Low dimensional embedding with compressed, incomplete and inaccurate measurements

Blake Hunter[1,2] Thomas Strohmer[1]

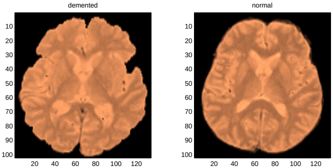
[1]Applied Mathematics, UC Davis
[2]Applied Mathematics, UCLA

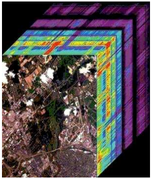
UC Davis Student-Run Seminar
April 6th, 2011

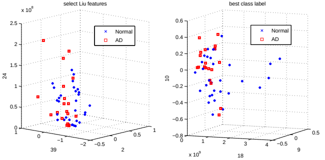
Outline

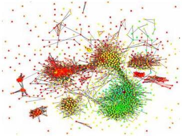
- Motivation
- Background
- Compressive Spectral Clustering
- Matrix Completion

Motivation


demented
normal
10
20
30
40
50
60
70
80
90
100
20
40
60
80
100
120




select Liu features
Normal
AD
best class label
Normal
AD



Spectral Clustering [Shi and Malik]

Spectral Clustering [Shi and Malik]

- Define a local distance between two data points $d(x_i, x_j)$.

Spectral Clustering [Shi and Malik]

- Define a local distance between two data points $d(x_i, x_j)$.
- Construct a graph, vertices are data points and edge weights are given by the Gaussian kernel :

$$W_{i,j} = e^{\frac{-d(x_i,x_j)^2}{2\sigma}}.$$

Spectral Clustering [Shi and Malik]

- Define a local distance between two data points $d(x_i, x_j)$.
- Construct a graph, vertices are data points and edge weights are given by the Gaussian kernel :

$$W_{i,j} = e^{\frac{-d(x_i,x_j)^2}{2\sigma}}.$$

- Let $\underbrace{P = D^{-1}W}_{\textit{random walk}}$ and $\underbrace{A = D^{-\frac{1}{2}}WD^{-\frac{1}{2}}}_{\textit{symmetric rw}}$ where $D_(i,i) = \sum_k W_{i,k}$.

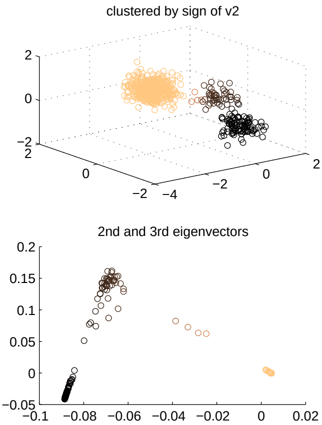
Spectral Clustering [Shi and Malik]

- Define a local distance between two data points $d(x_i, x_j)$.
- Construct a graph, vertices are data points and edge weights are given by the Gaussian kernel :

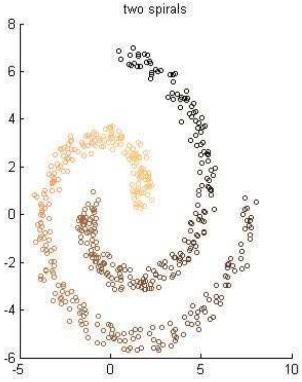
$$W_{i,j} = e^{\frac{-d(x_i,x_j)^2}{2\sigma}}.$$

- Let $\underbrace{P = D^{-1}W}_{random\ walk}$ and $\underbrace{A = D^{-\frac{1}{2}}WD^{-\frac{1}{2}}}_{symmetric\ rw}$ where $D_(i,i) = \sum_k W_{i,k}$.
- Partition the data using the coordinates of the top k eigenvectors v_ks, of the affinity matrix A.

