_1, s_1) \qquad \forall\ t_2 > t_1 \quad \text{and} \quad s_2 < s_1 .$$

Model interpretation (cont.)

- MVU ($\rho_2 = 2$ and $\Theta = I$): maximize all eigenvalues, i.e.,

$$t + s = \sum_{i=1}^{n} \lambda_i$$

Model interpretation (cont.)

- MVU ($\rho_2 = 2$ and $\Theta = I$): maximize all eigenvalues, i.e.,

$$t + s = \sum_{i=1}^{n} \lambda_i$$

- MVE-SDP ($\rho_2 = 2$ and $\Theta = \widetilde{P}_1 \widetilde{P}_1^T$): minimize

$$\langle I, -JDJ \rangle - 2\langle \widetilde{P}_1 \widetilde{P}_1^T, -JDJ \rangle \approx -t + s = -\sum_{i=1}^{r} \lambda_i + \sum_{i=r+1}^{n} \lambda_i.$$

Model interpretation (cont.)

- MVU ($\rho_2 = 2$ and $\Theta = I$): maximize all eigenvalues, i.e.,

$$t + s = \sum_{i=1}^{n} \lambda_i$$

- MVE-SDP ($\rho_2 = 2$ and $\Theta = \widetilde{P}_1 \widetilde{P}_1^T$): minimize

$$\langle I, -JDJ \rangle - 2\langle \widetilde{P}_1 \widetilde{P}_1^T, -JDJ \rangle \approx -t + s = -\sum_{i=1}^{r} \lambda_i + \sum_{i=r+1}^{n} \lambda_i.$$

- $\rho_2 = 0$: the nuclear norm penalized least squares (NNPLS)

Statistical error bounds

Suppose that the noise magnitude control factor satisfies $\eta < \omega = \|\mathcal{O}(\overline{D}^{(1/2)})\|_\infty$. Assume the sample size m satisfies $m > C_1 n \log(2n)$ for some constant $C_1 > 1$. For any given $\kappa > 1$, let $\rho_1 = O\left(\kappa\eta\omega\sqrt{\frac{\log(2n)}{mn}}\right)$ and $\rho_2 \geq 0$. Then, under uniform sampling rule, there exists a constant $C_2 > 0$ such that with probability at least $1 - 2/n$,

$$\frac{\|D^* - \overline{D}\|^2}{|\Omega|} \leq C_2\left(\left(\kappa\alpha(\rho_2) + 2\right)^2 + \frac{\kappa^2}{(\kappa-1)^2}\left(\alpha(\rho_2) + 2\right)^2\right)\frac{r|\Omega|\log(2n)}{nm},$$

where $|\Omega| := n(n-1)/2$ and

$$\alpha(\rho_2) := \frac{1}{\sqrt{2r}}\|\overline{P}_1\overline{P}_1 - \rho_2\Theta\|.$$

Statistical error bounds

Suppose that the noise magnitude control factor satisfies $\eta < \omega = \|\mathcal{O}(\overline{D}^{(1/2)})\|_\infty$. Assume the sample size m satisfies $m > C_1 n \log(2n)$ for some constant $C_1 > 1$. For any given $\kappa > 1$, let $\rho_1 = O\left(\kappa\eta\omega\sqrt{\frac{\log(2n)}{mn}}\right)$ and $\rho_2 \geq 0$. Then, under uniform sampling rule, there exists a constant $C_2 > 0$ such that with probability at least $1 - 2/n$,

$$\frac{\|D^* - \overline{D}\|^2}{|\Omega|} \leq C_2\left(\left(\kappa\alpha(\rho_2) + 2\right)^2 + \frac{\kappa^2}{(\kappa-1)^2}\left(\alpha(\rho_2) + 2\right)^2\right)\frac{r|\Omega|\log(2n)}{nm},$$

where $|\Omega| := n(n-1)/2$ and

$$\alpha(\rho_2) := \frac{1}{\sqrt{2r}}\|\overline{P}_1\overline{P}_1 - \rho_2\Theta\|.$$

- the sample size $m = O(r(n-1)\log(2n)/2) \ll n(n-1)/2$

Optimal Estimate of ρ_2

- MVU: $(\alpha_{MVU})^2 = \frac{1}{2r}\|\overline{P}_1\overline{P}_1^T - 2I\|^2 \geq \frac{1}{2}$.

