Exploiting Data-Dependent Structure for Improving Sensor Acquisition and Integration

Alexander Cloninger

Norbert Wiener Center
Department of Mathematics
University of Maryland, College Park
http://www.norbertwiener.umd.edu

April 14, 2014

Norbert Wiener Center for Harmonic Analysis and Applications

Outline

Outline

Data-Dependent Structure

- Advancements in sensor construction and production cost has led to a deluge of data
- Thesis utilizes data-dependent operators to discover efficient representations of data
- This focus on learning structure splits into three topics
  1. Building data-dependent graphs to capture structure and detect anomalous objects
  2. Fusing low-dimensional parameters from heterogeneous data sources
  3. Exploiting compressibility of data to reduce sampling requirements prior to collection

Reduced Acquisition Time

- Based on the theory of compressive sensing and matrix completion
  - Recover signal that is *sparse* in some basis
  - Key is that measurements are randomly made and *incoherent* with respect to sparsity basis
  - Utilizes convex relaxation and optimization schemes to reconstruct signal
  - Reconstruction only requires $\mathrm{O}(K \log N)$ measurements

Contributions of Thesis

- Proved *bounded norm Parseval frames* satisfy necessary conditions for matrix reconstruction
- Demonstrated use of matrix completion for solving 2D Fredholm integrals from incomplete measurements
- Improved acquisition time for nuclear magnetic resonance spectroscopy via reducing necessary number of samples

Fusing Low Dimensional Parameters of High Dimensional Data

- Based on graph and operator theoretic approaches to pattern recognition and machine learning
  - Builds operator that encodes similarity between data points
  - Takes data from high-dimensional data space and embeds into low-dimensional euclidean space
  - Allows common comparison across heterogeneous sensors

Contributions of Thesis

- Built approximate inversion algorithm for Laplacian eigenmaps that utilizes compressive sensing
- Used inversion along with Coifman and Hirn's graph rotation to create data fusion algorithm
- Reconstructed missing LIDAR data (altitudes) from hyperspectral camera images (electromagnetic spectrum frequencies)

Laplacian Eigenmaps

- Let $\Omega = \{x_1, ...x_n\} \subset \mathbb{R}^d$ be a set of data points, or *data space*
- Idea is to learn structure via inter-data similarities
- Encode relationships via symmetric kernel $k : \Omega \times \Omega \to [0, 1]$
  - Gaussian kernel, $k(x, y) = e^{-\frac{\|x-y\|_2^2}{2\sigma^2}}$
  - Mahalanobis distance, $k(x, y) = e^{-(x-y)^T S^{-1}(x-y)}$
  - Graph adjacency, $k(x, y) = \begin{cases} 1 & : \quad x \in \mathcal{N}(y), \\ 0 & : \quad \text{otherwise.} \end{cases}$
- Build graph $G = (\Omega, E, W)$, where $\{x, y\} \in E \iff k(x, y) \approx 1$
  - $W_{x,y} = k(x, y)$ if $\{x, y\} \in E$
  - k-Nearest Neighbors
  - ϵ-Nearest Neighbors
- Key is that G is sparse

Norbert Wiener Center for Harmonic Analysis and Applications

Laplacian Eigenmaps (cont.)

- Calculate the normalized graph Laplacian $L = I - D^{-1/2} W D^{-1/2}$, where $D_{x,x} = \sum_y W_{x,y}$
- Solve the eigenvalue problem

$$L\phi_i = \lambda_i \phi_i$$

  - $0 = \lambda_0 \leq \lambda_1 \leq ... \leq \lambda_{n-1} \leq 2$
  - $\langle \phi_i, \phi_j \rangle = 0$ for $i \neq j$

LE Embedding

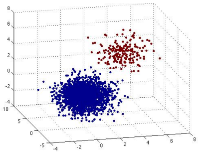
$$\begin{aligned} \Phi \quad &: \quad \Omega \to \mathbb{R}^m \\ x \quad &\mapsto \quad [\phi_1(x), ..., \phi_m(x)] \end{aligned}$$

- Forms low dimensional embedding that preserves local neighborhood structure
- Minimizes $\sum_{x,y} \|\Phi(x) - \Phi(y)\| \frac{W_{x,y}}{\sqrt{D_{x,x} D_{y,y}}}$

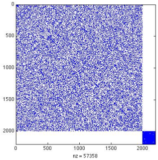
Graph Representation of Data Set

- Maps points from complicated data space to Euclidean feature space
  - $D_{LE}(x, y) = \|\Phi(x) - \Phi(y)\|_2$
- Can be used to reduce dimension of data