Spectral Embedding

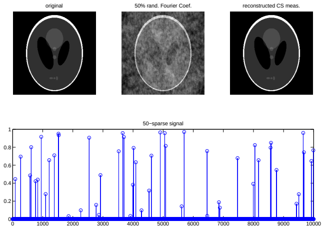


Spectral Embedding



Why it works

- The second eigenvector y of P is the real valued solution of normalized graph cut problem $\frac{e(x,x^c)}{e(x,x^c)} + \frac{e(x,V)}{e(x^c,V)}$,
$$\min_{\mathbf{x}} Ncut(\mathbf{x}) = \min_{\mathbf{y}} \frac{\mathbf{y}^T(D-W)\mathbf{y}}{\mathbf{y}^T D\mathbf{y}}$$
[Shi and Malik].
- The entries, $P_{i,j} = P(x(1) = x_j \mid x(0) = x_i)$, are the one step transition probabilities of moving from x_i to x_j.
- The first few eigenvectors of the normalized Laplacian $\mathcal{L} = I - A$ are discrete approximations of the eigenfunctions of the Laplace-Beltrami operator on a low dimensional manifold, when the data is sampled uniformly. [Belkin and Niyogi]



Compressed Sensing

- Solving underdetermined $Ax = b$ where x is r-sparse.
- Used for exact recovery of sparse signals using fewer measurements than the ambient dimension.
- Compressed sensing provides a bound on the error derived from making these few measurements of a signal.

restricted isometry property [Candès and Tao]

The *restricted isometry property (RIP)* holds for all r-sparse vectors x with parameters (r, δ) if

$$(1-\delta)\|x\|_2 \leq \|\Phi x\|_2 \leq (1+\delta)\|x\|_2.$$

restricted isometry property [Candès and Tao]

The *restricted isometry property (RIP)* holds for all r-sparse vectors x with parameters (r, δ) if

$$(1-\delta)\|x\|_2 \leq \|\Phi x\|_2 \leq (1+\delta)\|x\|_2.$$

Random Bernoulli [Candès, Romberg and Tao], Gaussian, and partial Fourier matrices [Rudelson and Vershynin] satisfy the RIP with exponentially high probability.

restricted isometry property [Candès and Tao]

The *restricted isometry property (RIP)* holds for all r-sparse vectors x with parameters (r, δ) if

$$(1-\delta)\|x\|_2 \leq \|\Phi x\|_2 \leq (1+\delta)\|x\|_2.$$

Theorem - Gaussian matrices [Rudelson and Vershynin]

Let Φ be a $m \times n$ Gaussian measurement matrix, and let $r \geq 1$, $0 < \delta < 1$, and $0 < \epsilon < 0.5$. Then with high probability the matrix $\frac{1}{\sqrt{m}}\Phi$ satisfies the restricted isometry property with parameters (r, δ) provided that the number of measurements m satisfies

$$m \geq \frac{r}{\epsilon^2} \log\left(\frac{n}{\epsilon^2 r}\right).$$

Can equivalent spectral clustering be done in the compressed domain?

Compressive Spectral Clustering

- Let Φ be a random $m \times n$ matrix, with Gaussian $\mathcal{N}(0, 1)$ entries.

Compressive Spectral Clustering

- Let Φ be a random $m \times n$ matrix, with Gaussian $\mathcal{N}(0,1)$ entries.
- Define $\tilde{d}(x_i, x_j) = \|\Phi x_i - \Phi x_j\|_2$ using m compressed sensing measurements, where $m << n$.

Compressive Spectral Clustering

- Let Φ be a random $m \times n$ matrix, with Gaussian $\mathcal{N}(0, 1)$ entries.
- Define $\tilde{d}(x_i, x_j) = \|\Phi x_i - \Phi x_j\|_2$ using m compressed sensing measurements, where $m << n$.
- Construct a graph with edge weights

$$\tilde{W}(x_i, x_j) = \exp\left(-\frac{\|\Phi x_i - \Phi x_j\|_2^2}{2\sigma}\right).$$

Compressive Spectral Clustering

- Let Φ be a random $m \times n$ matrix, with Gaussian $\mathcal{N}(0,1)$ entries.
- Define $\tilde{d}(x_i, x_j) = \|\Phi x_i - \Phi x_j\|_2$ using m compressed sensing measurements, where $m << n$.
- Construct a graph with edge weights

$$\tilde{W}(x_i, x_j) = \exp\left(-\frac{\|\Phi x_i - \Phi x_j\|_2^2}{2\sigma}\right).$$

- Define $\tilde{A} = \tilde{D}^{-\frac{1}{2}} \tilde{W} \tilde{D}^{-\frac{1}{2}}$ where $\tilde{D}_{i,i} = \sum_k \tilde{W}(x_i, x_k)$.