Optimal Estimate of ρ_2

- MVU: $(\alpha_{MVU})^2 = \frac{1}{2r}\|\overline{P}_1\overline{P}_1^T - 2I\|^2 \geq \frac{1}{2}$.
- NNPLS: $\rho_2 = 0$, $\alpha(0)^2 \leq \frac{1}{2}$.

Optimal Estimate of ρ_2

- MVU: $(\alpha_{MVU})^2 = \frac{1}{2r}\|\overline{P}_1\overline{P}_1^T - 2I\|^2 \geq \frac{1}{2}$.
- NNPLS: $\rho_2 = 0$, $\alpha(0)^2 \leq \frac{1}{2}$.
- the best choice ρ_2^* is the minimum of $\alpha(\rho_2)$, i.e.,

$$\rho_2^* := \frac{1}{r}\langle \overline{P}_1\overline{P}_1^T, \widetilde{P}_1\widetilde{P}_1^T\rangle = 1 + \frac{1}{r}\langle \overline{P}_1\overline{P}_1^T, \widetilde{P}_1\widetilde{P}_1^T - \overline{P}_1\overline{P}_1^T\rangle.$$

Optimal Estimate of ρ_2

- MVU: $(\alpha_{MVU})^2 = \frac{1}{2r}\|\overline{P}_1\overline{P}_1^T - 2I\|^2 \geq \frac{1}{2}$.
- NNPLS: $\rho_2 = 0$, $\alpha(0)^2 \leq \frac{1}{2}$.
- the best choice ρ_2^* is the minimum of $\alpha(\rho_2)$, i.e.,

$$\rho_2^* := \frac{1}{r}\langle \overline{P}_1\overline{P}_1^T, \widetilde{P}_1\widetilde{P}_1^T\rangle = 1 + \frac{1}{r}\langle \overline{P}_1\overline{P}_1^T, \widetilde{P}_1\widetilde{P}_1^T - \overline{P}_1\overline{P}_1^T\rangle.$$

Proposition

If $\|\widetilde{D} - \overline{D}\| < \overline{\lambda}_r/2$, *then*

$$\alpha(1) < \min\{\alpha(0), \alpha(2)\}.$$

Optimal Estimate of ρ_2

- MVU: $(\alpha_{MVU})^2 = \frac{1}{2r}\|\overline{P}_1\overline{P}_1^T - 2I\|^2 \geq \frac{1}{2}$.
- NNPLS: $\rho_2 = 0$, $\alpha(0)^2 \leq \frac{1}{2}$.
- the best choice ρ_2^* is the minimum of $\alpha(\rho_2)$, i.e.,

$$\rho_2^* := \frac{1}{r}\langle \overline{P}_1\overline{P}_1^T, \widetilde{P}_1\widetilde{P}_1^T\rangle = 1 + \frac{1}{r}\langle \overline{P}_1\overline{P}_1^T, \widetilde{P}_1\widetilde{P}_1^T - \overline{P}_1\overline{P}_1^T\rangle.$$

Proposition

If $\|\widetilde{D} - \overline{D}\| < \overline{\lambda}_r/2$, *then*

$$\alpha(1) < \min\{\alpha(0), \alpha(2)\}.$$

- NNPLS $\succeq$ MVU

Optimal Estimate of ρ_2

- MVU: $(\alpha_{MVU})^2 = \frac{1}{2r}\|\overline{P}_1\overline{P}_1^T - 2I\|^2 \geq \frac{1}{2}$.
- NNPLS: $\rho_2 = 0$, $\alpha(0)^2 \leq \frac{1}{2}$.
- the best choice ρ_2^* is the minimum of $\alpha(\rho_2)$, i.e.,

$$\rho_2^* := \frac{1}{r}\langle \overline{P}_1\overline{P}_1^T, \widetilde{P}_1\widetilde{P}_1^T\rangle = 1 + \frac{1}{r}\langle \overline{P}_1\overline{P}_1^T, \widetilde{P}_1\widetilde{P}_1^T - \overline{P}_1\overline{P}_1^T\rangle.$$

