Vector Diffusion Maps and the Connection Laplacian

Amit Singer

Princeton University, Department of Mathematics and PACM

Duke Workshop on Sensing and Analysis of High-Dimensional Data
July 26, 2011
Motivating Problem: Cryo-Electron Microscopy

- Projection images $P_i(x, y) = \int_{-\infty}^{\infty} \phi(xR_i^1 + yR_i^2 + zR_i^3) \, dz + \text{“noise”}$. 
- $\phi : \mathbb{R}^3 \mapsto \mathbb{R}$ is the electric potential of the molecule.
- Cryo-EM problem: Find $\phi$ and $R_1, \ldots, R_n$ given $P_1, \ldots, P_n$. 

$R_i = \begin{pmatrix} R_i^1 & R_i^2 & R_i^3 \end{pmatrix} \in \text{SO}(3)$
Toy Example
E. coli 50S ribosomal subunit: sample images
Class Averaging in Cryo-EM: Improve SNR
Current clustering method (Penczek, Zhu, Frank 1996)

- Projection images $P_1, P_2, \ldots, P_n$ with unknown rotations $R_1, R_2, \ldots, R_n \in SO(3)$
- Rotationally Invariant Distances (RID)

$$d_{RID}(i,j) = \min_{O \in SO(2)} \| P_i - OP_j \|$$

- Cluster the images using K-means.
- Images are not centered; also possible to include translations and to optimize over the special Euclidean group.
- Problem with this approach: outliers.
- At low SNR images with completely different viewing directions may have relatively small $d_{RID}$ (noise aligns well, instead of underlying signal).
Outliers: Small World Graph on $S^2$

- Define graph $G = (V, E)$ by $\{i, j\} \in E \iff d_{RID}(i, j) \leq \varepsilon$.

- Optimal rotation angles

  $$O_{ij} = \arg\min_{O \in SO(2)} \|P_i - OP_j\|, \quad i, j = 1, \ldots, n.$$ 

- Triplet consistency relation – good triangles

  $$O_{ij}O_{jk}O_{ki} \approx Id.$$ 

- How to use information of optimal rotations in a systematic way?

  Vector Diffusion Maps

Amit Singer (Princeton University)
In VDM, the relationships between data points (e.g., cryo-EM images) are represented as a weighted graph, where the weights $w_{ij}$ describing affinities between data points are accompanied by linear orthogonal transformations $O_{ij}$. 
Manifold Learning: Point cloud in $\mathbb{R}^p$

- $x_1, x_2, \ldots, x_n \in \mathbb{R}^p$.
- Manifold assumption: $x_1, \ldots, x_n \in \mathcal{M}^d$, with $d \ll p$.
- Local Principal Component Analysis (PCA) gives an approximate orthonormal basis $O_i$ for the tangent space $T_{x_i}\mathcal{M}$.
- $O_i$ is a $p \times d$ matrix with orthonormal columns: $O_i^T O_i = I_{d\times d}$.
- Alignment: $O_{ij} = \text{argmin}_{O \in O(d)} \| O - O_i^T O_j \|_{HS}$ (computed through the singular value decomposition of $O_i^T O_j$).
Parallel Transport

- $O_{ij}$ approximates the parallel transport operator $P_{x_i,x_j} : T_{x_j} \mathcal{M} \to T_{x_i} \mathcal{M}$
Vector diffusion mapping: $S$ and $D$

- Symmetric $nd \times nd$ matrix $S$:

$$S(i,j) = \begin{cases} w_{ij}O_{ij} & (i,j) \in E, \\ 0_{d \times d} & (i,j) \notin E. \end{cases}$$

$n \times n$ blocks, each of which is of size $d \times d$.

- Diagonal matrix $D$ of the same size, where the diagonal $d \times d$ blocks are scalar matrices with the weighted degrees:

$$D(i,i) = \deg(i)I_{d \times d},$$

and

$$\deg(i) = \sum_{j: (i,j) \in E} w_{ij}$$
The matrix $D^{-1}S$ can be applied to vectors $v$ of length $nd$, which we regard as $n$ vectors of length $d$, such that $v(i)$ is a vector in $\mathbb{R}^d$ viewed as a vector in $T_{x_i}\mathcal{M}$. The matrix $D^{-1}S$ is an averaging operator for vector fields, since

$$(D^{-1}Sv)(i) = \frac{1}{\deg(i)} \sum_{j: (i,j) \in E} w_{ij} O_{ij} v(j).$$

This implies that the operator $D^{-1}S$ transport vectors from the tangent spaces $T_{x_j}\mathcal{M}$ (that are nearby to $T_{x_i}\mathcal{M}$) to $T_{x_i}\mathcal{M}$ and then averages the transported vectors in $T_{x_i}\mathcal{M}$. 
Affinity between nodes based on consistency of transformations

