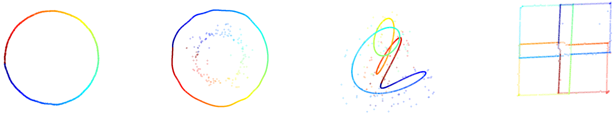
Spectral Frame Analysis and Learning through Graph Structure

Chae A. Clark

Scientific Computing
Norbert Wiener Center
Department of Mathematics
University of Maryland, College Park

April 6, 2016



Thesis Outline

- Spectral Analysis of Scalable Frames
- Generating Frame Scalings through Optimization
- Frames Drawn from Distributions
- Learning Graph Structure
- Learning Linear Structure
- Learning Nonlinear Structure

Outline

Norbert Wiener Center for Harmonic Analysis and Applications

Table of Contents

Norbert Wiener Center for Harmonic Analysis and Applications

Spectral Analysis

Let M be an $n \times m$ matrix with elements on $\mathbb{R}$. Then the Singular Value Decomposition of M is,

$$M = U\Sigma V^T.$$

Let $G = M^T M$ be an $m \times m$ symmetric matrix with elements on $\mathbb{R}$. Then the Eigen-decomposition of G is,

$$G = V\Lambda V^T.$$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$$

Spectral Analysis

Definition (Condition Number)

We define the condition number of an $n \times n$ matrix G to be,

$$\kappa(G) = \frac{\lambda_1}{\lambda_n}.$$

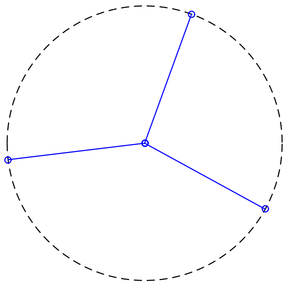
- If the smallest eigenvalue value is 0, we take the condition number to be ∞.
- We shall extend this definition of condition number to apply to non-square matrices using the singular values.
- The condition number of a $n \times m$ matrix M, is defined to be the square-root ratio of the largest and $\min(n, m)$th eigenvalues of $M^T M$.

Frame Definition

A finite frame for $\mathbb{R}^n$ is a set $\Phi = \{\varphi_k\}_{k=1}^m \subset \mathbb{R}^n$ such that there exist positive constants $0 < A \leq B < \infty$ for which

$$A\|f\|_2^2 \leq \sum_{k=1}^m |\langle f, \varphi_k \rangle|^2 \leq B\|f\|_2^2$$

for all $f \in \mathbb{R}^n$.



Norbert Wiener Center for Harmonic Analysis and Applications

Synthesis\Analysis

Synthesis Operator

Given a frame $\Phi \subset \mathbb{R}^n$, we denote, again by Φ, the $n \times m$ matrix whose k^{th} column is the vector φ_k. For a given set of coefficients $\{c_k\}_{k=1}^m$, we can construct/reconstruct a signal f,

$$f = \sum_{k=1}^{m} c_k \varphi_k.$$

Analysis Operator

The adjoint of the frame Φ^T denotes the analysis operator, that allows for the decomposition of signals into frame coefficients,

$$C = \{c_k\}_{k=1}^m = \{\langle \varphi_k, f \rangle\}_{k=1}^m.$$

Norbert Wiener Center for Harmonic Analysis and Applications

Dual Frames, Tight Frames, and the Frame Operator

Dual Frames

Given a frame Φ, we define the dual of Φ to be a frame Ψ such that

$$\Phi\Psi^T = I.$$

Frame Operator

Given a frame Φ, we define the frame operator to be

$$S = \Phi\Phi^T.$$

Tight Frames

A frame Φ is called tight if the frame operator is A times the identity,

$$S = \Phi\Phi^T = AI.$$

Norbert Wiener Center for Harmonic Analysis and Applications

Scalable Frames Definitions

Scalable frames were introduced in [KOPT13, KOP] as a way to create tight frames without changing the structure of the frame itself. More precisely:

Definition

Let $m \geq n$ be given. A frame $\Phi = \{\varphi_k\}_{k=1}^{m} \subset \mathbb{R}^n$ is scalable if there exist a subset $\Phi_J = \{\varphi_k\}_{k \in J}$ with $J \subseteq \{1, 2, \ldots, m\}$, and positive scalars $\{x_k\}_{k \in J}$ such that the system $\widetilde{\Phi}_J = \{x_k \varphi_k\}_{k \in J}$ is a tight frame for $\mathbb{R}^n$.

Characterization

- We can write the analysis operator of the scaled frame as a product of the original frame and a diagonal matrix X,

$$X\Phi^T.$$

- The frame operator then becomes

$$\widetilde{S} = \Phi X^T X \Phi^T = \Phi X^2 \Phi^T = AI.$$

- We can then rescale the coefficient matrix X so that $A = 1$.

Norbert Wiener Center for Harmonic Analysis and Applications

Characterization (cont.)

- One can covert the equation $\Phi X^2 \Phi^T = AI$ into a linear system of equations in m unknowns: x_k^2.
- we need the following function: $F : \mathbb{R}^n \to \mathbb{R}^d$ (called the Reduced Frame Transform)given by,

$$F(\varphi) = [F_0(\varphi), F_1(\varphi), \dots, F_{n-1}(\varphi)]^T,$$

$$F_0(\varphi) = \begin{bmatrix} \varphi_1^2 - \varphi_2^2 \\ \varphi_1^2 - \varphi_3^2 \\ \vdots \\ \varphi_1^2 - \varphi_n^2 \end{bmatrix}, F_k(\varphi) = \begin{bmatrix} \varphi_k \varphi_{k+1} \\ \varphi_k \varphi_{k+2} \\ \vdots \\ \varphi_k \varphi_n \end{bmatrix}$$

and $F_0(\varphi) \in \mathbb{R}^{n-1}$, $F_k(\varphi) \in \mathbb{R}^{n-k}$, $k = 1, 2, \dots, n-1$, where $d := \frac{(n-1)(n+2)}{2}$.

