Harmonic Analysis and Big Data: Data Dependent Representations

Wojciech Czaja

Institute for Mathematics and Its Applications 2015 PI Summer Graduate Program
College Park, August 5–7, 2015
Yesterday we have considered transformations which have an a priori structure which is independent from the given input data.

These types of transformations have some advantages in speed, or in provable accuracy for certain classes of data.

Drawbacks arise when data complexity increases over time and is no longer covered by our theoretical guarantees.

A possible solution is to develop methods of data analysis which are data-dependent.
There are many techniques for Data Organization and Manifold Learning, e.g., Principal Component Analysis (PCA), Locally Linear Embedding (LLE), Isomap, genetic algorithms, and neural networks.

We are interested in a subfamily of these techniques known as *Kernel Eigenmap Methods*. These include Kernel PCA, LLE, Hessian LLE (HLLE), and Laplacian Eigenmaps, Diffusion Maps, Diffusion Wavelets (all sometimes known as Output Normalized Methods).

Kernel eigenmap methods require two steps. Given data space $X$ of $N$ vectors in $\mathbb{R}^D$.

1. Construction of an $N \times N$ symmetric, positive semi-definite kernel, $K$, from these $N$ data points in $\mathbb{R}^D$.
2. Diagonalization of $K$, and then choosing $d \leq D$ significant eigenmaps of $K$. These become our new coordinates, and accomplish dimensionality reduction.

We are particularly interested in diffusion kernels $K$, which are defined by means of transition matrices.
Kernel eigenmap methods were introduced to address complexities not resolvable by linear methods.

The idea behind *kernel methods* is to express correlations or similarities between vectors in the data space $X$ in terms of a symmetric, positive semi-definite kernel function $K : X \times X \rightarrow \mathbb{R}$. Generally, there exists a Hilbert space $\mathcal{K}$ and a mapping $\Phi : X \rightarrow \mathcal{K}$ such that

$$K(x, y) = \langle \Phi(x), \Phi(y) \rangle.$$

Then, diagonalize by the spectral theorem and choose significant eigenmaps to obtain dimensionality reduction.

Kernels can be constructed by many kernel eigenmap methods. These include Kernel PCA, LLE, HLLE, and Laplacian Eigenmaps.
The second step in kernel eigenmap methods is the diagonalization of the kernel.

Let $e_j$, $j = 1, \ldots, N$, be the set of eigenvectors of the kernel matrix $K$, with eigenvalues $\lambda_j$.

Order the eigenvalues monotonically.

Choose the top $d << D$ significant eigenvectors to map the original data points $x_i \in \mathbb{R}^D$ to $(e_1(i), \ldots, e_d(i)) \in \mathbb{R}^d$, $i = 1, \ldots, N$. 
There are other alternative interpretations for the steps of our diagram:

1. Constructions of kernels $K$ may be independent from data and based on principles.

2. Redundant representations, such as frames, can be used to replace orthonormal eigendecompositions.

We need not select the target dimensionality to be lower than the dimension of the input. This leads to data expansion, or data organization, rather than dimensionality reduction.
Presented approach leads to analysis of operators on data-dependent structures, such as graphs or manifolds.

Locally Linear Embedding, Diffusion Maps, Diffusion Wavelets, Laplacian Eigenmaps, Schroedinger Eigenmaps.

Mathematical core:

- Pick a positive symmetric operator $A$ as the infinitesimal generator of a semigroup of operators, $e^{tA}$, $t > 0$.
- The semigroup can be identified with the Markov processes of diffusion or random walks, as is the case, e.g., with Diffusion Maps and Diffusion Wavelets of R. R. Coifman, S. Lafon, and M. Maggioni.
- The infinitesimal generator and the semigroup share the common representation, e.g., eigenbasis.