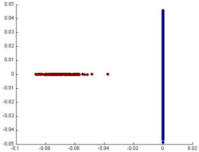
Original Data



Adjacency Matrix



LE Embedding



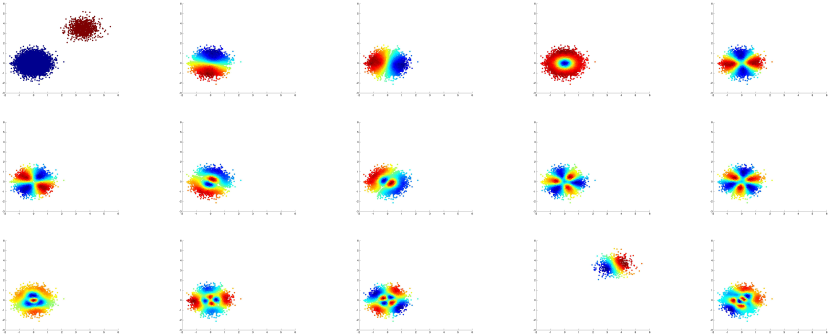
Norbert Wiener Center
for Harmonic Analysis and Applications

Complicated Distribution of Eigenvectors

- Most literature simply utilizes "first m eigenvectors with non-zero eigenvalue"
  - These correspond to "low frequency" information on graph
- When in doubt, simply be liberal with choice of m
- However, distribution of eigenvectors more complicated
  - Do not simply correspond to 1 eigenvector concentrated on each cluster
- Rest of talk is examination of eigenvector localization and order of emergence
  - Specifically when clusters are differing sizes

Examples of Small Clusters Failing to Emerge

Eigenvectors with non-zero eigenvalues



$|C_1| = 10,000, \ |C_2| = 1,000$

Norbert Wiener Center for Harmonic Analysis and Applications

Outline of Approach

- Assume graph $G = (\Omega, E)$ already formed from data, under some metric and using k-NN
- For simplicity, assume $\{x, y\} \in E \iff y \in \mathcal{N}(x)$ and $w_{x,y} = 1$
  - Approximates behavior of LE while utilizing vast literature on regular graph
- Wish to examine emergence of small clusters in eigenvectors
- Approach:
  1. Characterize eigenpairs of disjoint graphs with heterogeneous sized clusters
  2. Demonstrate that, upon adding edges to connect graph, eigenpairs do not deviate far from those of disjoint graph

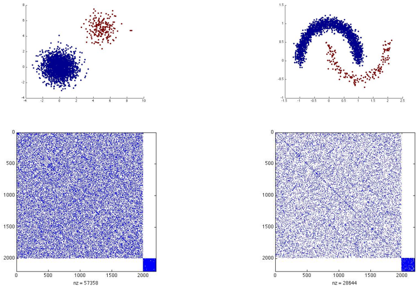
Outline of Approach

- Assume graph $G = (\Omega, E)$ already formed from data, under some metric and using k-NN
- For simplicity, assume $\{x, y\} \in E \iff y \in \mathcal{N}(x)$ and $w_{x,y} = 1$
  - Approximates behavior of LE while utilizing vast literature on regular graph
- Wish to examine emergence of small clusters in eigenvectors
- Approach:
  1. Characterize eigenpairs of disjoint graphs with heterogeneous sized clusters
  2. Demonstrate that, upon adding edges to connect graph, eigenpairs do not deviate far from those of disjoint graph

Outline

Similarity of Data Generated Graphs

- By analyzing graph, can bypass specifics of data set
- Characteristics such as convexity and scale can be ignored



Clusters as Regular Graphs

- Need way to characterize data clusters
- Define data cluster on n nodes to be *random regular graph*

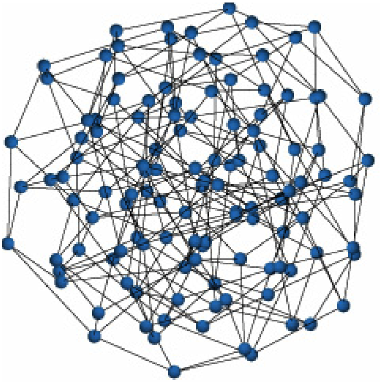
Definition (Family of Regular Graphs)

The family of regular graphs $\mathcal{G}_{n,k}$ is the set of all graphs $G = (V, E)$ such that:

1. V contains n nodes
2. $\forall x \in V, \deg(x) \equiv |\{y \in V : \{x, y\} \in E\}| = k$.

- Random regular graph is $G \in \mathcal{G}_{n,k}$ chosen uniformly at random from all graphs
- With high probability, G does not have large cycles or large complete subgraphs

Random Regular Graphs



Donetti, Neri, and Muño 2006

Validity of Regular Graph Assumption

- Ties into k-Nearest Neighbors edges for graph
- If ignoring need for weights to be symmetric, then exactly generates k-regular graph
- Following theory also applies for Erdös Renyi graph

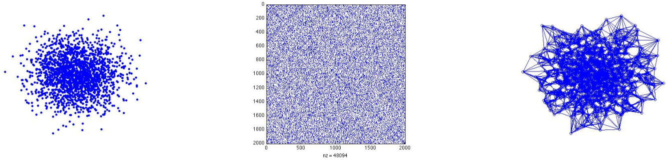
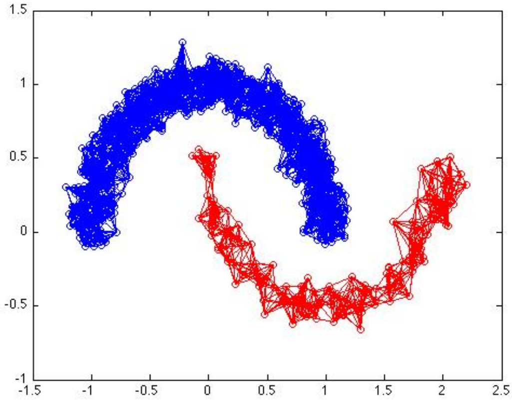