- Let $F(\Phi)$ be the $d \times m$ matrix given by

$$F(\Phi) = (F(\varphi_1) \ F(\varphi_2) \ \dots \ F(\varphi_m)).$$

Norbert Wiener Center for Harmonic Analysis and Applications

Previous Result

Proposition

[KOP, Proposition 3.7] Given a frame $\Phi \subset \mathbb{R}^n$. Φ is scalable if and only if there exists a non-negative $u \in \ker F(\Phi) \backslash \{0\}$. Moreover, the scaling matrix X is a diagonal operator where the elements are the square-roots of the solution u.

Norbert Wiener Center for Harmonic Analysis and Applications

Convex Geometry

- First consider the sets $\mathcal{S}_1$ and $\mathcal{S}_2$ given by

$$\mathcal{S}_1 := \{u \in \mathbb{R}^m \mid F(\Phi)u = 0\,,\, u \geq 0\,,\, u \neq 0\},$$

  and

$$\mathcal{S}_2 := \{v \in \mathbb{R}^m \mid F(\Phi)v = 0\,,\, v \geq 0\,,\, \|v\|_1 = 1\}.$$

- $\mathcal{S}_1$ is a subset of the null space of $F(\Phi)$, and each $u \in \mathcal{S}_1$ is associated a scaling matrix X_u, defined as

$$X_u := (X_{ij})_u = \begin{cases} \sqrt{u_i} & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

- $\mathcal{S}_2 \subset \mathcal{S}_1 \cap B_{\ell^1}$, where B_{ℓ^1} is the unit ball under the ℓ^1 norm.

Norbert Wiener Center for Harmonic Analysis and Applications

Generating Frame Scalings through Optimization

Theorem

Let $\Phi = \{\varphi_k\}_{k=1}^m \subset \mathbb{R}^n$ *be a frame, and let* $g : \mathbb{R}^m \to \mathbb{R}$ *be a convex objective function. Then the program*

$$\begin{aligned} \textit{minimize: } & g(u) \\ \textit{subject to: } & F(\Phi)u = 0 \\ & \|u\|_1 = 1 \\ & u \geq 0 \end{aligned}$$

has a solution if and only if the frame Φ *is scalable.*

Norbert Wiener Center for Harmonic Analysis and Applications

Proof Sketch

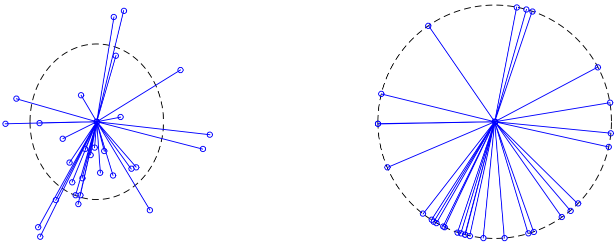
Proof.

Any feasible solution u^* is contained in the set $\mathcal{S}_2$, which itself is contained in $\mathcal{S}_1$, and thus corresponds to a scaling matrix X_u.

Conversely, any $u \in \mathcal{S}_1$ can be mapped to a $v \in \mathcal{S}_2$ by appropriate scaling factor. This provides an initial feasible solution, and so there must exist a minimizer on $\mathcal{S}_2$. □

Norbert Wiener Center for Harmonic Analysis and Applications

Gaussian Frames



Norbert Wiener Center for Harmonic Analysis and Applications

Sparse Results

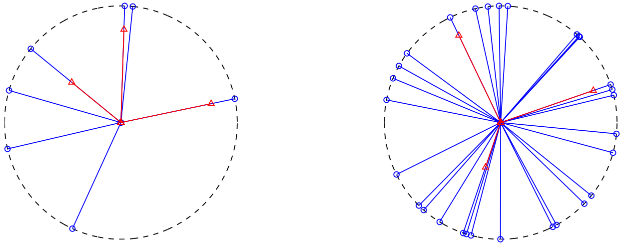


Figure: Results on two random Gaussian frames before and after scaling.

Barrier Results

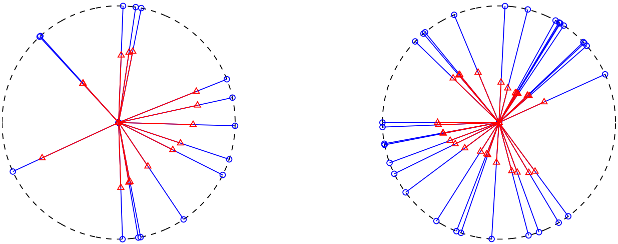


Figure: Results on two random Gaussian frames before and after scaling.

Table of Contents

Spectral Analysis of Frames

- We can relate the scaling weights to the spectrum of the frame
- We can also view the scaling weights as analogs of the spectrum

Theorem (Spectral Frame Decomposition)

Let Φ be a frame in $\mathbb{R}^n$ with m elements, and assume Φ is scalable with diagonal scaling matrix X. Furthermore, let V be an $m \times m$ matrix of the right singular vectors of Φ, such that the singular value decomposition is,

$$\Phi = U\Sigma V^T.$$

Then there exists an $m \times n$ sub-block of V (denoted $\widetilde{V}$) such that

$$\widetilde{V}^T X^2 \widetilde{V} = \Lambda^{-1}.$$

Norbert Wiener Center for Harmonic Analysis and Applications

Proof Sketch

$$\Phi X^2 \Phi^T = I.$$

Using a singular value decomposition of Φ, we have

$$U\Sigma V^T X^2 V \Sigma^T U^T = I.$$

We can simplify this system by performing left and right matrix multiplications of U^T and U respectively.