A. Pazy, Semi-groups of Linear Operators And Applications to Partial Differential Equations, Lecture Notes #10, University of Maryland College Park, January 25, 1974

A. Pazy, Semi-groups of Linear Operators And Applications to Partial Differential Equations, in Applied Mathematical Sciences (1983), Springer

S. Lafon, Diffusion maps and geometric harmonics, Ph.D. Thesis, Yale University, 2004
Principal Component Analysis (PCA), also known as Karhunen–Loéve transform (KLT) or the Hotelling transform.

We present a data-inspired model for PCA.

- Assume we have $D$ observed (measured) variables:
  $y = [y_1, \ldots, y_D]^T$. This is our data.

- Assume we know that our data is obtained by a linear transformation $W$ from $d$ unknown variables $x = [x_1, \ldots, x_d]^T$:
  \[
  y = W(x).
  \]

Typically we assume $d < D$.

- Assume moreover that the $D \times d$ matrix $W$ is a change of a coordinate system, i.e., columns of $W$ (or rows of $W^T$) are orthonormal to each other:
  \[
  W^T W = I_{d_d}.
  \]

Note that $WW^T$ need not be an identity.
Given the above assumptions the problem of PCA can be stated as follows:

*How can we find the transformation $W$ and the dimension $d$ from a finite number of measurements $y$?*

We shall need 2 additional assumptions:

- Assume that the unknown variables are Gaussian;
- Assume that both the unknown variables and the observations have mean zero (this is easily guaranteed by subtracting the mean, or the sample mean).
PCA from minimizing the reconstruction error

For a noninvertible matrix, we have its pseudoinverse defined as

$$W^+ = (W^T W)^{-1} W^T$$

In our case, $W^+ = W^T$, Thus, if $y = Wx$, we have

$$WW^T y = WW^T Wx = Wld_d x = y,$$

or, equivalently,

$$y - WW^T y = 0.$$  

With the presence of noise, we cannot assume anymore the perfect reconstruction, hence, we shall minimize the reconstruction error defined as

$$E_y (\|y - WW^T y\|_2^2).$$

It is not difficult to see that

$$E_y (\|y - WW^T y\|_2^2) = E_y (y^T y) - E_y (y^T WW^T y).$$
As $E_y(y^T y)$ is constant, our minimization of error reconstruction turns into a maximization of $E_y(y^T WW^T y)$. In reality, we know little about $y$, so we have to rely on the measurements $y(k), k = 1, \ldots, N$. Then,

$$E_y(y^T WW^T y) \sim \frac{1}{N} \sum_{n=1}^{N} (y(n))^T WW^T (y(n)) \sim \frac{1}{N} \text{tr}(Y^T WW^T Y),$$

where $Y$ is the matrix whose columns are the measurements $y(n)$ (hence $Y$ is a $D \times N$ matrix).

Using SVD for $Y$: $Y = V\Sigma U^T$, we obtain:

$$E_y(y^T WW^T y) \sim \frac{1}{N} \text{tr}(U\Sigma^T V^T WW^T V\Sigma U^T).$$

Therefore, after some computations we obtain:

$$\arg\max_W E_y(y^T WW^T y) \sim V \text{Id}_{D \times d},$$

and so $x \sim \text{Id}_{d \times D} V^T y$. 

Wojciech Czaja  
Harmonic Analysis and Big Data

Points close on the manifold should remain close in $\mathbb{R}^d$.

Use Laplace-Beltrami operator $\Delta_M$ to control the embedding.

Use discrete approximations for practical problems.

Proven convergence (Belkin and Niyogi, 2003 – 2008).

Generalized by Diffusion Maps and Diffusion Wavelets.


M. Belkin, P. Niyogi, Laplacian Eigenmaps for Dimensionality Reduction and Data Representation, Neural Computation, vol.15 no. 6 (2003), pp. 1373–1396.

1. Put an edge between nodes $i$ and $j$ if $x_i$ and $x_j$ are close. Precisely, given a parameter $k \in \mathbb{N}$, put an edge between nodes $i$ and $j$ if $x_i$ is among the $k$ nearest neighbors of $x_j$ or vice versa.