Compressive Spectral Clustering

- Let Φ be a random $m \times n$ matrix, with Gaussian $\mathcal{N}(0,1)$ entries.
- Define $\tilde{d}(x_i, x_j) = \|\Phi x_i - \Phi x_j\|_2$ using m compressed sensing measurements, where $m << n$.
- Construct a graph with edge weights

$$\tilde{W}(x_i, x_j) = \exp\left(-\frac{\|\Phi x_i - \Phi x_j\|_2^2}{2\sigma}\right).$$

- Define $\tilde{A} = \tilde{D}^{-\frac{1}{2}} \tilde{W} \tilde{D}^{-\frac{1}{2}}$ where $\tilde{D}_{i,i} = \sum_k \tilde{W}(x_i, x_k)$.
- Use the first k eigenvectors of $\tilde{A}$ as a k low-dimensional representation of the data and coordinates for clustering.

- traditional spectral clustering $x \to A \to v_i$

- traditional spectral clustering $x \rightarrow A \rightarrow v_i$
- compressive spectral clustering $\Phi x \rightarrow \tilde{A} \rightarrow \tilde{v}_i$.

- traditional spectral clustering $x \to A \to v_i$
- compressive spectral clustering $\Phi x \to \tilde{A} \to \tilde{v}_i$.
- $\|\Phi x_i - \Phi x_j\| \approx \|x_i - x_j\|$ by the RIP
- $v_i \approx \tilde{v}_i$?

- traditional spectral clustering $x \to A \to v_i$
- compressive spectral clustering $\Phi x \to \tilde{A} \to \tilde{v}_i$.
- $\|\Phi x_i - \Phi x_j\| \approx \|x_i - x_j\|$ by the RIP
- $v_i \approx \tilde{v}_i$?

Question

Can the first k eigenvectors of $\tilde{A}$ provide the same spectral coordinates as the first k eigenvectors of traditional spectral clustering?

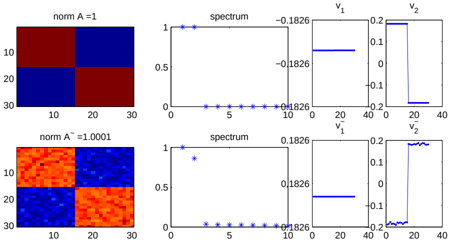
Earlier results have shown that spectral clustering using the second eigenvector is robust to small perturbation of the data.

Theorem [Stewart]

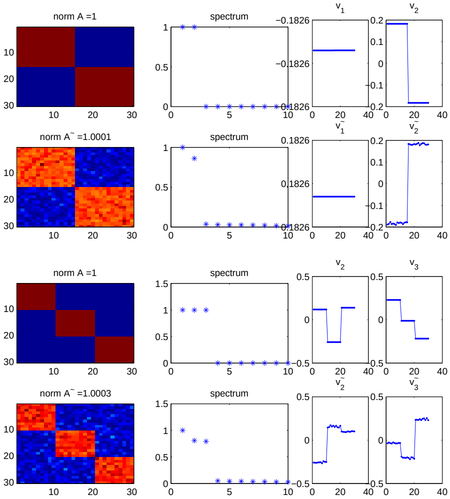
Let $\tilde{A} = A + E$ be a perturbation of A and let λ_i and v_i be the i^{th} eigenvalue and eigenvector of A and $\tilde{v}_i$ be the i^{th} eigenvector of $\tilde{A}$ respectively, then

$$\|\tilde{v}_2 - v_2\| \leq \frac{1}{\lambda_2 - \lambda_3}\|E\| + O\left(\|E\|^2\right).$$