$$\Sigma V^T X^2 V \Sigma^T = I.$$

$$\Sigma^T \Sigma V^T X^2 V \Sigma^T \Sigma = \Sigma^T I \Sigma,$$

$$\begin{bmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} V^T X^2 V \begin{bmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

$$\Lambda \widetilde{V}^T X^2 \widetilde{V} \Lambda = \Lambda,$$

$$\widetilde{V}^T X^2 \widetilde{V} = \Lambda^{-1} \Lambda \Lambda^{-1},$$

$$\widetilde{V}^T X^2 \widetilde{V} = \Lambda^{-1}.$$

Spectral Analysis of Frames

Corollary (Spectral Frame Decomposition)

Let Φ be a frame in $\mathbb{R}^n$ with m elements, and assume Φ is scalable with diagonal scaling matrix X. Furthermore, let V be an $m \times m$ matrix of the right singular vectors of Φ, such that the singular value decomposition is

$$\Phi = U \Sigma V^T.$$

Then the inverse of each eigenvalue of the frame operator $S = \Phi \Phi^T$ can be written as the sum of squares of the right singular vectors $(v_i)_k = v_{ik}$ and the scaling weights $X_{kk} = x_k$,

$$\frac{1}{\lambda_i} = \langle v_i \odot v_i, x \odot x \rangle = \sum_{k=1}^{m} (v_{ik} x_k)^2 \quad \text{for } i = 1, \ldots, n.$$

Norbert Wiener Center for Harmonic Analysis and Applications

Perturbed Spectral Analysis of Frames

- It will often occur that a frame will not be exactly scalable
- We can bound approximately scalable frames
- We present worst-case bounds for non-exact scalings

Norbert Wiener Center for Harmonic Analysis and Applications

Perturbed Spectral Analysis of Frames

Theorem (Perturbed Spectral Decomposition)

Let Φ be a frame in $\mathbb{R}^n$ with m elements. Also, let $\widetilde{V}$ denote an $m \times n$ sub-block of V. Given a non-trivial, non-negative diagonal matrix Y, we shall write the general scalability equality as

$$\Phi Y^2 \Phi^T = I + E,$$

with an error matrix, E, bounded by

$$E \preceq \delta \mathbb{1}\mathbb{1}^T,$$

for some $\delta > 0$. Then the following inequality holds,

$$\left\| \widetilde{V}^T Y^2 \widetilde{V} \right\|_2 \leq \frac{1 + \delta n}{\lambda_n}.$$

Proof Sketch

$$
\begin{aligned}
\Phi Y^2 \Phi^T &= I + E, \\
(U\Sigma V^T) Y^2 (U \Sigma V^T)^T &= I + E, \\
\widetilde{V}^T Y^2 \widetilde{V} &= \Lambda^{-1} \Lambda \Lambda^{-1} + \Lambda^{-1/2} U^T E U \Lambda^{-1/2}, \\
\widetilde{V}^T Y^2 \widetilde{V} &= \Lambda^{-1} + \Lambda^{-1/2} U^T E U \Lambda^{-1/2}.
\end{aligned}
$$

$$
\begin{aligned}
\|\widetilde{V}^T Y^2 \widetilde{V}\|_2 &= \|\Lambda^{-1} + \Lambda^{-1/2} U^T E U \Lambda^{-1/2}\|_2, \\
\|\widetilde{V}^T Y^2 \widetilde{V}\|_2 &\le \frac{1}{\lambda_n} + \frac{\delta n}{\lambda_n}.
\end{aligned}
$$

Approximate Spectral Analysis of Frames

Corollary (Approximate Spectral Decomposition)

Let Φ be a frame in $\mathbb{R}^n$ with m elements. Also, let $\widetilde{V}$ denote an $m \times n$ sub-block of V. Given a non-trivial, non-negative diagonal matrix Y, we shall write the general scalability equality as

$$\Phi Y^2 \Phi^T = I + E,$$

with an error matrix, E, bounded by

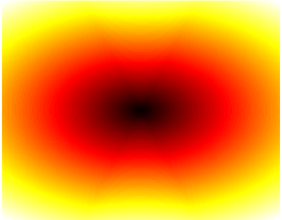
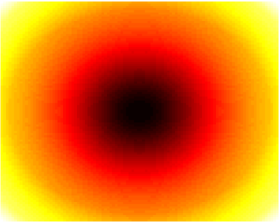
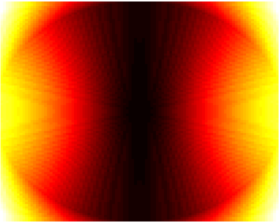
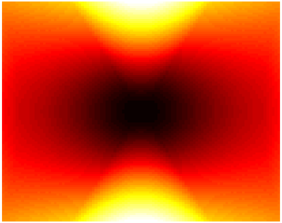
$$E \preceq \delta \mathbb{1}\mathbb{1}^T,$$

for some $\delta > 0$. If Φ is scalable with scaling matrix X, and the difference between X and Y is denoted $D^2 := X^2 - Y^2$, then the following inequality holds,

$$\left\| \widetilde{V}^T D^2 \widetilde{V} \right\|_2 \leq \frac{\delta n}{\lambda_n}.$$