2. Given a parameter $t > 0$, if nodes $i$ and $j$ are connected, set

$$[W_{t,n}]_{i,j} = e^{-\frac{\|x_i - x_j\|^2}{4t}}.$$  

3. Set $[D_{t,n}]_{i,i} = \sum_j [W_{t,n}]_{i,j}$, and let $L_{t,n} = D_{t,n} - W_{t,n}$. Solve $L_{t,n}f = \lambda D_{t,n}f$, under the constraint $y^\top D_{t,n}y = Id$. Let $f_0, f_1, \ldots, f_d$ be $d + 1$ eigenvector solutions corresponding to the first eigenvalues $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_d$. Discard $f_0$ and use the next $d$ eigenvectors to embed in $d$-dimensional Euclidean space using the map $x_i \rightarrow (f_1(i), f_2(i), \ldots, f_d(i))$. 

Nobert Wiener Center for Harmonic Analysis and Applications

Wojciech Czaja  
Harmonic Analysis and Big Data
Swiss Roll

Figure: a) Original, b) PCA in $\mathbb{R}^2$

A. Halevy, Extensions of Laplacian Eigenmaps for Manifold Learning, PhD. Thesis, University of Maryland College Park, 2011
Figure: a) Original, b) LE outputs in $\mathbb{R}^2$ for different choices of parameters

Image source: M. Belkin, P. Niyogi, Laplacian Eigenmaps for Dimensionality Reduction and Data Representation, Neural Computation, vol.15 no. 6 (2003), pp. 1373–1396.
Given \( n \) data points \( x_1, \ldots, x_n \) from a manifold \( M \)

We constructed an operator \( L_{t,n} \) which acts on a function \( f \) defined on the data points as follows:

\[
L_{t,n}(f)(x_i) = \sum_j f(x_j) e^{-\frac{||x_i - x_j||^2}{4t}} - \sum_j f(x_j) e^{-\frac{||x_i - x_j||^2}{4t}}
\]

Our goal is to show that such defined operator is related in a strong sense to the Laplace-Beltrami operator on the manifold \( M \).
Laplace operator and heat equation

- Heat equation \( \frac{\partial}{\partial t} u(x, t) + \Delta u(x, t) = 0 \), with \( u(x, 0) = f(x) \).
- Solution given by \( u(x, t) = H_t(f)(x) \), where

\[
H_t(f)(x) = \int_M h_t(x, y)f(y)d\mu(y).
\]

On \( \mathbb{R}^k \), we have

\[
h_t(x, y) = (4\pi t)^{-k/2}e^{-\frac{||x-y||^2}{4t}}.
\]

- Key to approximation is the following equality

\[
\Delta(f)(x) = - \frac{\partial}{\partial t} u(x, t) \bigg|_{t=0} = - \frac{\partial}{\partial t} H_t(f)(x) \bigg|_{t=0} = \lim_{t\to 0} \frac{1}{t}(f(x) - H_t(f)(x)).
\]
Key to approximation

\[
\Delta(f)(x) = \lim_{t \to 0} \frac{1}{t} (4 \pi t)^{-k/2} \left( \int_M f(x) e^{-\frac{\|x-y\|^2}{4t}} \, dy - \int_M f(y) e^{-\frac{\|x-y\|^2}{4t}} \, dy \right)
\]

\[
\tilde{\Delta}(f)(x) = \lim_{t \to 0} \frac{1}{t} (4 \pi t)^{-k/2} \left( \frac{1}{n} \sum_j f(x_j) e^{-\frac{\|x_j-x\|^2}{4t}} - \frac{1}{n} \sum_j f(x_j) e^{-\frac{\|x_j-x\|^2}{4t}} \right)
\]

\[
= \frac{1}{t} (4 \pi t)^{-k/2} \frac{1}{n} L_{t,n}(f)(x).
\]
Define $\tilde{L}_{t,n} : C(M) \rightarrow C(M)$:

$$
\tilde{L}_{t,n}(f)(x) = \frac{1}{t} (4\pi t)^{-k/2} \left( \frac{1}{n} \sum_j f(x_j) e^{-\frac{\|x_j - x \|^2}{4t}} - \frac{1}{n} \sum_j f(x_j) e^{-\frac{\|x_j - x \|^2}{4t}} \right)
$$