Figure: $\mu_{degree} = 24.05$, $\sigma_{degree} = 1.41$

Data Structure to Keep in Mind for Section



$$|C_1| = 2000, |C_2| = 200$$

Eigenvalues Determine Order of LE Feature Vectors

- Order in which LE eigenvectors appear determined by eigenvalue order
- Goal:
  - Characterize eigenvalues of two graph clusters separately
  - Examine interlacing of eigenvalues to determine order of features emerging
- Eigenvalue distribution of k-regular graph is well studied question
  - McKay, 1981 - showed empirical spectral distribution of $\frac{1}{\sqrt{k-1}}A_n$ converges to

$$f_{semi}(x) = \frac{1}{2\pi}\sqrt{4-x^2}, \quad -2 < x < 2$$

  - Dumitriu, Pal 2013 - found deviation from f_{semi} for finite graph
  - Independently found by Tran, Vu, and Wang 2013

Eigenvalues Determine Order of LE Feature Vectors

- Order in which LE eigenvectors appear determined by eigenvalue order
- Goal:
    - Characterize eigenvalues of two graph clusters separately
    - Examine interlacing of eigenvalues to determine order of features emerging
- Eigenvalue distribution of k-regular graph is well studied question
    - McKay, 1981 - showed empirical spectral distribution of $\frac{1}{\sqrt{k-1}}A_n$ converges to

$$f_{semi}(x) = \frac{1}{2\pi}\sqrt{4-x^2}, \quad -2 < x < 2$$

    - Dumitriu, Pal 2013 - found deviation from f_{semi} for finite graph
    - Independently found by Tran, Vu, and Wang 2013

Eigenvalues of Regular Graph (cont.)

Theorem (Dumitriu, Pal, 2012)

Fix $\delta > 0$ and let $k = (\log(n))^{\gamma}$, and let $\eta = \frac{1}{2}(\exp(k^{-\alpha}) - \exp(-k^{-\alpha}))$, for $0 < \alpha < \min(1, 1/\gamma)$. Then for

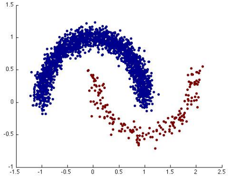
- *$G \in \mathcal{G}_{n,k}$ chosen randomly with adjacency matrix A, and*
- *interval $\mathcal{I} \subset \mathbb{R}$ such that $|\mathcal{I}| \geq \max\{2\eta, \eta/(-\delta \log \delta)\}$,*

there exists an N large enough such that $\forall n > N$,

$$|\mathcal{N}_{\mathcal{I}} - n \int_{\mathcal{I}} f_{semi}(x)dx| < n\delta|\mathcal{I}|$$

with probability at least $1 - o(1/n)$. Here, $\mathcal{N}_I$ is the number of eigenvalues of $\frac{1}{\sqrt{k-1}}A$ in the interval $\mathcal{I}$.

Eigenvalues for Disjoint Clusters



- Consider two clusters C_1 and C_2 with $|C_1| = D|C_2|$
  - G_1 and G_2 are generated graphs on C_1 and C_2, respectively
  - $G_1 \in \mathcal{G}_{n,k}$ and $G_2 \in \mathcal{G}_{\frac{n}{D},k}$
- $\sigma\left(\frac{1}{\sqrt{k-1}}A_1\right)$ and $\sigma\left(\frac{1}{\sqrt{k-1}}A_2\right)$ distributed similarly due to Dumitriu and Pal
  - $\sigma(L_1)$ and $\sigma(L_2)$ distributed similarly on $[0, 2]$
- Eigenvalues interweave in way that depends on D

Eigenvalues for Disjoint Clusters (cont.)

Theorem (C., 2014)

Let $G = (\Omega, E)$ be graph. Suppose Ω can be split into two disjoint regular graph clusters C_1 and C_2 such that $|C_1| = D|C_2| = n$. Choose any interval $\mathcal{I} \subset [0, 2]$ such that

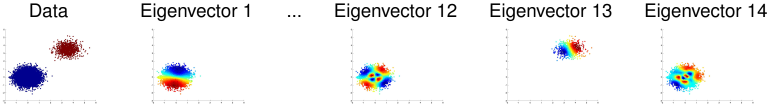
$$|\mathcal{I}| \geq \frac{\sqrt{k-1}}{k} \max\{2\eta, \eta/(-\delta \log \delta)\}.$$