Norbert Wiener Center for Harmonic Analysis and Applications

Scalability Projections

Scalability Projections

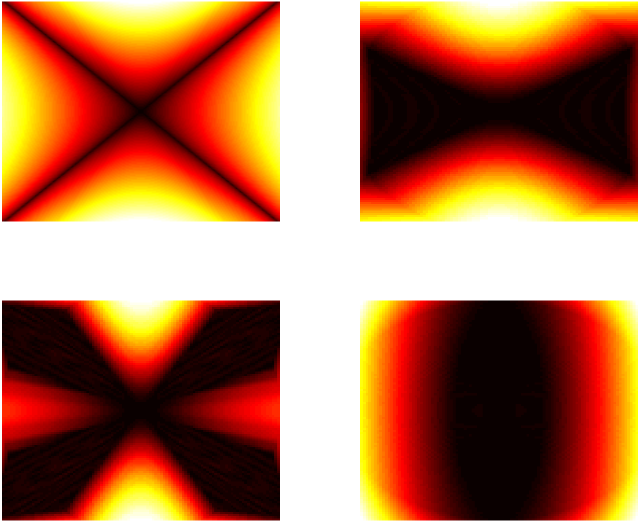


Table of Contents

Graph Background

- Denote a graph by $\mathcal{G} := (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} := \{\nu_1, \nu_2, \ldots, \nu_n\}$ is a set of vertices on the graph.
- $\mathcal{E} := \{e_1, e_2, \ldots, e_m\}$ is the ordered set of edge pairs that denotes a connection between two nodes.
- A weight $0 \leq \omega_{ij} \leq 1$ denotes the similarity between two nodes (ν_i, ν_j), and $\omega_{ij} = 0$ if the nodes are not connected.
- The matrix of these weights is referred to as the adjacency matrix W.
- We denote the degree of a node, ν_i, as $d_i := \sum_{j=1}^{n} \omega_{ij}$.
- The degree matrix D is then a diagonal matrix with entries $D_{ii} = d_i$ for $i = 1, \ldots, n$.

Norbert Wiener Center for Harmonic Analysis and Applications

Graph Background

- We now define the Lapalcian matrix on the graph as

$$L_{\mathcal{G}} := L = D - W. \tag{1}$$

- The incidence matrix, $B = [b_1, b_2, \ldots, b_m]$, is defined as an $n \times m$ matrix where every column in B represents an edge (ν_i, ν_j) in $\mathcal{E}$.
- For a column in the incidence matrix, b_k, we have

$$b_k(i) := \begin{cases} \sqrt{\omega_{ij}} & : (\nu_i, \nu_j) \in \mathcal{E}\,,\ i < j \\ -\sqrt{\omega_{ij}} & : (\nu_i, \nu_j) \in \mathcal{E}\,,\ i > j \\ 0 & : \text{else} \end{cases}. \tag{2}$$

- The Laplacian can now be defined as $L := BB^T$.

Norbert Wiener Center for Harmonic Analysis and Applications

Graph Conditioning

Definition (Graph Condition Number)

$$\kappa(L_{\mathcal{G}}) := \frac{\lambda_1}{\lambda_r} \text{ for } \lambda_1 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_n = 0. \qquad (3)$$

- This function is simply the condition number of L (ignoring the zero eigenvalues).
- Where $\lambda_i = 0$ can lead to numerically unstable solutions for linear systems, $\lambda_i = 0$ in this setting, disconnects the graph.
- Scaling leads to the well-conditioned graphs, as sets of complete sub-graphs.
- This encourages the use of the incidence matrix B as the frame Φ, which leads to the Laplacian L as the frame operator S.

Graph Conditioning

Definition

Let $\mathcal{G}(\mathcal{V}, \mathcal{E}, \omega)$ be a graph with incidence matrix B, and Laplacian matrix $L = BB^T$. Then B is scalable if there exists a non-negative, non-zero diagonal matrix X, such that the graph condition number $\kappa(\widetilde{\mathcal{L}})$ of the scaled Laplacian $\widetilde{L} = BX^2B^T$ is equal to 1,

$$\widetilde{\lambda}_1 = \cdots = \widetilde{\lambda}_r > \widetilde{\lambda}_{r+1} = \ldots \lambda_n = 0. \tag{4}$$

Proposition

Let $\mathcal{G}(\mathcal{V}, \mathcal{E}, \omega)$ be a complete graph with incidence matrix B, and Laplacian matrix $L = BB^T$. Then B is scalable with scaling weights $x_k = \dfrac{1}{\sqrt{w_{ij}}}$, such that,

$$\widetilde{L}_{\mathcal{G}} = BX^2B^T = nI - \mathbb{1}\mathbb{1}^T,$$

and $\kappa(\widetilde{L}_{\mathcal{G}}) = 1$.

Norbert Wiener Center for Harmonic Analysis and Applications

Proof Sketch

- The graph is complete, and if we scale all of the edges to have weight 1, the degree of each node will be equal to $n-1$, and the resulting graph will be complete.
- As the complete graph has Laplacian eigenvalues $\lambda_i = n$ for $i = 1, \ldots, n-1$, the graph has condition number 1.

Problem Formulation

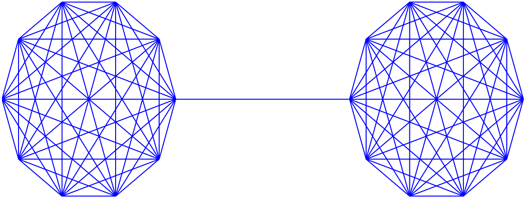
- As every complete graph is trivially scalable, we move on to more complicated graphs
- We can cast the problem of scaling a graph as finding a non-negative solution to the following:

$$\begin{aligned} F(B)u &= [\mathbf{0}, -\mathbb{1}]^T, \\ \mathbb{1}^T u &\geq 1, \\ u &\geq \mathbf{0}, \end{aligned}$$

$$\begin{aligned} \text{minimize:} \quad & g(u) \\ \text{subject to:} \quad & F_0(B)u = \mathbf{0}, \\ & \mathbb{1}^T u \geq 1, \\ & u \geq \mathbf{0}. \end{aligned}$$