Perturbation of Eigenvectors



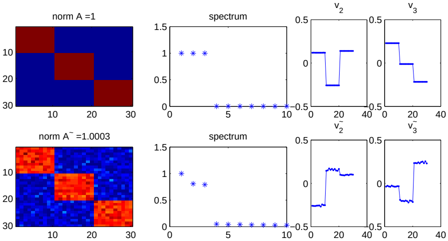
Perturbation of Eigenvectors



$$\left(V_1 V_2\right)^H A \left(V_1 V_2\right) = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix},$$

and

$$\left(\tilde{V}_1 \tilde{V}_2\right)^H \tilde{A} \left(\tilde{V}_1 \tilde{V}_2\right) = \begin{pmatrix} \tilde{\Sigma}_1 & 0 \\ 0 & \tilde{\Sigma}_2 \end{pmatrix}.$$



$$\left(V_1 V_2\right)^H A \left(V_1 V_2\right) = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix},$$

and

$$\left(\tilde{V}_1 \tilde{V}_2\right)^H \tilde{A} \left(\tilde{V}_1 \tilde{V}_2\right) = \begin{pmatrix} \tilde{\Sigma}_1 & 0 \\ 0 & \tilde{\Sigma}_2 \end{pmatrix}.$$

Canonical Angles

Let $\mathcal{V}_m$ and $\tilde{\mathcal{V}}_m$ be subspaces spanned by the orthonormal eigenvectors $v_i, \ldots, v_{i+m}$ and $\tilde{v}_i, \ldots, \tilde{v}_{i+m}$. And let $\gamma_1 \leq \ldots \leq \gamma_m$ be the singular values of $\left[v_i \cdots v_{i+m}\right]^T \left[\tilde{v}_i \cdots \tilde{v}_{i+m}\right]$. Then the values,

$$\theta_i = \cos^{-1} \gamma_i$$

are called the **canonical angles** between $\mathcal{V}_m$ and $\tilde{\mathcal{V}}_m$. Define $\mathcal{V}_m$ and $\tilde{\mathcal{V}}_m$ to be close if the largest canonical angle, θ_1, is small.

Let λ_i, v_i, $\tilde{\lambda}_i$, $\tilde{v}_i$ be the i^{th} eigenvalues and eigenvectors of A and $\tilde{A}$ respectively, and let Θ be the canonical angles between the column space of $V_k = [v_1, v_2, \ldots, v_k]$ and $\tilde{V}_k = [\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_k]$.

Theorem [Davis and Kahan]

If there is an $\alpha > 0$ such that

$$|\tilde{\lambda}_k - \lambda_{k+1}| \geq \alpha$$

and

$$\tilde{\lambda}_k \geq \alpha$$

then

$$\|\sin\Theta\|_F \leq \frac{1}{\alpha}\|A\tilde{V}_k - \tilde{V}_k\tilde{\Sigma}_k\|_F.$$

Corollary

Let P_{V_k} and $P_{\tilde{V}_k}$ be the orthogonal projection on to V_k and $\tilde{V}_k$. If there is a $\alpha > 0$ such that $\lambda_k - \lambda_{k+1} \geq \alpha$ and $\lambda_k \geq \alpha$, then,

$$\|P_{V_k} - P_{\tilde{V}_k}\|_F \leq \frac{\sqrt{2}}{\alpha}\|A - \tilde{A}\|_F.$$

Corollary

Let P_{V_k} and $P_{\tilde{V}_k}$ be the orthogonal projection on to V_k and $\tilde{V}_k$. If there is a $\alpha > 0$ such that $\lambda_k - \lambda_{k+1} \geq \alpha$ and $\lambda_k \geq \alpha$, then,

$$\|P_{V_k} - P_{\tilde{V}_k}\|_F \leq \frac{\sqrt{2}}{\alpha}\|A - \tilde{A}\|_F.$$

Theorem

Let V_k be the matrix formed by the top k column eigenvectors of A and $\tilde{V}_k$, the matrix formed by the top k eigenvectors of $\tilde{A}$ defined above. If Q is the orthogonal matrix that minimizes $\|\tilde{V}_k - V_k Q\|_F$, then

$$\|\tilde{V}_k - V_k Q\|_2 \leq (1 + \sqrt{2})\frac{1}{\alpha}\|A - \tilde{A}\|_F.$$