Define $L_t : L^2(M) \rightarrow L^2(M)$:

$$
L_t(f)(x) = \frac{1}{t} (4\pi t)^{-k/2} \left( \int_M f(x) e^{-\frac{\|x - y \|^2}{4t}} \, dy - \int_M f(y) e^{-\frac{\|x - y \|^2}{4t}} \, dy \right)
$$
A few more notations

- Laplace-Beltrami operator on the manifold $M$

$$\Delta_M : C^2(M) \rightarrow L^2(M), \quad \Delta_M(f) = -\text{div}(\nabla f)$$

- Heat kernel $h_t(x, y) : M \times M \rightarrow \mathbb{R}$

- Heat operator $H_t : L^2(M) \rightarrow L^2(M)$

$$H_t(f)(x) = \int_M h_t(x, y)f(y)\,dy$$

- Alternatively $H_t(f) = \exp(-t\Delta_M)(f)$

- Remainder $R_t = \frac{1}{t}(\text{Id} - H_t) - L_t$. 
Theorem (Pointwise Convergence - Belkin and Niyogi, 2007)

Let the data points $x_1, x_2, \ldots, x_n$ be sampled independently from a uniform distribution on a smooth compact manifold $M \subset \mathbb{R}^d$. Let $\alpha > 0$, and let $t_n = n^{-1/(k+2+\alpha)}$. For $f \in C^\infty(M)$ we then have that:

$$\lim_{n \to \infty} \tilde{L}_{t_n,n}(f)(x) = \frac{1}{\text{vol}(M)} \Delta_M(f)(x)$$

in probability, where $\text{vol}(M)$ is the volume of the manifold with respect to the canonical measure.

M. Belkin, P. Niyogi, Towards a Theoretical Foundation for Laplacian-Based Manifold Methods, COLT 2005.
Overview of the proof

- Usually we do not know the exact form of the heat kernel on the manifold.
- Usually we do not know the geodesic distance between points, only the distance in the ambient space.
- Henceforth, our strategy is as follows:

$$\tilde{L}_{t,n}(f)(p) \xrightarrow{n \to \infty} L_t(f)(p) \xrightarrow{t \to 0} \frac{1}{\text{vol}(M)} \Delta_M(f)(p).$$
Sample approximations

- Fix $f \in C^\infty(M)$, $p \in M$, $t > 0$.
- $\tilde{L}_{t,n}$ is nothing else but the sample average of $n$ i.i.d. random variables with expectation $E[\tilde{L}_{t,n}(f)(p)] = L_t(f)(p)$.

**Lemma (Hoeffding’s Inequality)**

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables such that $|X_i| \leq K$. Then

$$P \left\{ \left| \frac{\sum_i X_i}{n} - E[X_i] \right| > \epsilon \right\} \leq 2 \exp\left( -\frac{\epsilon^2 n}{2K^2} \right)$$

Thus $P \left[ |\tilde{L}_{t,n}(f)(p) - L_t(f)(p)| > \epsilon \right] \leq 2 \exp\left( -\epsilon^2 nt^2(4\pi t)^k / 2K^2 \right)$

Note that Hoeffding’s inequality yields a stronger result than, e.g., Chebyshev’s inequality.

Let $t = t_n = n^{-1/(k+2+\alpha)}$, $\alpha > 0$. Then for any $\epsilon > 0$,

$$\lim_{n \to \infty} P \left[ |\tilde{L}_{t,n}(f)(p) - L_t(f)(p)| > \epsilon \right] = 0.$$  

Therefore, if we can show that the following holds:

$$L_t(f)(p) \xrightarrow{t \to 0} \frac{1}{\text{vol}(M)} \Delta_M(f)(p),$$

then we can finish the proof of the theorem.
We can show the last approximation in 3 major steps:

- Reduce the integral to a ball $B$ in $M$, i.e., show that

$$\lim_{t \to 0} L_t(f)(p) = \lim_{t \to 0} \frac{1}{t} (4\pi t)^{-k/2} \int_B (f(p) - f(y)) e^{-\frac{\|p-y\|^2}{4t}} d\mu(y).$$

- Apply a change of coordinates using the exponential map to rewrite the integral in $\mathbb{R}^k$.