Norbert Wiener Center for Harmonic Analysis and Applications

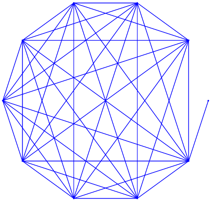
Graph Examples I

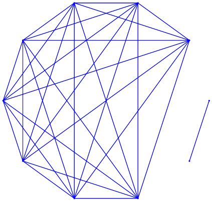


Two-Complete Graphs Spectrum				
	original	g1	g2	g3
$\sigma_{\max}$	3.4396	3.4396	3.3977	3.1623
$\sigma_{\min}$	0.4112	0.4112	0.4089	3.1623
$\kappa(L_{\mathcal{G}})$	8.3648	8.3648	8.3094	1.0000
x_k	-	0.0101	0.0112	0.0090

Norbert Wiener Center for Harmonic Analysis and Applications

Graph Examples II





Outlier Complete Graph Spectrum				
	original	g1	g2	g3
$\sigma_{\max}$	3.1623	3.0000	3.1623	3.0927
$\sigma_{\min}$	1.0000	3.0000	1.0000	0.9909
$\kappa(L_{\mathcal{G}})$	10.000	1.0000	10.000	9.7412
x_k	-	0.2000	0.0286	0.0000

Norbert Wiener Center for Harmonic Analysis and Applications

Table of Contents

Low-Rank Embeddings and Pre-Image Problems

- Principal Component Analysis (PCA) is a standard tool for data analysis and low-rank approximations [Jol02, LV07].
- Viewing the eigenvalues as variance indicators is a clear and concise explanation of the projection, and when data's true manifold is linear/affine, PCA is optimal in its representation.
- For notation, and for connections with later notions, we shall denote PCA as an eigen-value/vector problem of a data matrix M,

$$M^T M V = V \Lambda,$$

where the eigen-decomposition of the gram matrix $M^T M$ is,

$$M^T M = V \Lambda V^T.$$

Low-Rank Embeddings and Pre-Image Problems

- The assumption of linearity on the manifold is generally violated for complex datasets.
- Nonlinear dimension reduction techniques were designed to alleviate this drawback.
- The canonical example being Kernel PCA [SSM97, LV07].
- Instead of analyzing the data directly, kernel methods analyze the relationship between data points.

Low-Rank Embeddings and Pre-Image Problems

- In [HLMS04, BDLR+04], various non-linear dimension reduction methods (Isomap [TDSL00], Laplacian Eigenmaps [BN03], Locally Linear Embeddings [SR00], etc.) are shown to fall under the Kernel PCA model.
- The Kernel PCA problem for a dataset M, with respect to a kernel $K(M) = K_M$, shall be denoted,

$$K_M V = V\Lambda.$$

- The embedding Θ of this dataset, shall be denoted,

$$\Theta = \Lambda^{\frac{1}{2}} V^T,$$

Robust Principal Component Analysis [RPCA]

- We have at our disposal, sparsity methods, and spectral methods.
- Both are useful signal processing techniques when dealing with large datasets.
- Robust Principal Component Analysis was devised as a technique to take advantage of sparsity and low intrinsic dimensionality of datasets.

Robust Principal Component Analysis [RPCA]

- Standard practice, when dealing with search over such a large space, is to formulate an optimization problem
- Given a data matrix $\widetilde{\Phi}$, we want a sparsity E and low-rank Φ decomposition.

$$\begin{aligned} \text{minimize: } & \text{rank}(\Phi) + \gamma \|E\|_0 \\ \text{subject to: } & \Phi + E = \widetilde{\Phi} \end{aligned}$$

ℓ_1 Norm Approximation

- Let's first consider minimizing $\|E\|_0$.
- This problem is NP-hard, so the standard approach is to find a convex relaxation that approximately solves the problem.
- The well known convex relaxation is the ℓ_1 norm.

Norbert Wiener Center for Harmonic Analysis and Applications

Nuclear Norm Approximation

- Now consider minimizing rank(Φ).
- We first notice that minimizing the rank of a matrix is also NP-hard.
- We need a convex relaxation of the rank function.

Nuclear Norm Approximation

- Using an analogous approximation from the ℓ_0-ℓ_1 derivation, the Nuclear Norm becomes the convex relaxation

$$\|\Phi\|_* := \sigma_1 + \sigma_2 + \sigma_3 + \cdots + \sigma_n.$$

The Complete Formulation

Original Problem

$$\begin{aligned} \text{minimize: } & \operatorname{rank}(\Phi) + \gamma \|E\|_0 \\ \text{subject to: } & \Phi + E = \widetilde{\Phi} \end{aligned}$$

Norbert Wiener Center for Harmonic Analysis and Applications

The Complete Formulation

Original Problem

$$\text{minimize: } \operatorname{rank}(\Phi) + \gamma \|E\|_0$$
$$\text{subject to: } \Phi + E = \widetilde{\Phi}$$

Convex Relaxation

$$\text{minimize: } \|\Phi\|_* + \gamma \|E\|_1$$
$$\text{subject to: } \Phi + E = \widetilde{\Phi}$$

Norbert Wiener Center for Harmonic Analysis and Applications

Robust Manifold Learning

- The natural extension of PCA is to compare similarities between nonlinear transformations of the dataset in the form of kernels (KPCA).
- In this same vein, we may wish to add a notion of robustness to KPCA by employing an error regularizing term.
- This motivates the introduction of the Robust Manifold Learning (RML) problem,

$$\begin{aligned} \text{minimize: } & \text{rank}(K(\Phi)) + \gamma \|E\|_0 \\ \text{subject to: } & \Phi + E = \widetilde{\Phi}, \end{aligned}$$

Robust Manifold Learning

$$\begin{aligned} \text{minimize: } & \operatorname{rank}(K(\Phi)) + \gamma\|E\|_0 \\ \text{subject to: } & \Phi + E = \widetilde{\Phi}, \end{aligned}$$

- As with many formulations, we shall study the convex relaxation of this problem,

$$\begin{aligned} \text{minimize: } & \|K(\Phi)\|_* + \gamma\|E\|_1 \\ \text{subject to: } & \Phi + E = \widetilde{\Phi}. \end{aligned}$$

- Much of the intuition behind this approach can be gleamed from an understanding of robust PCA and its variations.