Proposition

If $\|\widetilde{D} - \overline{D}\| < \overline{\lambda}_r/2$, *then*

$$\alpha(1) < \min\{\alpha(0), \alpha(2)\}.$$

- NNPLS $\succeq$ MVU
- EDM $\succeq$ NNPLS, MVE-SDP

A convergent 3-block ADMM algorithm

Consider the following matrix optimization problem:

$$\begin{array}{ll} \min & \frac{1}{2}\|\mathcal{A}(X) - \mathbf{a}\|^2 + \langle C, X\rangle \\ \text{s.t.} & \mathcal{B}(X) = \mathbf{c}, \;\; X \in \mathcal{K}_+^n, \;\; \|X\|_\infty \le b, \end{array}$$

where $\mathcal{K}_+^n$ is the almost positive semidefinite cone.

A convergent 3-block ADMM algorithm

Consider the following matrix optimization problem:

$$\begin{array}{ll} \min & \frac{1}{2}\|\mathcal{A}(X) - \mathbf{a}\|^2 + \langle C, X\rangle \\ \text{s.t.} & \mathcal{B}(X) = \mathbf{c}, \;\; X \in \mathcal{K}_+^n, \;\; \|X\|_\infty \le b, \end{array}$$

where $\mathcal{K}_+^n$ is the almost positive semidefinite cone.

By introducing a new variable $\mathbf{t} = \mathcal{A}(X) - \mathbf{a}$ and a slack variable $W \in \mathcal{S}^n$, we can rewrite it as the following equivalent form primal and dual form:

$$\begin{array}{lll} (P) & \min & \frac{1}{2}\|\mathbf{t}\|^2 + \langle C, X\rangle + \delta_{\mathcal{K}_+^n}(X) + \delta_{\mathbb{B}_b^\infty}(W) \\ & \text{s.t.} & \mathcal{A}(X) - \mathbf{t} = \mathbf{a}, \;\; \mathcal{B}(X) = \mathbf{c}, \;\; X = W \\ (D) & \max & \frac{1}{2}\|\mathbf{y}_1\|^2 + \langle \mathbf{a}, \mathbf{y}_1\rangle + \langle \mathbf{c}, \mathbf{y}_2\rangle + \delta_{(\mathcal{K}_+^n)^*}(S) - \delta^*_{\mathbb{B}_b^\infty}(-Z) \\ & \text{s.t.} & Z + \mathcal{A}^*\mathbf{y}_1 + \mathcal{B}^*\mathbf{y}_2 - S = C \end{array}$$

A convergent 3-block ADMM algorithm

Consider the following matrix optimization problem:

$$\begin{array}{ll} \min & \frac{1}{2}\|\mathcal{A}(X) - \mathbf{a}\|^2 + \langle C, X\rangle \\ \text{s.t.} & \mathcal{B}(X) = \mathbf{c}, \;\; X \in \mathcal{K}_+^n, \;\; \|X\|_\infty \le b, \end{array}$$

where $\mathcal{K}_+^n$ is the almost positive semidefinite cone.

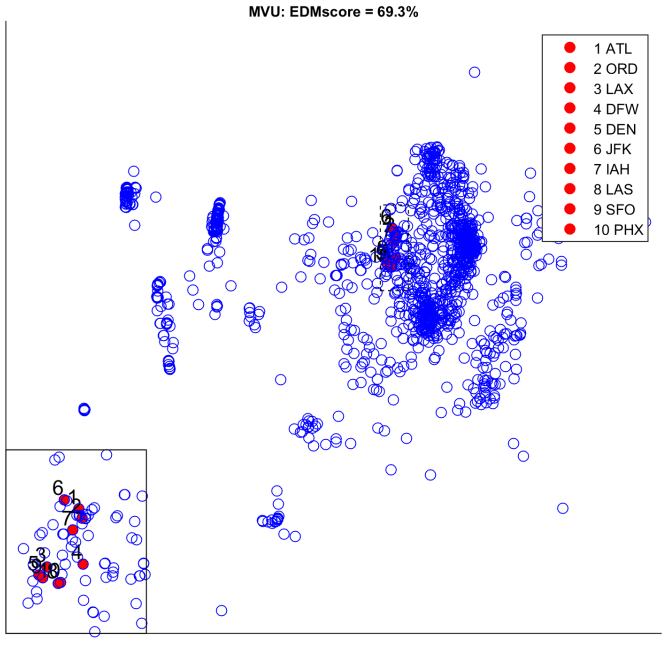
By introducing a new variable $\mathbf{t} = \mathcal{A}(X) - \mathbf{a}$ and a slack variable $W \in \mathcal{S}^n$, we can rewrite it as the following equivalent form primal and dual form:

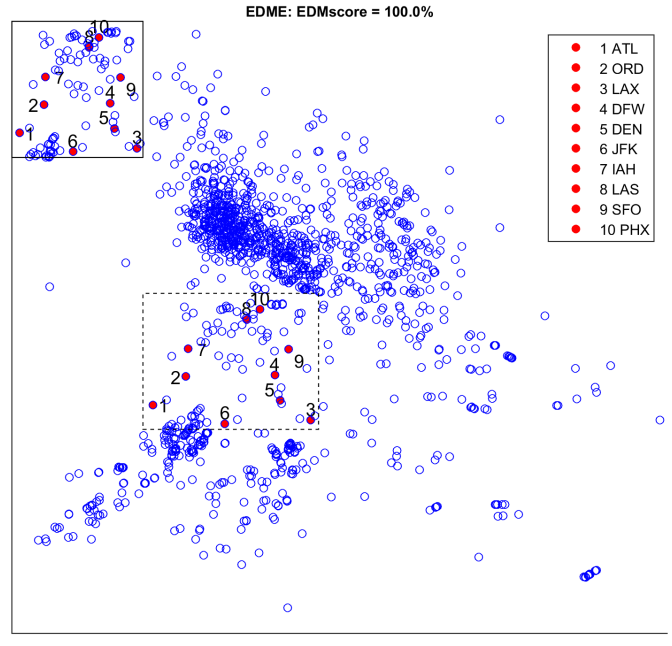
$$\begin{array}{lll} (P) & \min & \frac{1}{2}\|\mathbf{t}\|^2 + \langle C, X\rangle + \delta_{\mathcal{K}_+^n}(X) + \delta_{\mathbb{B}_b^\infty}(W) \\ & \text{s.t.} & \mathcal{A}(X) - \mathbf{t} = \mathbf{a}, \;\; \mathcal{B}(X) = \mathbf{c}, \;\; X = W \\ (D) & \max & \frac{1}{2}\|\mathbf{y}_1\|^2 + \langle \mathbf{a}, \mathbf{y}_1\rangle + \langle \mathbf{c}, \mathbf{y}_2\rangle + \delta_{(\mathcal{K}_+^n)^*}(S) - \delta^*_{\mathbb{B}_b^\infty}(-Z) \\ & \text{s.t.} & Z + \mathcal{A}^*\mathbf{y}_1 + \mathcal{B}^*\mathbf{y}_2 - S = C \end{array}$$

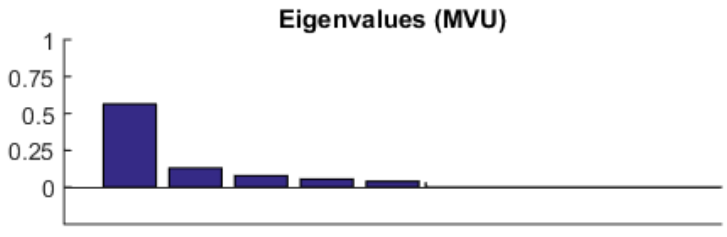
It can be solved by an efficient 3-block ADMM method, which is inspired by the recent work of Li et al., Math. Prog. 2016.

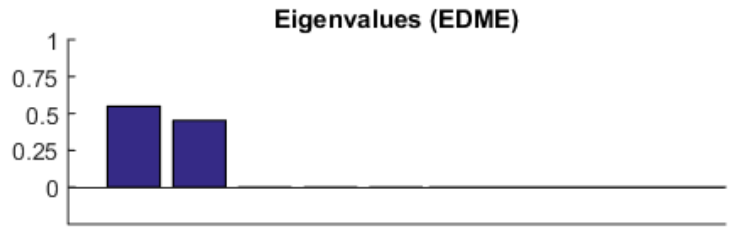
Part I: Social Network

(SN1) US airport network









(SN1) US airport network (cont.)