- Use a Taylor expansion to analyze the integral in $\mathbb{R}^k$. 
Step 1: Reduction to a ball

- Let \( d = \inf_{x \notin B} \|p - x\|^{2} \) and \( m = \mu(M \setminus B) \).
- Since \( B \) is open, we have that \( d > 0 \).
- Moreover,

\[
\left| \int_{B} f(y) e^{-\frac{\|p-y\|^{2}}{4t}} \, d\mu(y) - \int_{M} f(y) e^{-\frac{\|p-y\|^{2}}{4t}} \, d\mu(y) \right| \leq m \sup_{x \in M} |f(x)| e^{-d^2/4t}
\]

- Thus, the difference can be estimates as \( o(t^{a}) \) for any \( a > 0 \).
Step 2: Change of coordinates

- \(\exp_p : T_pM(\sim \mathbb{R}^k) \to M, \exp_p(0) = p.\)
- \(\exp_p\) sends straight lines through the origin to geodesics.
- We can choose a small ball \(\tilde{B} \subset \mathbb{R}^k\) such that \(\exp_p\) is a diffeomorphism onto its image \(B \subset M\).
- Rewrite \(f : M \to \mathbb{R}^k\) as \(\tilde{f}(x) = f(\exp_p(x)).\)
- We thus obtain the following:

\[
\Delta_M(f)(p) = \Delta(\tilde{f})(0) = -\sum_i \frac{\partial^2 \tilde{f}}{\partial x_i^2}(0).
\]

**Lemma (Beylkin, Niyogi, 2007)**

There exists a constant \(C > 0\) such that for any pair \(p, y \in M, \) where \(y = \exp_p(x)\), we have

\[
0 \leq \|x\|_k^2 - \|y - p\|_d^2 = g(x) \leq C\|x\|^4_k.
\]
Step 2: Change of coordinates

\[
\frac{1}{\text{vol}(M)} \frac{1}{t} (4\pi t)^{-k/2} \int_{\tilde{B}} (\tilde{f}(0) - \tilde{f}(x)) e^{-\frac{\|p - \exp_p(x)\|^2}{4t}} \sqrt{\text{det}(g_{i,j})} dx
\]

For the metric tensor \( g \) we have:

\[
\text{det}(g_{i,j}) = 1 + O(\|x\|^2)
\]
The key component of the previous approximations is the following term:

\[ A_t = \frac{1}{\text{vol}(M)} \frac{1}{t} (4\pi t)^{-k/2} \int_{\tilde{B}} (\tilde{f}(0) - \tilde{f}(x)) e^{-\|x\|^2/4t} \, dx. \]

\( \tilde{f}(x) - \tilde{f}(0) = x \nabla \tilde{f} + \frac{1}{2} x^T H x + O(\|x\|^3), \) where \( H \) denotes the Hessian.

Thus,

\[ A_t = -\frac{1}{\text{vol}(M)} \frac{1}{t} (4\pi t)^{-k/2} \int_{\tilde{B}} (x \nabla \tilde{f} + \frac{1}{2} x^T H x + O(\|x\|^3)) e^{-\|x\|^2/4t} \, dx. \]

Note that \( \int_{\tilde{B}} x_i e^{-\|x\|^2/4t} \, dx = 0 \) and \( \int_{\tilde{B}} x_i x_j e^{-\|x\|^2/4t} \, dx = 0 \) for \( i \neq j \).