Let L denote the graph Laplacian, with eigenpairs $\{(\sigma_i, v_i)\}_{i=1}^m$ that lie in $\mathcal{I}$. Let $\mathcal{N}_{\mathcal{I}}^1 = |\{i : \text{supp}(v_i) \subset C_1\}|$ and $\mathcal{N}_{\mathcal{I}}^2 = |\{i : \text{supp}(v_i) \subset C_2\}|$. Then $\mathcal{N}_{\mathcal{I}}^1 + \mathcal{N}_{\mathcal{I}}^2 = m$, and $\forall n > N$, with probability at least $1 - o(1/n)$,

$$|\mathcal{N}_{\mathcal{I}}^1 - D\mathcal{N}_{\mathcal{I}}^2| \leq 2\delta n \frac{k}{\sqrt{k-1}} |\mathcal{I}|.$$

Eigenvector Localization on Disjoint Graphs

- Theorem implies each eigenvector is localized on either C_1 or C_2
- Up to error, eigenvector on C_2 appears approximately $1 : D + 1$ times
  - Implies most of energy from LE embedding lies in C_1
- Applies for any interval $\mathcal{I} \subset [0, 2]$
- Can be generalized to larger number of clusters
- Argument explains initial example shown $(D = 10)$



Sketch of Proof for Disjoint Graphs

- $L = \begin{pmatrix} L_1 & 0 \\ 0 & L_2 \end{pmatrix}$
  - $\sigma(L) = \sigma(L_1) \cup \sigma(L_2)$
  - $L \begin{pmatrix} v \\ 0 \end{pmatrix} = \lambda \begin{pmatrix} v \\ 0 \end{pmatrix} \iff L_1 v = \lambda v$
  - $L \begin{pmatrix} 0 \\ v \end{pmatrix} = \lambda \begin{pmatrix} 0 \\ v \end{pmatrix} \iff L_2 v = \lambda v$
- Thus all eigenvectors v of L concentrated on one cluster C_i
  - Order determined by $\sigma(L_1)$ and $\sigma(L_2)$
- Rescale $\sigma\left(\frac{1}{\sqrt{k-1}}A\right)$ from Dumitriu and Pal Theorem to $\sigma(L)$
  - Because G is k-regular,

$$\frac{1}{\sqrt{k-1}}Av_i = \lambda_i v_i \iff Lv_i = (1 - \frac{\sqrt{k-1}}{k}\lambda_i)v_i$$

Sketch of Proof for Disjoint Graphs

- $L = \begin{pmatrix} L_1 & 0 \\ 0 & L_2 \end{pmatrix}$
  - $\sigma(L) = \sigma(L_1) \cup \sigma(L_2)$
  - $L \begin{pmatrix} v \\ 0 \end{pmatrix} = \lambda \begin{pmatrix} v \\ 0 \end{pmatrix} \iff L_1 v = \lambda v$
  - $L \begin{pmatrix} 0 \\ v \end{pmatrix} = \lambda \begin{pmatrix} 0 \\ v \end{pmatrix} \iff L_2 v = \lambda v$
- Thus all eigenvectors v of L concentrated on one cluster C_i
  - Order determined by $\sigma(L_1)$ and $\sigma(L_2)$
- Rescale $\sigma\left(\frac{1}{\sqrt{k-1}} A\right)$ from Dumitriu and Pal Theorem to $\sigma(L)$
  - Because G is k-regular,

$$\frac{1}{\sqrt{k-1}} A v_i = \lambda_i v_i \iff L v_i = (1 - \frac{\sqrt{k-1}}{k} \lambda_i) v_i$$

Sketch of Proof (cont.)

- Design parameters from Dumitriu and Pal Theorem that are constant across both clusters C_i
  - Interval $\mathcal{I}$ for $\sigma(L)$ has corresponding interval $\mathcal{I}_A$ for $\sigma(\frac{1}{\sqrt{k-1}}A)$
- Theorem guarantees that

$$|\mathcal{N}^1_{\mathcal{I}_A} - n\int_{\mathcal{I}_A} f_d(x)dx| < n\delta|\mathcal{I}_A|,$$
$$|\mathcal{N}^2_{\mathcal{I}_A} - \frac{n}{D}\int_{\mathcal{I}_A} f_d(x)dx| < \frac{n}{D}\delta|\mathcal{I}_A|.$$

- This means

$$\begin{aligned}|\mathcal{N}^1_{\mathcal{I}} - D\mathcal{N}^2_{\mathcal{I}}| &\leq |\mathcal{N}^1_{\mathcal{I}_A} - n\int_{\mathcal{I}_A} f_d(x)dx| + |D\mathcal{N}^2_{\mathcal{I}_A} - n\int_{\mathcal{I}_A} f_d(x)dx| \\ &\leq 2n\delta|\mathcal{I}_A| \\ &= 2n\delta\frac{k}{\sqrt{k-1}}|\mathcal{I}|\end{aligned}$$

Norbert Wiener Center for Harmonic Analysis and Applications

Sketch of Proof (cont.)