From the EDME embedding, we can observe that these ten airports are naturally separated into four groups:

- Group 1 $= \{1(\texttt{ATL}), 2(\texttt{ORD}), 7(\texttt{IAH})\}$: three southeast US cities: `Atlanta`, `Orlando` and `Houston`
- Group 2 $= \{6(\texttt{JFK})\}$: one big east-coast city: `New York`
- Group 3 $= \{8(\texttt{LAS}), 10(\texttt{PHX})\}$: two closed related southwest cities: `Las Vegas` and `Phoenix`
- Group 4 $= \{3(\texttt{LAX}), 4(\texttt{DFW}), 5(\texttt{DEN}), 9(\texttt{SFO})\}$: four west cities: `Los Angeles`, `Dallas`, `Denver` and `San Francisco`

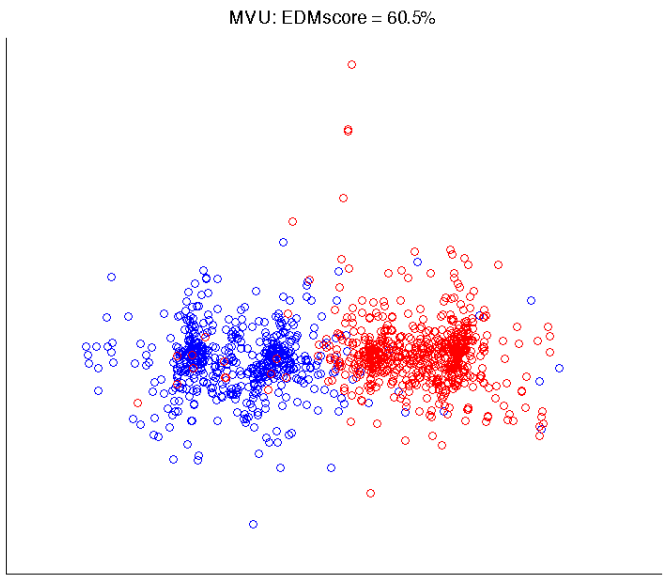
(SN1) US airport network (cont.)

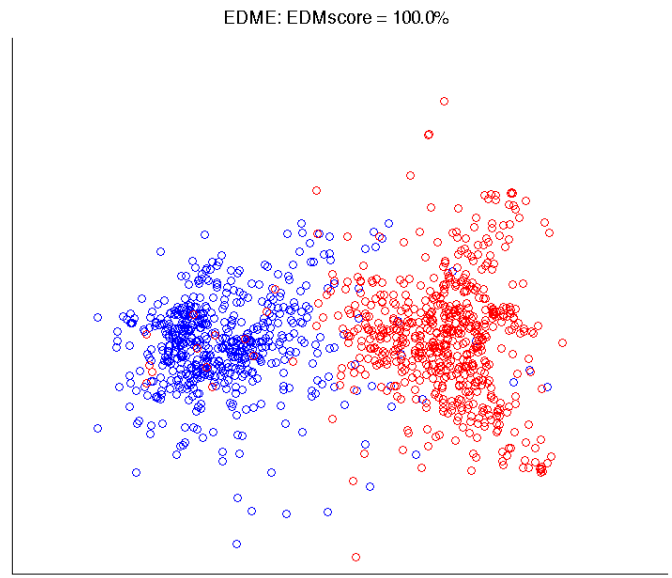
From the EDME embedding, we can observe that these ten airports are naturally separated into four groups:

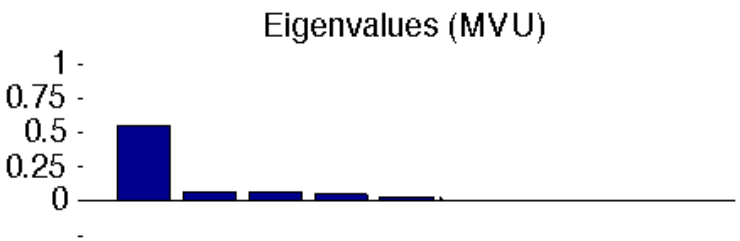
- Group 1 $= \{1(\texttt{ATL}), 2(\texttt{ORD}), 7(\texttt{IAH})\}$: three southeast US cities: `Atlanta`, `Orlando` and `Houston`
- Group 2 $= \{6(\texttt{JFK})\}$: one big east-coast city: `New York`
- Group 3 $= \{8(\texttt{LAS}), 10(\texttt{PHX})\}$: two closed related southwest cities: `Las Vegas` and `Phoenix`
- Group 4 $= \{3(\texttt{LAX}), 4(\texttt{DFW}), 5(\texttt{DEN}), 9(\texttt{SFO})\}$: four west cities: `Los Angeles`, `Dallas`, `Denver` and `San Francisco`

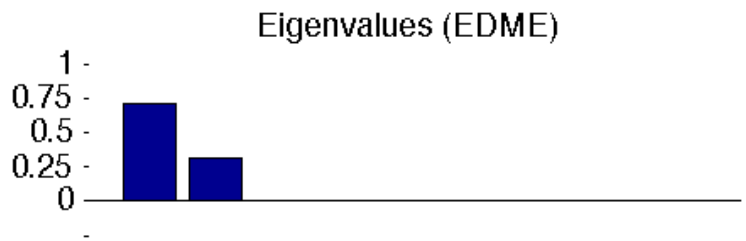
However, by MVU, we can only obtain two groups: one consists of the east cities: $\{\texttt{ATL}, \texttt{ORD}, \texttt{IAH}, \texttt{JFK}\}$, and another consists of the west ones: $\{\texttt{DFW}, \texttt{LAS}, \texttt{PHX}, \texttt{LAX}, \texttt{DEN}, \texttt{SFO}\}$.