Hence

\[ \text{vol}(M) \lim_{t \to 0} A_t = -\text{tr}(H) = \Delta_M(f)(p), \]

which completes the proof of the point wise convergence.
We can also prove stronger results about spectral convergence of Laplacian Eigenmaps.

\[ \text{Eig}(\tilde{L}_{t,n}) \rightarrow_{n \to \infty} \text{Eig}(L_t) \rightarrow_{t \to 0} \text{Eig}(\Delta_M) \]

Theorem (Spectral Convergence - von Luxburg, Bousquet, Belkin, 2004)

For a fixed, sufficiently small \( t \), let \( \lambda_{t,n}^i \) and \( \lambda_t^i \) denote the \( i \)th eigenvalues of \( \tilde{L}_{t,n} \) and \( L_t \), resp. Similarly, let \( e_{t,n}^i \) and \( e_t^i \) denote the corresponding eigenvectors. If \( \lambda_t^i < \frac{1}{2t} \), then

\[ \lim_{n \to \infty} \lambda_{t,n}^i = \lambda_t^i \quad \text{and} \quad \lim_{n \to \infty} \| e_{t,n}^i - e_t^i \| = 0. \]

Theorem (Spectral Convergence - Belkin, Niyogi, 2008)

Let $\lambda^i$ and $\lambda^i_t$ denote the $i$th eigenvalues of $\Delta_M$ and $L_t$, resp. Similarly, let $e^i$ and $e^i_t$ denote their corresponding eigenvectors. Then

$$\lim_{t \to 0} |\lambda^i - \lambda^i_t| = 0, \quad \lim_{t \to 0} \|e^i - e^i_t\| = 0.$$

M. Belkin, P. Niyogi, Convergence of Laplacian Eigenmaps, NIPS 2008
Sketch of proof of spectral convergence

- Denote the $i$th eigenvalue of $\frac{Id - H_t}{t}$ by $\lambda^i((Id - H_t)/t)$
- Because $H_t(f) = \exp(-t\Delta_M)(f)$, we have
  \[ \lambda^i((Id - H_t)/t) = \frac{1 - e^{-\lambda^i t}}{t} \quad \text{and} \quad \lim_{t \to 0} \lambda^i((Id - H_t)/t) = \lambda^i \]
- Eigenvectors are also equal: $e^i = e^i((Id - H_t)/t)$
- We now need to show that for each fixed $i$ and small $t$, the $i$th eigenpair of $L_t$ is close to that of $\frac{Id - H_t}{t}$

Wojciech Czaja
Harmonic Analysis and Big Data
Sketch of proof of spectral convergence

Lemma (Belkin, Niyogi, 2008)

Let $A, B$ be positive, self-adjoint operators with discrete spectrum arranged in increasing order. Let $R = A - B$ and assume that for all $f \in L^2(X)$, we have

$$\left| \frac{\langle R(f), f \rangle}{\langle A(f), f \rangle} \right| \leq \epsilon.$$

Then, for all $k$, $1 - \epsilon \leq \lambda^k(B)/\lambda^k(A) \leq 1 + \epsilon$.

- $|\langle A(f), f \rangle| \leq |\langle B(f), f \rangle| + |\langle R(f), f \rangle| \leq |\langle B(f), f \rangle| + \epsilon |\langle A(f), f \rangle|$
- $|\langle A(f), f \rangle| \geq |\langle B(f), f \rangle| - |\langle R(f), f \rangle| \geq |\langle B(f), f \rangle| - \epsilon |\langle A(f), f \rangle|$
- $(1 - \epsilon)|\langle A(f), f \rangle| \leq |\langle B(f), f \rangle| \leq (1 + \epsilon)|\langle A(f), f \rangle|$
- If $H$ is a $k$-dimensional subspace of $L^2(X)$, then

$$(1 - \epsilon) \max_{H} \min_{f \in H^\perp} |\langle B(f), f \rangle| \leq \max_{H} \min_{f \in H^\perp} |\langle B(f), f \rangle|$$

- The result follows now from Courant-Fisher (min-max) theorem.
Combining the previous lemma with the following theorem, completes the proof of spectral convergence.