Inverse Mapping

- This brings us to the major issue with nonlinear methods; there is in general no well-defined inverse for an embedding obtained from KPCA.
- The kernel matrix K is computed, and an embedding is formed and thresholded for the kernel.
- Once the threshold has been applied, an inverse operation is performed as follows,

$$\varphi_k = \sum_{i \in \Omega(\theta_k)} \frac{a}{\|\theta_k - \theta_i\|_2^2} \varphi_i. \tag{5}$$

Norbert Wiener Center for Harmonic Analysis and Applications

RML Algorithm

Before presenting the algorithm, we define following operators.

Definition (Spatial Shrinkage Operator)

We denote by $\mathcal{S}_\gamma$ the Spatial Shrinkage Operator, which performs a soft thresholding on a given $n \times m$ matrix by subtracting positive constant, γ, from each element and thresholding all negative values to 0:

$$\mathcal{S}_\gamma[A] = \max\{A - \gamma \mathbb{1}\mathbb{1}^T \,, \mathbf{0}\},$$

where we use the entry-wise $\max$ function.

Definition (Spectral Shrinkage Operator)

We denote by $\widehat{\mathcal{S}}_\mu$ the Spectral Shrinkage Operator, which performs a soft thresholding on a given $n \times m$ matrix by subtracting positive constant, μ, from each singular value and thresholding all negative values to 0:

$$\widehat{\mathcal{S}}_\mu[A] = U \cdot \max\{\Sigma - \mu I \,, \mathbf{0}\} \cdot V^T,$$

where $A = U\Sigma V^T$.

Norbert Wiener Center for Harmonic Analysis and Applications

RML Algorithm

Definition (Embedding Operator)

Define $\mathcal{E}$ as the embedding operator such that,

$$\Theta = \mathcal{E}[K(\Phi)],$$

where $K(\Phi)$ is the kernel matrix on the dataset Φ.

Definition (Inverse Operator)

Define $\mathcal{E}^{-1}$ as the embedding operator such that,

$$\Phi = \mathcal{E}^{-1}[\Theta],$$

where Θ is an embedding formed from a kernel matrix. The inverse is performed using the interpolation formula presented previously.

Norbert Wiener Center for Harmonic Analysis and Applications

RML Algorithm

while not converged do:

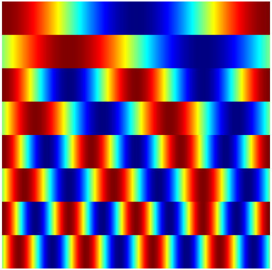
1. $K \leftarrow K(\widetilde{\Phi} - E)$
2. $K \leftarrow \widehat{\mathcal{S}}_\mu[K]$
3. $\Theta \leftarrow \mathcal{E}[K]$
4. $\Phi \leftarrow \mathcal{E}^{-1}[\Theta]$
5. $E \leftarrow \mathcal{S}_\gamma[\widetilde{\Phi} - \Phi]$

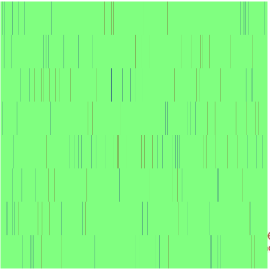
end while

Norbert Wiener Center for Harmonic Analysis and Applications

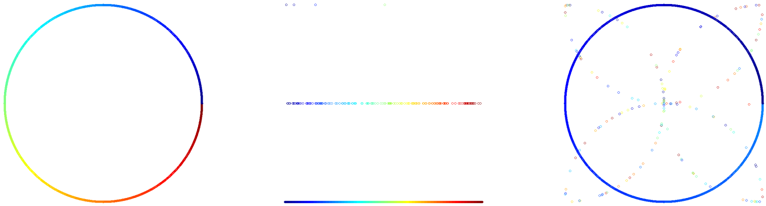
Circle Embeddings

- We start with a clear example of when standard RPCA fails by adding sparse noise to an embedding of a circle.
- We sample sine and cosine functions $m = 1000$ times using $n = 4$ frequencies for each, resulting in a dataset Φ of size $2n$-by-m.
- Sparse noise E is then added to the dataset by randomly selecting 80 indices and biasing the location.

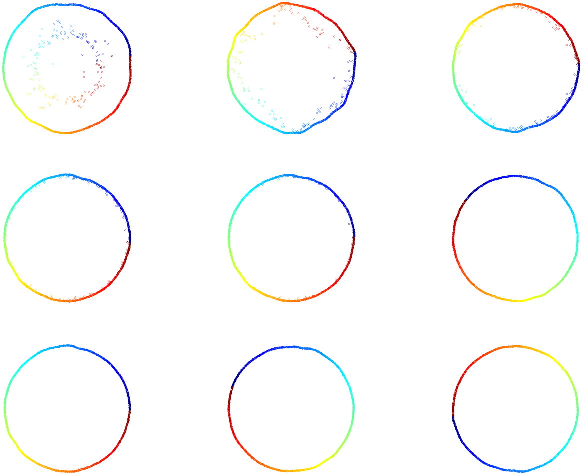




Circle Embeddings



Circle Embeddings



Circle Embeddings

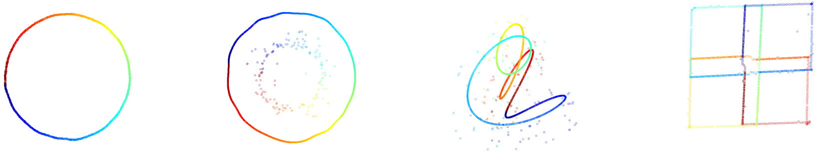
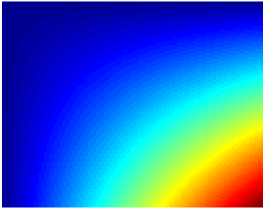