- Design parameters from Dumitriu and Pal Theorem that are constant across both clusters C_i
  - Interval $\mathcal{I}$ for $\sigma(L)$ has corresponding interval $\mathcal{I}_A$ for $\sigma(\frac{1}{\sqrt{k-1}}A)$
- Theorem guarantees that

$$|\mathcal{N}^1_{\mathcal{I}_A} - n\int_{\mathcal{I}_A} f_d(x)dx| < n\delta|\mathcal{I}_A|,$$
$$|\mathcal{N}^2_{\mathcal{I}_A} - \frac{n}{D}\int_{\mathcal{I}_A} f_d(x)dx| < \frac{n}{D}\delta|\mathcal{I}_A|.$$

- This means

$$\begin{aligned}|\mathcal{N}^1_{\mathcal{I}} - D\mathcal{N}^2_{\mathcal{I}}| &\leq |\mathcal{N}^1_{\mathcal{I}_A} - n\int_{\mathcal{I}_A} f_d(x)dx| + |D\mathcal{N}^2_{\mathcal{I}_A} - n\int_{\mathcal{I}_A} f_d(x)dx| \\ &\leq 2n\delta|\mathcal{I}_A| \\ &= 2n\delta\frac{k}{\sqrt{k-1}}|\mathcal{I}|\end{aligned}$$

Disjoint Graph Conclusions

- Important notes from Theorem
  - Characterizes order of feature vectors from LE
  - Demonstrates that, among first m eigenvectors, $\frac{D}{D+1}$ of them are concentrated in largest cluster
  - Attempt to design LE similarity kernel such that graph as disjoint as possible
  - Arguments generalize to larger number of clusters
- Drawbacks
  - In practice, cannot design disconnected graph from data
  - Need to add edges to connect graph for better theory
  - Already know Fiedler vector is highly sensitive to connecting edge

Disjoint Graph Conclusions

- Important notes from Theorem
  - Characterizes order of feature vectors from LE
  - Demonstrates that, among first m eigenvectors, $\frac{D}{D+1}$ of them are concentrated in largest cluster
  - Attempt to design LE similarity kernel such that graph as disjoint as possible
  - Arguments generalize to larger number of clusters
- Drawbacks
  - In practice, cannot design disconnected graph from data
  - Need to add edges to connect graph for better theory
  - Already know Fiedler vector is highly sensitive to connecting edge

Outline

Weakly Connected Clusters

Definition

A graph with *weakly connected clusters of order t* is a connected graph with adjacency matrix

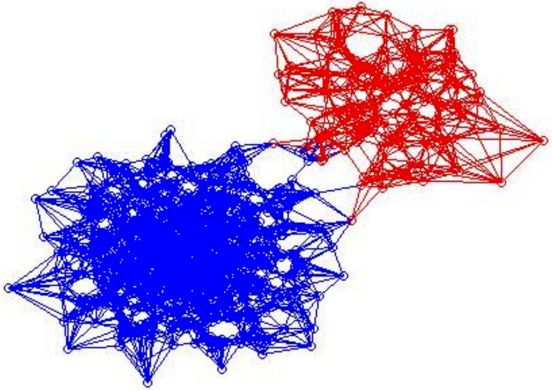
$$A = \begin{pmatrix} A_1 & B_{1,2} \\ B_{1,2}^T & A_2 \end{pmatrix},$$

where

1. A_1 and A_2 are adjacency matrices of k-regular graphs, and
2. $B_{1,2}$ has t non-zero entries.

We shall refer to the nodes of A_1 as C_1 and the nodes of A_2 as C_2

Weakly Connected Clusters Example



Weakly Connected Clusters as Matrix Perturbation

- Now problem is characterized by perturbation of known matrix
  - G is weakly connected graph with adjacency $A = \begin{pmatrix} A_1 & B_{1,2} \\ B_{1,2}^T & A_2 \end{pmatrix}$
  - H is disjoint graph with adjacency $A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$
- Let L_G be normalized Laplacian of G, and similar for L_H

Perturbation of L_H

$L_G = L_H + E$, where $\|E\|_F \ll \|L_H\|_F$

- Questions:
  1. Is eigenvalue ordering of L_G drastically affected?
  2. Are eigenvectors of L_G still concentrated on clusters?

Eigenvalue Distribution for GWCC

Theorem (Chen, et. al., 2012)

Let $G = (\Omega, E_G)$ and $H = (\Omega, E_H)$ be spanning graphs such that $|E(G - H)| \leq t$. If

$$\lambda_1 \leq \ldots \leq \lambda_n, \text{ and } \theta_1 \leq \ldots \leq \theta_n$$

are the eigenvalues of the normalized Laplacians L_G and L_H respectively, then

$$\theta_{i-t} \leq \lambda_i \leq \theta_{i+t}, \qquad 1 \leq i \leq n,$$

with the convention that $\theta_{-t} = \ldots = \theta_0 = 0$ and $\theta_{n+1} = \ldots = \theta_{n+t} = 2$.

- Related to Weyl's inequality and Courant-Fischer theorem
- Shows why lowest eigenvalues difficult to predict
- Will lead to issues with Fiedler vector

Eigenvalue Distribution for GWCC (cont.)

Lemma (C., 2014)

Let $G = (\Omega, E)$ be a graph with weakly connected clusters of order t, with

- $|C_1| = n$,
- $|C_2| = \frac{n}{D}$.