**Theorem (Belkin, Niyogi, 2008)**

For small enough $t$, there exists a constant $C > 0$, independent of $t$, such that

$$
\sup_{f \in L^2(M)} \left| \frac{\langle R(f), f \rangle}{\langle \frac{\text{Id} - H_t}{t}(f), f \rangle} \right| \leq Ct^{\frac{2}{\kappa+6}}.
$$
For the following considerations we fix $t$ and $n$, hence we can discard for the moment the subscript $t, n$, without loss of generality. Consider the following minimization problem, $y \in \mathbb{R}^d$,

$$
\min_{y^\top Dy = ld} \frac{1}{2} \sum_{i,j} \|y_i - y_j\|^2 W_{i,j} = \min_{y^\top Dy = E} tr(y^\top Ly).
$$

Its solution is given by the $d$ minimal non-zero eigenvalue solutions of $Lf = \lambda Df$ under the constraint $y^\top Dy = ld$.

Similarly, for diagonal $\alpha \cdot V$, $\alpha > 0$, consider the problem

$$
\min_{y^\top Dy = ld} \frac{1}{2} \sum_{i,j} \|y_i - y_j\|^2 W_{i,j} + \alpha \sum_i \|y_i\|^2 V_{i,i} = \min_{y^\top Dy = E} tr(y^\top (L + \alpha \cdot V)y),
$$

which leads to solving equation $(L + \alpha V)f = \lambda Df$. 

(1)
With M. Ehler, 2011
Want to go from un-supervised to semi-supervised learning
In SE, replace $L$ by $L + V$, where $V$ is, e.g., a barrier potential
Schroedinger Eigenmaps allow for the use of labeled data
Enforce certain relations between the nodes
Allow to utilize expert input or templates in otherwise fully automated techniques such as LE
Have proven convergence results (with A. Halevy)


A. Halevy, Extensions of Laplacian Eigenmaps for Manifold Learning, PhD. Thesis, University of Maryland College Park, 2011
Let the data graph be connected and let $V$ be a symmetric positive semi-definite matrix.

**Theorem (with M. Ehler)**

Let the data graph be connected, let $V$ be a symmetric positive semi-definite, and let $n \leq \dim(\text{Null}(V))$. Then the minimizer of (1) satisfies:

$$\|y^{(\alpha)}\|^2_V = \text{trace}^2(y^{(\alpha)^T}Vy^{(\alpha)}) \leq C \frac{1}{\alpha}. $$

In particular, if $V = \text{diag}(v_1, \ldots, v_N)$, then

$$v_i \|y_i^{(\alpha)}\|^2 \leq \sum_{i=1}^{N} v_i \|y_i^{(\alpha)}\|^2 \leq C_1 \frac{1}{\alpha}, \quad \text{for all } i = 1, \ldots, N.$$
Given \( n \) data points \( x_1, x_2, \ldots, x_n \) sampled independently from a uniform distribution on a smooth, compact, \( d \)-dimensional manifold \( \mathcal{M} \subset \mathbb{R}^D \), define the operator \( \hat{L}_{t,n} : C(\mathcal{M}) \to C(\mathcal{M}) \) by

\[
\hat{L}_{t,n}(f)(x) = \frac{1}{(4\pi t)^{d/2}t} \left( \frac{1}{n} \sum_j f(x) e^{-\frac{\|x-x_j\|^2}{4t}} - \frac{1}{n} \sum_j f(x_j) e^{-\frac{\|x-x_j\|^2}{4t}} \right).
\]

Let \( \nu \in C(\mathcal{M}) \) be a potential. For \( x \in \mathcal{M} \), let \( y_n(x) = \arg\min_{x_1,x_2,\ldots,x_n} \|x-x_i\| \) \( x_i \in \{x_1,x_2,\ldots,x_n\} \) and define \( V_n : C(\mathcal{M}) \to C(\mathcal{M}) \) by \( V_nf(x) = \nu(y_n(x))f(x) \).