Corollary

Let P_{V_k} and $P_{\tilde{V}_k}$ be the orthogonal projection on to V_k and $\tilde{V}_k$. If there is a $\alpha > 0$ such that $\lambda_k - \lambda_{k+1} \geq \alpha$ and $\lambda_k \geq \alpha$, then,

$$\|P_{V_k} - P_{\tilde{V}_k}\|_F \leq \frac{\sqrt{2}}{\alpha}\|A - \tilde{A}\|_F.$$

Theorem

Let V_k be the matrix formed by the top k column eigenvectors of A and $\tilde{V}_k$, the matrix formed by the top k eigenvectors of $\tilde{A}$ defined above. If Q is the orthogonal matrix that minimizes $\|\tilde{V}_k - V_k Q\|_F$, then

$$\|\tilde{V}_k - V_k Q\|_2 \leq (1 + \sqrt{2})\frac{1}{\alpha}\|A - \tilde{A}\|_F.$$

- $\tilde{V}_k$ can be made arbitrarily close to a unitary transform of V_k

Assume that there is an underlying r-sparse representation $y_i = Bx_i$ of the data x_i.

Theorem

Let $W_{i,j} = e^{-\frac{\|x_i - x_j\|_2^2}{2\sigma}}$ and $A = D^{-1/2} W D^{-1/2}$ where $D_{i,i} = \sum_{k=1}^{N} W(x_i, x_k)$. And let $\tilde{A} = \tilde{D}^{-\frac{1}{2}} \tilde{W} \tilde{D}^{-\frac{1}{2}}$ where $\tilde{D}_{i,i} = \sum_{k=1}^{N} \tilde{W}(x_i, x_k)$. If $\tilde{W}_{i,j} = e^{-\frac{\|\Phi x_i - \Phi x_j\|_2^2}{2\sigma}}$ where the x_is are r-sparse and Φ satisfies the RIP with

$$\delta = \frac{\epsilon}{4 \max_{i,j} \left\{ \frac{\|x_i - x_j\|_2^2}{2\sigma} \right\}},$$

then for $0 < \epsilon < 1$,

$$|\tilde{A}_{i,j} - A_{i,j}| \leq \epsilon.$$

Theorem

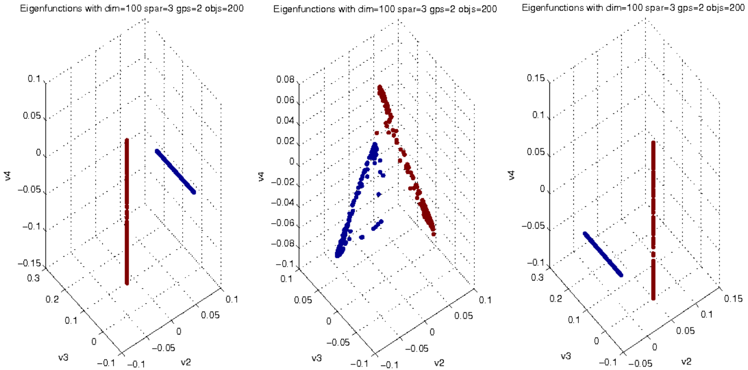
Let $v(i)$ be the i^{th} row of the matrix $V_k =$ formed by the top k column eigenvectors of A and $\tilde{v}(i)$ be the i^{th} row of $\tilde{V}_k$ formed by the top k eigenvectors of $\tilde{A}$ from taking compresses measurements. If there exists an $\alpha > 0$ such that $\lambda_k - \lambda_{k+1} \geq \alpha$ and $\lambda_k \geq \alpha$, then