Fix δ, k, α, η, and $\mathcal{I}$ as in Theorem for disjoint clusters. Let L denote the graph Laplacian, and $\sigma_1, \ldots, \sigma_m$ denote the m eigenvalues of L that lie in $\mathcal{I}$. Then m satisfies

$$|m - (n + \frac{n}{D}) \int_{\mathcal{I}} f_{semi}(x)dx| < \delta(n + \frac{n}{D}) \frac{k}{\sqrt{k-1}} |\mathcal{I}| + 2t, \qquad (1)$$

again with probability at least $1 - o(1/n)$.

Invariant Subspace Perturbations

- Eigenvectors under perturbation require more careful treatment
- Dependent on separation of spectrum

Example

Let $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \implies \sigma(A) = \{1\}$, $V = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Let $\widetilde{A} = \begin{pmatrix} 1 & \epsilon \\ \epsilon & 1 \end{pmatrix} \implies \sigma(\widetilde{A}) = \{1-\epsilon, 1+\epsilon\}$, $\widetilde{V} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$.

- Introduced by Davis 1963 for single eigenvector
- Generalized by Davis and Kahan 1970
- “Matrix Perturbation Theory” by Stewart and Sun 1990
- Localization of QR, LU, and Cholesky by Krishtal, Strohmer, and Wertz
- Studied on graphs by Rajapakse 2013

Invariant Subspace Perturbations (cont.)

Theorem (Davis, 1963)

Let $A, E \in \mathbb{C}^{n\times n}$. Let (λ, x) be an eigenpair of A such that

$$\mathrm{sep}(\lambda, \sigma(A) \setminus \lambda) = \min\{|\lambda - \gamma| : \gamma \in \sigma(A) \setminus \lambda\} = \delta.$$

Let

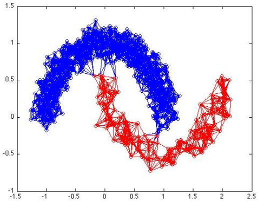
- *P be a spectral projector of A such that $Px = x$*
- *P' be the corresponding spectral projector of $A + E$, and*
- *$\overline{P'}z = z - P'z$.*

Then if $\|E\| \leq \epsilon \leq \delta/2$,

$$\|\overline{P'}P\| \leq \frac{\epsilon}{\delta - \epsilon}.$$

Poor Prediction Using Existing Theory

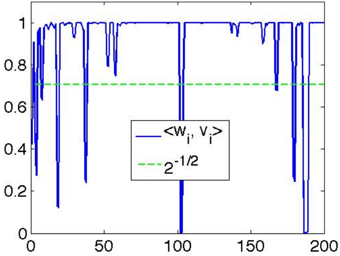
- Consider two moons example
  - $\{(\widetilde{\lambda}_i, v_i)\}$ eigenpairs of weakly connected graph L_G
  - $\{(\lambda_i, w_i)\}$ eigenpairs of disjoint graph L_H
  - Generate L_H by removing off block diagonal entries



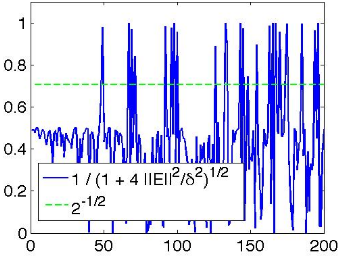
- $\sigma(L_H) \in [0, 2]$ is not sufficiently separated for existing theory
- Assumptions in literature are too strict for problem
- Also we are interested in localization, not angle

Poor Prediction Using Existing Theory (cont.)

Actual Vector Angles



Predicted Vector Angles



Eigengap Dependence on Similar Eigenvectors

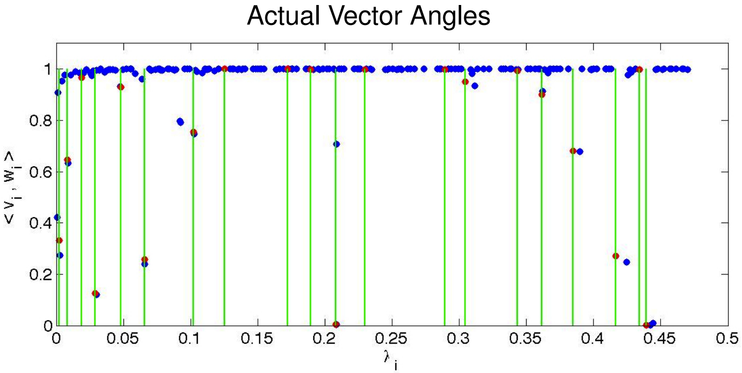
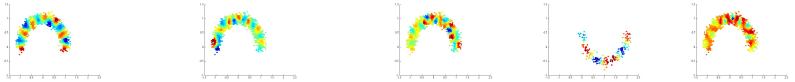


Figure: Green Line Denotes Eigenvector of L_H Concentrated on Smaller C_2 Cluster

Eigengap Dependence on Similar Eigenvectors (cont.)