(SN2) Political blogs



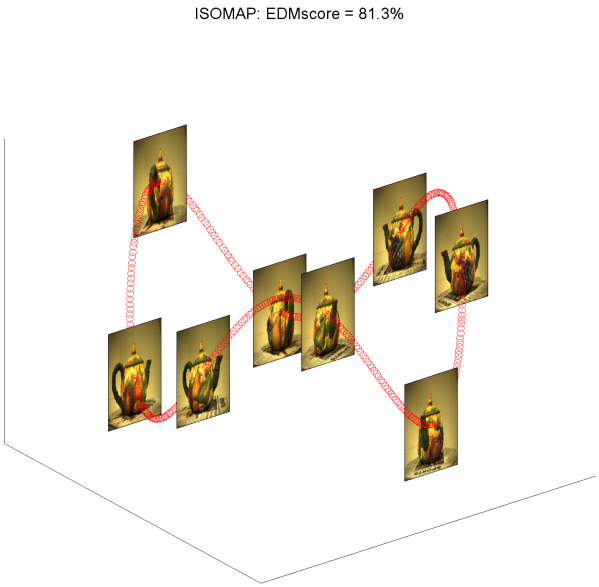


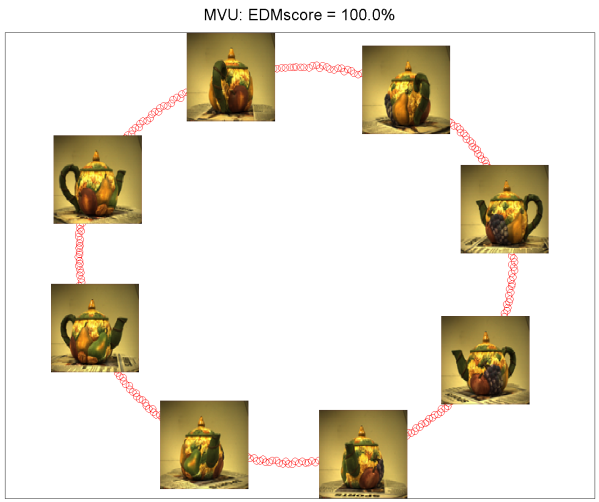


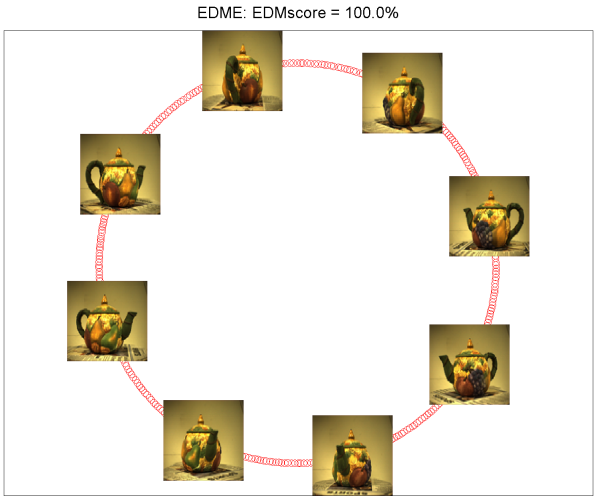


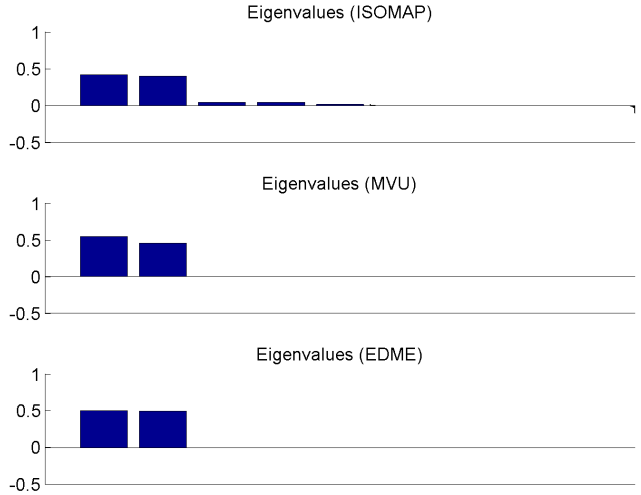
Part II: Manifold learning

(ML1) Teapots data

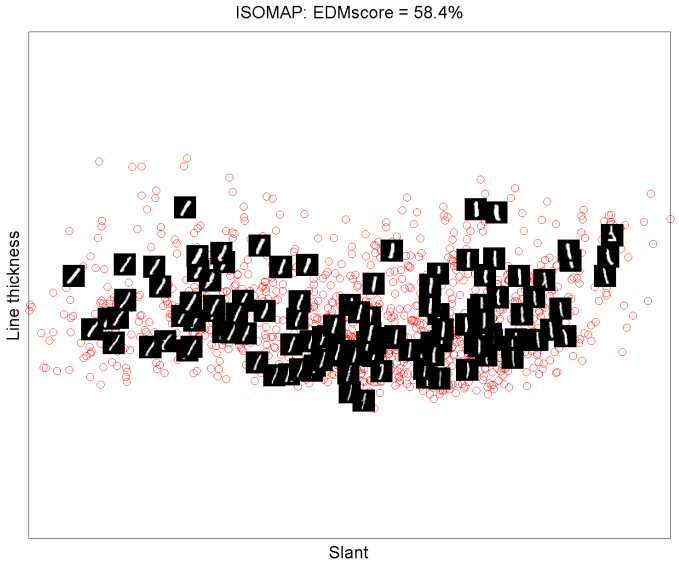


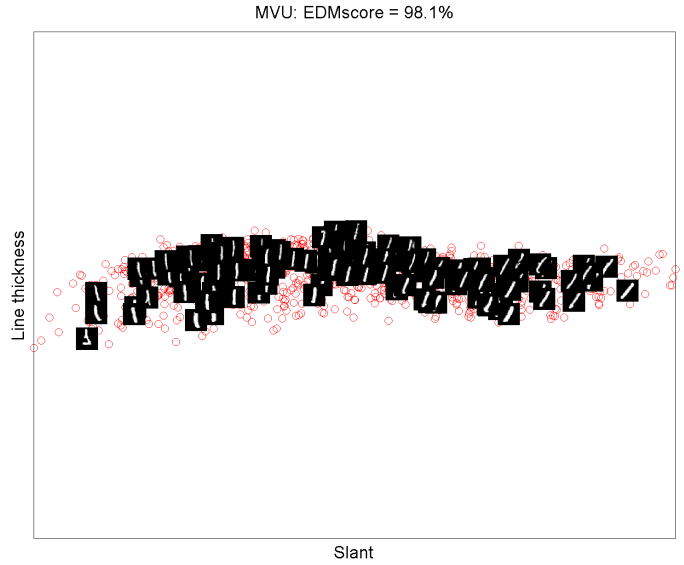


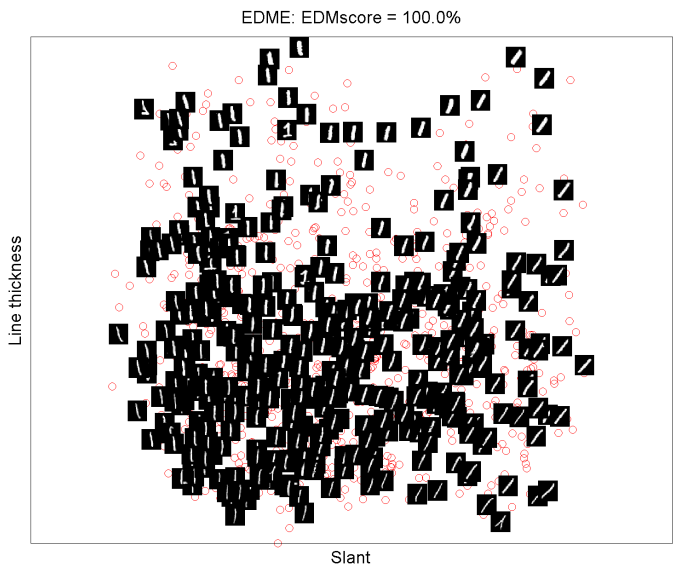


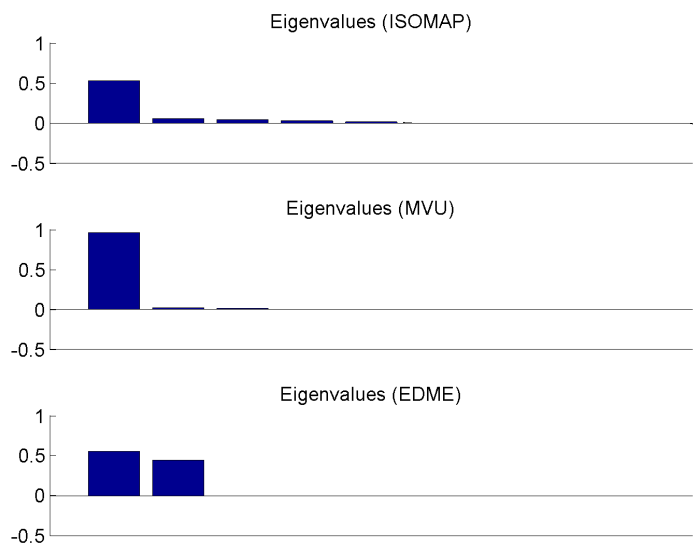


(ML2) The digits Data









Problems		MVU (SDPT3)			EDME		
	n/edges	relgap	EDMscore	cpu(s)	R	EDMscore	cpu(s)
Enron	182/2097	3.25e-04	48.1%	5.46	9.92e-04	100%	1.05
Facebook-like	1893/13835	4.40e-04	20.6%	624.18	9.78e-04	100%	165.34
TrainBombing	64/243	9.38e-04	91.8%	0.49	9.99e-04	100%	0.56
USairport2010	1572/17214	7.30e-04	69.3%	31587.08	9.98e-04	100%	80.53
Blogs	1222/16714	4.71e-04	60.5%	12173.31	8.40e-04	100%	83.66
	k/n/edges	relgap	EDMscore	cpu(s)	R	EDMscore	cpu(s)
Teapots400	5/400/1050	8.45e-04	100%	3.44	9.75e-04	100%	3.77
Face698	5/698/2164	2.96e-04	100%	14.25	9.94e-04	100%	29.73
Digit1	6/1135/4885	7.49e-04	98.1%	68.85	9.95e-04	100%	39.62
Digits19	6/1000/4394	6.57e-04	94.0%	51.02	9.83e-04	100%	27.66
FreyFace	5/1965/6925	9.48e-04	86.2%	214.41	8.72e-04	100%	187.56

Table: Numerical performance comparison of the MVU and the EDME

Thank You.