$$\|\tilde{v}(i) - v(i)Q\|_2 \leq (1+\sqrt{2})\frac{N}{\alpha}\epsilon$$

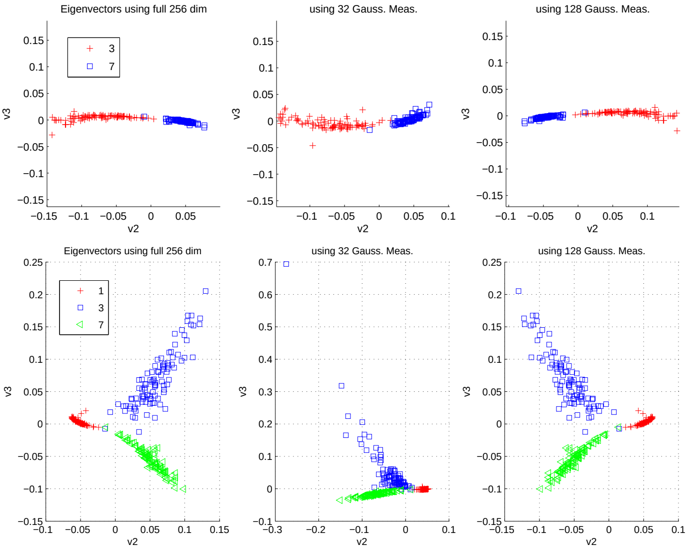
where Q is the orthogonal matrix that minimizes $\|\tilde{V}_k - V_k Q\|_F$.

- The first k eigenvectors of compressive spectral clustering provide the same spectral coordinates as the first k eigenvectors of traditional spectral clustering.

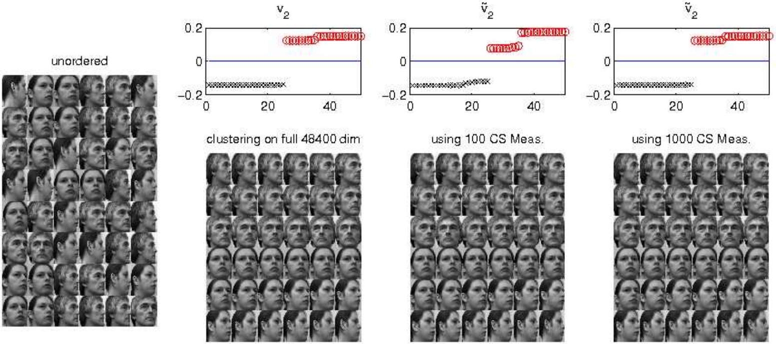
Results


Eigenfunctions with dim=100 spar=3 gps=2 objs=200
v4
v3
v2
Eigenfunctions with dim=100 spar=3 gps=2 objs=200
v4
v3
v2
Eigenfunctions with dim=100 spar=3 gps=2 objs=200
v4
v3
v2

Results


Eigenvectors using full 256 dim
using 32 Gauss. Meas.
using 128 Gauss. Meas.
3
7
v3
v2
Eigenvectors using full 256 dim
using 32 Gauss. Meas.
using 128 Gauss. Meas.
1
3
7
v3
v2

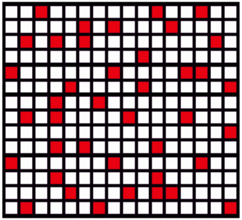
Results


unordered
v_2
0.2
0
−0.2
0
20
40
clustering on full 48400 dim
$\tilde{v}_2$
using 100 CS Meas.
$\tilde{v}_2$
using 1000 CS Meas.

- traditional spectral clustering $x \rightarrow A \rightarrow v_i$
- compressive spectral clustering $\Phi x \rightarrow \tilde{A} \rightarrow \tilde{v}_i$.
- $\|\Phi x_i - \Phi x_j\| \approx \|x_i - x_j\| \rightarrow v_i Q \approx \tilde{v}_i$
- The first k eigenvectors of compressive spectral clustering provide the same spectral coordinates as the first k eigenvectors of traditional spectral clustering.

Question

Can this be extended to incomplete data via matrix completion?



Matrix Completion

Can we recover a matrix $M \in \mathbb{R}^{n_1 \times n_2}$ from a small subset of its entries?

In general this is impossible.

Surprisingly we can recover many matrices of interest from "incomplete" sampled entries, via convex programming.

Which Matrices?

$$M = e_1 e_n^* = \begin{bmatrix} 0 & \dots & 0 & 1 \\ 0 & \dots & 0 & 0 \\ \vdots & \dots & \vdots & \vdots \\ 0 & \dots & 0 & 0 \end{bmatrix} \qquad \text{rank } 1$$

can not be recovered from a sampling set.