Figure: This figure shows the embeddings obtained after the various techniques are employed. From left to right, we present the final results after our robust manifold learning technique, kernel PCA, standard PCA, and standard robust PCA.

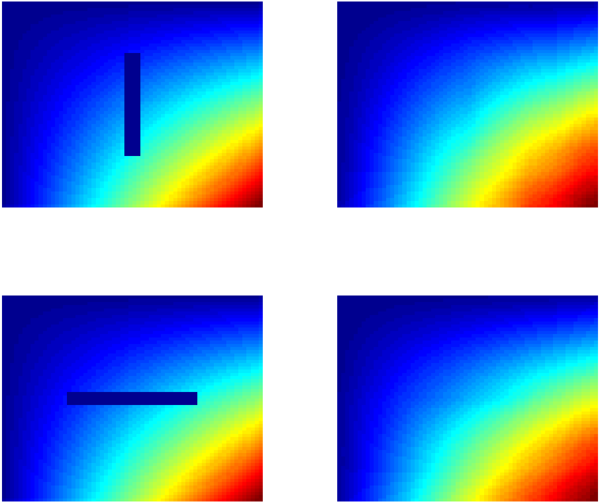
Inpainting Background

- Image inpainting interpolates across corrupted of missing data in an image
- Sapiro and Bertalmio 2000
- Igehy and Pereira 1997



Norbert Wiener Center for Harmonic Analysis and Applications

Inpainting



Inpainting

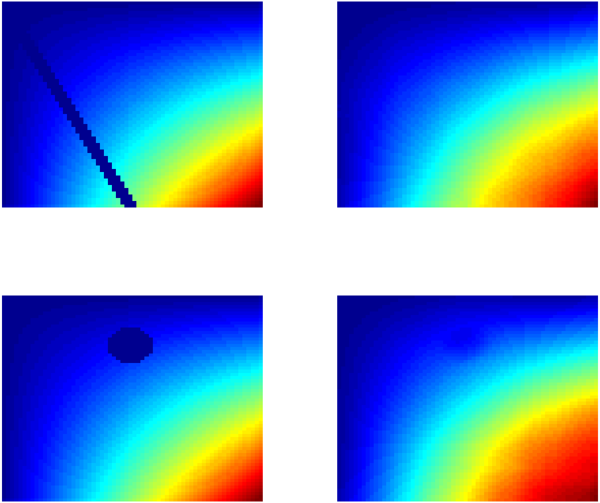


Table of Contents

References I

- Y. Bengio, O. Delalleau, N. Le Roux, J. Paiement, P. Vincent, and M. Ouimet, *Learning eigenfunctions links spectral embedding and kernel pca*, Neural Computation **16** (2004), no. 10, 2197–2219.
- M. Belkin and P. Niyogi, *Laplacian eigenmaps for dimensionality reduction and data representation*, Neural computation **15** (2003), no. 6, 1373–1396.
- J. Ham, D. Lee, S. Mika, and B. Schölkopf, *A kernel view of the dimensionality reduction of manifolds*, Proceedings of the twenty-first international conference on Machine learning, ACM, 2004, p. 47.
- I. Jolliffe, *Principal component analysis*, Wiley Online Library, 2002.
- G. Kutyniok, K. Okoudjou, and F. Philipp, *Scalable Frames and Convex Geometry*, Contemp. Math. 626 (2014), 19-32.
- G. Kutyniok, K. Okoudjou, F. Philipp, and E. Tuley, *Scalable frames*, Linear Algebra and its Applications **438** (2013), no. 5, 2225–2238.

Norbert Wiener Center for Harmonic Analysis and Applications

References II

- J. Lee and M. Verleysen, *Nonlinear dimensionality reduction*, Springer Science & Business Media, 2007.
- L. Saul and S. Roweis, *An introduction to locally linear embedding*, Tech. report, 2000.
- B. Schölkopf, A. Smola, and K.R. Müller, *Kernel principal component analysis*, Artificial Neural NetworksICANN'97, Springer, 1997, pp. 583–588.
- J. Tenenbaum, V. De Silva, and J. Langford, *A global geometric framework for nonlinear dimensionality reduction*, science **290** (2000), no. 5500, 2319–2323.

Norbert Wiener Center for Harmonic Analysis and Applications

Extra Slides: Error Analysis

$$
\begin{aligned}
\widetilde{G}\widetilde{f} - Gf &= 0 \\
(G+E)\widetilde{f} - Gf &= 0 \\
G(\widetilde{f} - f) &= -E\widetilde{f} \\
\widetilde{f} - f &= -G^{-1}E\widetilde{f} \\
\|\widetilde{f} - f\|_2 &= \|G^{-1}E\widetilde{f}\|_2 \\
\|\widetilde{f} - f\|_2 &\leq \|G^{-1}\|_2\|E\|_2\|\widetilde{f}\|_2 \\
\frac{\|\widetilde{f} - f\|_2}{\|\widetilde{f}\|_2} &\leq \frac{\|G^{-1}\|_2\|E\|_2\|G\|_2}{\|G\|_2} \\
\frac{\|\widetilde{f} - f\|_2}{\|\widetilde{f}\|_2} &\leq (\|G^{-1}\|_2\|G\|_2)\frac{\|E\|_2}{\|G\|_2}.
\end{aligned}
$$

Norbert Wiener Center for Harmonic Analysis and Applications

Extra Slides: Error Analysis (cont.)