- Problem eigenvector w_i follows pattern
  - $\text{supp}(w_i) \subset C_2$
  - $\text{supp}(w_{i-1}) \subset C_1$
  - $\text{supp}(w_{i+1}) \subset C_1$
- This is case for most eigenvectors from smaller C_2 cluster
  - $|C_1| = D|C_2| \implies \frac{D}{D+1}$ eigenvectors of L_H concentrated on C_1
- Consider eigenvector w_{25} of L_H as example

$\lambda_{20} = 0.0287$ ... $\lambda_{25} = 0.0371$ ... $\lambda_{30} = 0.0479$ $\lambda_{31} = 0.0481$ $\lambda_{32} = 0.0494$



Eigenvector Localization

Theorem (C., 2014)

*Let $L_H \in \mathbb{R}^{n\times n}$ be symmetric with eigendecomposition $L_H = V\Sigma V^*$.
Let (λ_i, v_i) be an eigenpair of L_H.
Partition $V = [V_1, V_2, v_i, V_3, V_4]$ where $V_2, V_3 \in \mathbb{R}^{n\times s}$.
Moreover, assume $\exists C \subsetneq \{1, ..., n\}$ such that $supp(v_i) \subset C$ and $supp(v_j) \subset C$ where v_j is a column of V_2, V_3.
Let $(\widetilde{\lambda}, x)$ an eigenvector of the perturbed matrix $L_G = L_H + E$, where $x = [x_1, ..., x_n]$. Then*

$$\sum_{j\in C^c} |x_j|^2 \leq \frac{\|(\widetilde{\lambda} - \lambda_i)x - Ex\|_2^2}{\min(\lambda_i - \lambda_{i-s}, \lambda_{i+s} - \lambda_i)^2}.$$

- Apply SVD Theorem to symmetric matrix $L_H - \lambda_i I$
  - SVD equivalence with eigendecomposition up to parity

Singular Vector Localization

Theorem (C., 2014)

Let $A \in \mathbb{R}^{n \times n}$ with SVD $A = U\Sigma V^*$. Partition

$$V = [V_1, V_2, v_n],$$

where $v_n \in \mathbb{R}^n$, $V_2 \in \mathbb{R}^{n \times s}$.
Moreover, assume $\exists C \subsetneq \{1, ..., n\}$ such that

$$\mathrm{supp}(v_i) \subset C \quad \text{for} \quad i \in \{n-s, ..., n\}.$$

Let $x \in \mathbb{R}^n$ such that $\|x\|_2 = 1$. Then

$$\sum_{i \in C^c} |x_i|^2 \leq \frac{\|Ax\|_2^2 - \|Av_n\|_2^2}{\sigma_{n-s-1}^2(A) - \sigma_n^2(A)}.$$

Sketch of Proof for SVD Localization

Assume $x = V_1c_1 + V_2c_2 + v_nc_3$.
Bound

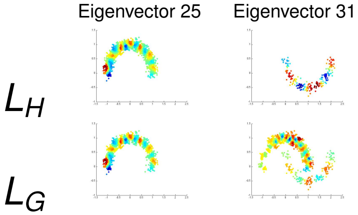
$$\|Ax\|_2^2 = \|U_1\Sigma_1V_1^*x + U_2\Sigma_1V_2^*x + u_n\sigma_nv_n^*x\|_2^2$$
$$\implies \|Ax\|_2^2 - \|Av_n\|_2^2 \geq (\sigma_{n-s-1}^2 - \sigma_n^2)\|c_1\|_2^2$$
$$\implies \|c_1\|_2^2 \leq \frac{\|Ax\|_2^2 - \|Av_n\|_2^2}{\sigma_{n-s-1}^2 - \sigma_n^2}.$$

Using the localization of V_2,

$$\sum_{i\in C^c}|x_i|^2 \leq \sum_{i=1}^{n}\sum_{j=1}^{n-s-1}|(V_1)_{i,j}c_j|^2$$
$$= \sum_{j=1}^{n-s-1}|c_j|^2$$
$$\leq \frac{\|Ax\|_2^2 - \|Av_n\|_2^2}{\sigma_{n-s-1}^2 - \sigma_n^2}.$$

Norbert Wiener Center for Harmonic Analysis and Applications

Eigenvector Localization Conclusions

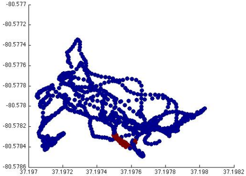


- Important notes from Theorem
  - Theorem implies 124 of 180 eigenvectors supported on C_1 remain concentrated
  - Only 3 of 20 eigenvectors supported on C_2 remain concentrated
  - Makes determining inter-cluster differences difficult for small clusters

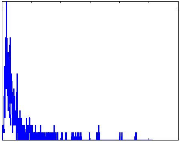
Outline

Nuclear Data

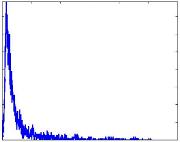
- Wish to detect anomalous material that emit radiation
- Build LE graph from radiological spectra