**Theorem (Pointwise Convergence, with A. Halevy)**

Let \( \alpha > 0 \), and set \( t_n = \left( \frac{1}{n} \right)^{\frac{1}{\alpha+2+\alpha}} \). For \( f \in C^{\infty}(\mathcal{M}) \),

\[
\lim_{n \to \infty} \hat{L}_{t_n,n}f(x) + V_nf(x) = C\Delta_{\mathcal{M}}f(x) + \nu(x)f(x) \text{ in probability.}
\]
Let $\hat{L}_{t,n}$ be the unnormalized discrete Laplacian.

**Theorem (Spectral Convergence of SE, with A. Halevy)**

Let $\lambda_{t,n}^i$ and $e_{t,n}^i$ be the $i$th eigenvalue and corresponding eigenfunction of $\hat{L}_{t,n} + V_n$. Let $\lambda_i$ and $e_i$ be the $i$th eigenvalue and corresponding eigenfunction of $\Delta_M + V$. Then there exists a sequence $t_n \to 0$ such that, in probability,

$$\lim_{n \to \infty} \lambda_{t,n}^i = \lambda_i \quad \text{and} \quad \lim_{n \to \infty} \|e_{t,n}^i - e_i\| = 0.$$
The Schroedinger Eigenmaps with diagonal potential $V = \text{diag}(0, \ldots, 0, 1, 0, \ldots, 0)$ only acting in one point $y_{i_0}$ in the middle of the arc for $\alpha = 0.05, 0.1, 0.5, 5$. This point is pushed to zero.
By applying the potential to the end points of the arc for \( \alpha = 0.01, 0.05, 0.1, 1 \), we are able to control the dimension reduction such that we obtain an almost perfect circle.
(left) The Indian Pines image is a $145 \times 145$ pixel image with 224 spectral bands. It was acquired using an AVIRIS spectrometer. (right) The Pavia University image is a $610 \times 340$ pixel image that contains 115 spectral bands. It was acquired using a ROSIS sensor.

Data courtesy of:
Prof. Landgrebe (Purdue University, USA): the Indian Pines dataset,

Prof. Paolo Gamba (Pavia University, Italy): the Pavia University dataset.
Let $\iota = \{i_1, \ldots, i_m\}$ be a collection of nodes. A cluster potential over $\iota$ is defined by:

$$V[\iota, \iota] := \begin{pmatrix}
1 & -1 \\
-1 & 2 & -1 \\
& \ddots & \ddots & \ddots \\
& & -1 & 2 & -1 \\
& & & -1 & 1
\end{pmatrix}.$$  

For such a $V$, the minimization problem is equivalent to:

$$\min_{y^T D y = I} \frac{1}{2} \sum_{i,j} \|y_i - y_j\|^2 W_{i,j} + \alpha \sum_{k=1}^{m-1} \|y_{i_k} - y_{i_{k+1}}\|^2.$$
Use k-means clustering to cluster the hyperspectral image. The clustering was initialized randomly so that the construction of the potential would be completely unsupervised. The Euclidean metric was used.

Define a cluster potential, $V_k$ over each of the $k = 1, \ldots, K$ clusters. The order of the pixels in each cluster is determined by their index (smallest to largest).

Aggregate into one potential by summing the individual cluster potentials:

$$V_K = \sum_{k=1}^{K} V_k.$$
Impact of SE on Cluster analysis

Pavia University: Dimensions 4 and 5 of the LE and SE embeddings for classes 2 (meadows), 3 (gravel), and 7 (bitumen)

Impact of SE on Cluster analysis

Indian Pines: Dimensions 17 and 22 of the LE and SE embeddings for classes 2 (corn 1), 3 (corn 2), and 10 (soybeen 1)
We have introduced several examples of data dependent representation methods.

These methods are well suited for the analysis of complex, noisy, high-dimensional data.

As a drawback, we are going to encounter increased computational requirements relative to fast a priori methods, such as FFT or DWT.

Also, the nonlinear dimensionality reduction methods are typically non-invertible.