- The relative error in the approximate solution is bounded by the error matrix E, but also properties of the matrix G.
- The matrix norms of G and G^{-1} are the largest and reciprocal smallest eigenvalues respectively,

$$\|G\|_2 = \lambda_1 \quad , \quad \|G^{-1}\|_2 = \frac{1}{\lambda_n}.$$

Norbert Wiener Center for Harmonic Analysis and Applications

Extra Slides: Spectral Frame Decomposition

$$\Phi X^2 \Phi^T = I.$$

Using a singular value decomposition of Φ, we have

$$(U\Sigma V^T)X^2(U\Sigma V^T)^T = I,$$
$$U\Sigma V^T X^2 V \Sigma^T U^T = I.$$

We can simplify this system by performing left and right matrix multiplications of U^T and U respectively.

$$U^T U\Sigma V^T X^2 V \Sigma^T U^T U = U^T I U,$$
$$\Sigma V^T X^2 V \Sigma^T = I.$$

We shall now perform left and right matrix multiplications by Σ^T and Σ respectively. In this case, we obtain block matrices where the upper-left $n \times n$ block is a diagonal matrix of non-zero eigenvalues Λ and all other blocks are zero matrices.

Extra Slides: Spectral Frame Decomposition (cont.)

$$\Sigma^T \Sigma V^T X^2 V \Sigma^T \Sigma = \Sigma^T I \Sigma,$$
$$\begin{bmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} V^T X^2 V \begin{bmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

We write V in block form as well,

$$\begin{bmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} V_1 & V_2 \\ V_3 & V_4 \end{bmatrix}^T \begin{bmatrix} X_1 & \mathbf{0} \\ \mathbf{0} & X_2 \end{bmatrix}^2 \begin{bmatrix} V_1 & V_2 \\ V_3 & V_4 \end{bmatrix} \begin{bmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Given this structure, we shall further simplify the system by removing the zero matrices and obtain the result,

$$\begin{aligned} \Lambda \widetilde{V}^T X^2 \widetilde{V} \Lambda &= \Lambda, \\ \widetilde{V}^T X^2 \widetilde{V} &= \Lambda^{-1} \Lambda \Lambda^{-1}, \\ \widetilde{V}^T X^2 \widetilde{V} &= \Lambda^{-1}. \end{aligned}$$

Extra Slides: Perturbed Spectral Decomposition

$$
\begin{aligned}
\Phi Y^2 \Phi^T &= I + E, \\
(U\Sigma V^T) Y^2 (U\Sigma V^T)^T &= I + E, \\
U\Sigma V^T Y^2 V \Sigma^T U^T &= I + E, \\
U^T U \Sigma V^T Y^2 V \Sigma^T U^T U &= U^T I U + U^T E U, \\
I \Sigma V^T Y^2 V \Sigma^T I &= U^T U + U^T E U, \\
\Sigma V^T Y^2 V \Sigma^T &= I + U^T E U, \\
\Sigma^T \Sigma V^T Y^2 V \Sigma^T \Sigma &= \Sigma^T I \Sigma + \Sigma^T U^T E U \Sigma, \\
\Lambda \widetilde{V}^T Y^2 \widetilde{V} \Lambda &= \Lambda + \Lambda^{1/2} U^T E U \Lambda^{1/2}, \\
\widetilde{V}^T Y^2 \widetilde{V} &= \Lambda^{-1} \Lambda \Lambda^{-1} + \Lambda^{-1/2} U^T E U \Lambda^{-1/2}, \\
\widetilde{V}^T Y^2 \widetilde{V} &= \Lambda^{-1} + \Lambda^{-1/2} U^T E U \Lambda^{-1/2}.
\end{aligned}
$$

Extra Slides: Perturbed Spectral Decomposition (cont.)

Taking the norm of both sides of the equation, and applying the bound, we have on the error matrix E,

$$\|\widetilde{V}^T Y^2 \widetilde{V}\|_2 = \|\Lambda^{-1} + \Lambda^{-1/2} U^T E U \Lambda^{-1/2}\|_2,$$
$$\|\widetilde{V}^T Y^2 \widetilde{V}\|_2 \leq \|\Lambda^{-1}\|_2 + \|\Lambda^{-1/2} U^T E U \Lambda^{-1/2}\|_2,$$
$$\|\widetilde{V}^T Y^2 \widetilde{V}\|_2 \leq \|\Lambda^{-1}\|_2 + \|\Lambda^{-1/2} U^T (\delta \mathbb{1}\mathbb{1}^T) U \Lambda^{-1/2}\|_2,$$
$$\|\widetilde{V}^T Y^2 \widetilde{V}\|_2 \leq \frac{1}{\lambda_n} + \frac{\delta}{\lambda_n}\|U^T (\mathbb{1}\mathbb{1}^T) U\|_2,$$
$$\|\widetilde{V}^T Y^2 \widetilde{V}\|_2 \leq \frac{1}{\lambda_n} + \frac{\delta}{\lambda_n}\|\mathbb{1}\mathbb{1}^T\|_2,$$
$$\|\widetilde{V}^T Y^2 \widetilde{V}\|_2 \leq \frac{1}{\lambda_n} + \frac{\delta n}{\lambda_n}.$$

Norbert Wiener Center for Harmonic Analysis and Applications

Examples