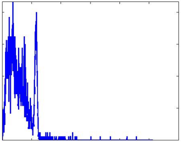
Background 1



Background 2



Anomalous Material

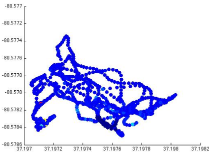


Data courtesy of Kevin Kochersberger, Virginia Tech

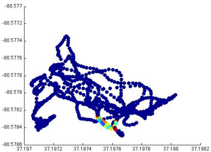
Anomalous Clusters Washed Out

- Use simple L^2 distance to create binary similarity kernel, 10 nearest neighbors
- $\#$ background measurements $= 1137$, $\#$ anomalous measurements $= 23$
  - $D = 49.4 \implies$ energy on anomalous cluster shows up $1:50$ times at best
  - All other eigenvectors (not shown) are only noise

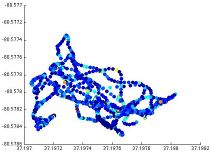
Eigenvector 1



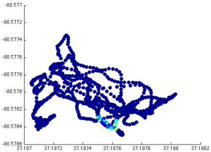
Eigenvector 3



Random Eigenvector



Eigenvector 81



Conclusions

- Takeaways:
  - Small clusters in disjoint graphs appear rarely in LE feature vectors
    - Proportional to number of data points in cluster
  - LE eigenvectors concentrated on small cluster rarely remain localized as graph becomes connected
  - Eigenvectors concentrated on larger cluster almost always remain localized
    - Leads to points on small cluster being forced to zero
  - Phenomenon is supported by simulated and real-world data
- Future Directions:
  - Upper bound on localization theory that doesn't require $A + E$ eigenvector
  - Theory for anomalous clusters that are smaller than k data points
  - Alter selection to subset of "low-frequency" eigenfunctions
    - Subset of indices originally introduced by Jones, Maggioni, Schul 2010

Conclusions

- Takeaways:
  - Small clusters in disjoint graphs appear rarely in LE feature vectors
    - Proportional to number of data points in cluster
  - LE eigenvectors concentrated on small cluster rarely remain localized as graph becomes connected
  - Eigenvectors concentrated on larger cluster almost always remain localized
    - Leads to points on small cluster being forced to zero
  - Phenomenon is supported by simulated and real-world data
- Future Directions:
  - Upper bound on localization theory that doesn't require $A + E$ eigenvector
  - Theory for anomalous clusters that are smaller than k data points
  - Alter selection to subset of "low-frequency" eigenfunctions
    - Subset of indices originally introduced by Jones, Maggioni, Schul 2010

Thank you!

Extra Slides

Eigenvalues of Regular Graph

Theorem (McKay, 1981)

Let $X_1, X_2, \ldots$ be a sequence of random k-regular graphs with adjacency matrices $A_1, A_2, \ldots$.
Let the family $\{X_i\}$ satisfy

1. $n(X_i) \to \infty$
2. $c_k(X_i)/n(X_i) \to 0.$

Then the empirical spectral distribution

$$F_n(x) = |\{i : \lambda_i \left(\frac{1}{\sqrt{k-1}} A_n \right) < x\}|/n$$

converges pointwise to the semicircle law

$$f_{semi}(x) = \frac{1}{2\pi}\sqrt{4 - x^2}, \quad -2 < x < 2.$$

Invariant Subspace Perturbations (cont.)

Theorem (Davis, 1963)

Let $A, E \in \mathbb{C}^{n\times n}$*. Let* (λ, x) *be an eigenpair of* A *such that*

$$\mathrm{sep}(\lambda, \sigma(A) \setminus \lambda) = \min\{|\lambda - \gamma| : \gamma \in \sigma(A) \setminus \lambda\} = \delta.$$

Let

- P *be a spectral projector of* A *such that* $Px = x$
- P' *be the corresponding spectral projector of* $A + E$*, and*
- $\overline{P'}z = z - P'z$.

Then if $\|E\| \leq \epsilon \leq \delta/2$,

$$\|\overline{P'}P\| \leq \frac{\epsilon}{\delta - \epsilon}.$$

Invariant Subspace Perturbations (cont.)

Theorem (Stewart, 1973)

Let $A, E \in \mathbb{C}^{n\times n}$. *Let* $X = [X_1, X_2]$ *be a unitary matrix with* $X_1 \in \mathbb{C}^{n\times l}$, *and suppose* $\mathcal{R}(X_1)$ *is an invariant subspace of* A. *Let*

$$X^*AX = \begin{pmatrix} A_{1,1} & A_{1,2} \\ 0 & A_{2,2} \end{pmatrix}, \quad X^*EX = \begin{pmatrix} E_{1,1} & E_{1,2} \\ E_{2,1} & E_{2,2} \end{pmatrix}.$$

Let $\delta = \operatorname{sep}(A_{1,1}, A_{2,2}) - \|E_{1,1}\| - \|E_{2,2}\|$. *Then if*

$$\frac{\|E_{2,1}\|(\|A_{1,2}\| + \|E_{1,2}\|)}{\delta^2} \leq \frac{1}{4},$$

there is a matrix P *satisfying* $\|P\| \leq 2\frac{\|E_{2,1}\|}{\delta}$ *such that*

$$\widetilde{X_1} = (X_1 + X_2P)(I + P^*P)^{-1//2}$$

is an invariant subspace of $A + E$.