Which entries are observed?

$$M = xy^* = \begin{bmatrix} \times & \times & \dots & \times & \times & \times \\ \\ \times & \times & \dots & \times & \times & \times \\ \vdots & & \dots & & \vdots & \vdots \\ \times & \times & \dots & \times & \times & \times \end{bmatrix} \quad \text{outer product}$$

$$M_{ij} = x_i y_j$$

If a single row (or column)is not sampled $\rightarrow$ recovery is not possible.
What happen for almost all sampling sets?
Ω subset of entries selected uniformly at random.

$$M = e_1 x^* = \begin{bmatrix} x_1 & x_2 & & \dots & x_n \\ 0 & 0 & \dots & & 0 \\ \vdots & \vdots & \dots & & \vdots \\ 0 & 0 & \dots & & 0 \end{bmatrix} \quad \text{rank } 1$$

can not be recovered from a sampling set.

intuition

The column and row space can not supported on a small set of the indices.

The singular vectors can not be aligned with basis vectors.

strong incoherence property [Candès, Recht and Tao]

A matrix X obeys the **strong incoherence property** with parameter μ if the following hold.

1. Let P_Z be the orthogonal projection onto the singular vectors $z_1, \ldots, z_r$ of $X \in \mathbb{R}^{N \times n}$ of rank r. For all pairs $(a, a') \in [N] \times [N]$ and $(b, b') \in [n] \times [n]$,
$$\left|\langle e_a, P_Z e_{a'}\rangle - \frac{r}{N} 1_{a=a'}\right| \leq \mu \frac{\sqrt{r}}{N}, \quad \left|\langle e_b, P_Y e_{b'}\rangle - \frac{r}{n} 1_{b=b'}\right| \leq \mu \frac{\sqrt{r}}{n}.$$
2. Let Ξ be the "sign matrix" defined by $\Xi = \sum_{i \in [r]} z_i y_i^*$. For all $(a, b) \in [N] \times [n]$,
$$|\Xi_{a,b}| \leq \mu \frac{\sqrt{r}}{\sqrt{Nn}}.$$

Theorem [Candés and Plan]

Let $X \in \mathbb{R}^{N \times n}$ be a fixed matrix of rank r obeying the strong incoherence property with parameter μ. Suppose we observe a fraction $p = (\#\text{ of entries observed})/(Nn)$ of entries of X with locations sampled uniformly at random with noise $\|P_\Omega(X) - P_\Omega(\hat{X})\|_F \leq \delta$. Then with high probability the solution to the matrix completion problem $\hat{X}$ satisfies,

$$\|X - \hat{X}\|_F \leq 4\sqrt{\frac{(2+p)\min(N,n)}{p}}\delta + 2\delta.$$

Perfect recovery of most low-rank matrices can be achieved with $O(r \times N \log^2 N)$ samples [Candès and Tao].

Theorem

Let $W_{i,j} = e^{-\frac{\|x_i - x_j\|_2^2}{2\sigma}}$ and $A = D^{-1/2} W D^{-1/2}$ where $D_{i,i} = \sum_{k=1}^{N} W(x_i, x_k)$. And let $\hat{W}_{i,j} = e^{-\frac{\|\hat{x}_i - \hat{x}_j\|_2^2}{2\sigma}}$ and $\hat{A} = \hat{D}^{-\frac{1}{2}} \hat{W} \hat{D}^{-\frac{1}{2}}$ where $\hat{D}_{i,i} = \sum_{k=1}^{N} \hat{W}(x_i, x_k)$. Assume the data matrix X obeys the strong incoherence property with parameter μ under the same assumptions of Candés-Plan. If

$$4\sqrt{\frac{(2+p)\min(N,n)}{p}}\delta + 2\delta \leq \frac{\epsilon}{16 \max_{i,j}\left\{\frac{\|x_i - x_j\|_2}{2\sigma}\right\}}$$

then with high probability,

$$|\hat{A}_{i,j} - A_{i,j}| \leq \epsilon.$$

Theorem

Let $v(i)$ be the i^{th} row of the matrix $V_k =$ formed by the top k column eigenvectors of A and $\hat{v}(i)$ be the i^{th} row of $\hat{V}_k$ formed by the top k eigenvectors of $\hat{A}$ from taking compresses measurements. Given there is a $\alpha > 0$ such that $\lambda_k - \lambda_{k+1} \geq \alpha$ and $\lambda_k \geq \alpha$, if

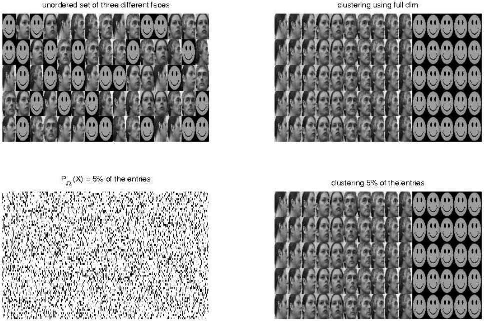
$$4\sqrt{\frac{(2+p)\min(N,n)}{p}}\delta + 2\delta \leq \frac{\epsilon}{16\max_{i,j}\left\{\frac{\|x_i - x_j\|_2}{2\sigma}\right\}}$$

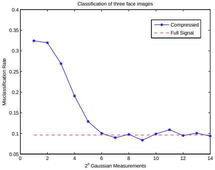
then with high probability

$$\|\hat{v}(i) - v(i)Q\|_2 \leq (1+\sqrt{2})\frac{N\epsilon}{\alpha}.$$

- The first k eigenvectors of spectral clustering with matrix completion provide the same spectral coordinates as the first k eigenvectors of traditional spectrral clustering.

Results


unordered set of three different faces
clustering using full dim
$P_\Omega(X)$ = 5% of the entries
clustering 5% of the entries


Classification of three face images
Compressed
Full Signal
Misclassification Rate
2^x Gaussian Measurements
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
2
4
6
8
10
12
14

- traditional spectral clustering $x \to A \to \mathcal{V}_k$
- compressive spectral clustering $\Phi x \to \tilde{A} \to \tilde{\mathcal{V}}_k$.
- spectral clustering with matrix completion $\hat{x} \to \hat{A} \to \hat{\mathcal{V}}_k$.
- $\mathcal{V}_k \approx \tilde{\mathcal{V}}_k Q_1 \approx \hat{\mathcal{V}}_k Q_2$
- $v(i) \approx \tilde{v}(i) Q_1 \approx \hat{v}(i) Q_2$

Theorem

The first k eigenvectors of the adjacency matrix formed from compressed sensing measurements or matrix completion provide the same spectral coordinates as the first k eigenvectors of traditional spectral clustering.

For more information

E-mail: blakehunter@math.ucdavis.edu
Web: http://www.math.ucdavis.edu/~blakehunter

References:

- M. Belkin and P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (2003), no. 6, 13731396.
- D. L. Donoho, Compressed sensing, Communications on Pure and Applied Mathematics 52.
- **B. Hunter**, T. Strohmer, "Performance Analysis of Spectral Clustering on Compressed, Incomplete and Inaccurate Measurements." Submitted for publication.
- J. Romberg E. Candes and T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Communications on Pure and Applied Mathematics 59.
- J. Shi and J. Malik, Normalized cuts and image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (1997), 888905.
- G. W. Stewart and J. Sun, Matrix perturbation theory, Academic Press,

Diffusion Distance

- W induces a distance between two data points, based on their dynamic proximity.

$$D_t^2(x_i, x_j) = \sum_k \lambda_k^{2t} \left(\psi_k(x_i) - \psi_k(x_j)\right)^2$$

Diffusion Map

- the mapping $x \mapsto (\lambda_1^m \psi_1(x), \lambda_2^m \psi_2(x), \dots, \lambda_k^m \psi_k(x))$
- gives k dimensional parametrization of the geometry and density of the data
- normalized graph Laplacian
- first non-trivial eigenvector $\approx$ optimal normalized cut
- connection to the underlying potential $U(x)$ and geometry of Ω