- In the VDM framework, we define the affinity between $i$ and $j$ by considering all paths of length $t$ connecting them, but instead of just summing the weights of all paths, we sum the transformations.
- Every path from $j$ to $i$ may result in a different transformation (like parallel transport due to curvature).
- When adding transformations of different paths, cancelations may happen.
- We define the affinity between $i$ and $j$ as the consistency between these transformations.
- $D^{-1}S$ is similar to the symmetric matrix $\tilde{S}$

$$\tilde{S} = D^{-1/2} S D^{-1/2}$$

- We define the affinity between $i$ and $j$ as

$$\|\tilde{S}^{2t}(i,j)\|_{HS}^{2} = \frac{\text{deg}(i)}{\text{deg}(j)} \| (D^{-1}S)^{2t}(i,j) \|_{HS}^{2}.$$
Embedding into a Hilbert Space

- Since $\tilde{S}$ is symmetric, it has a complete set of eigenvectors \( \{v_l\}_{l=1}^{nd} \) and eigenvalues \( \{\lambda_l\}_{l=1}^{nd} \) (ordered as \( |\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_{nd}| \)).
- Spectral decompositions of $\tilde{S}$ and $\tilde{S}^{2t}$:

\[
\tilde{S}(i,j) = \sum_{l=1}^{nd} \lambda_l v_l(i) v_l(j)^T, \quad \text{and} \quad \tilde{S}^{2t}(i,j) = \sum_{l=1}^{nd} \lambda_l^{2t} v_l(i) v_l(j)^T,
\]

where $v_l(i) \in \mathbb{R}^d$ for $i = 1, \ldots, n$ and $l = 1, \ldots, nd$.
- The HS norm of $\tilde{S}^{2t}(i,j)$ is calculated using the trace:

\[
\|\tilde{S}^{2t}(i,j)\|_{HS}^2 = \sum_{l,r=1}^{nd} (\lambda_l \lambda_r)^{2t} \langle v_l(i), v_r(i) \rangle \langle v_l(j), v_r(j) \rangle.
\]
- The affinity $\|\tilde{S}^{2t}(i,j)\|_{HS}^2 = \langle V_t(i), V_t(j) \rangle$ is an inner product for the finite dimensional Hilbert space $\mathbb{R}^{(nd)^2}$ via the mapping $V_t$:

\[
V_t : i \mapsto ((\lambda_l \lambda_r)^{t} \langle v_l(i), v_r(i) \rangle)_{l,r=1}^{nd}.
\]
Vector Diffusion Distance

- The vector diffusion mapping is defined as
  \[ V_t : i \mapsto ((\lambda_l \lambda_r)^t \langle v_l(i), v_r(i) \rangle)_{l,r=1}^{nd}. \]

- The vector diffusion distance between nodes \( i \) and \( j \) is denoted \( d_{\text{VDM},t}(i,j) \) and is defined as
  \[ d^2_{\text{VDM},t}(i,j) = \langle V_t(i), V_t(i) \rangle + \langle V_t(j), V_t(j) \rangle - 2 \langle V_t(i), V_t(j) \rangle. \]

- Other normalizations of the matrix \( S \) are possible and lead to slightly different embeddings and distances (similar to diffusion maps).

- The matrices \( I - \tilde{S} \) and \( I + \tilde{S} \) are positive semidefinite, because
  \[ v^T(I \pm D^{-1/2} SD^{-1/2})v = \sum_{(i,j) \in E} \left\| \frac{v(i)}{\sqrt{\deg(i)}} \pm \frac{w_{ij}O_{ij}v(j)}{\sqrt{\deg(j)}} \right\|^2 \geq 0, \]
  for any \( v \in \mathbb{R}^{nd} \). Therefore, \( \lambda_l \in [-1, 1] \). As a result, the vector diffusion mapping and distances can be well approximated by using only the few largest eigenvalues and their corresponding eigenvectors.
Application to the class averaging problem in Cryo-EM

(a) Neighbors are identified using $d_{\text{RID}}$  
(b) Neighbors are identified using $d_{\text{VDM, } t=2}$

Figure: SNR=1/64: Histogram of the angles (x-axis, in degrees) between the viewing directions of each image (out of 40000) and its 40 neighboring images. Left: neighbors are identified using the original rotationally invariant distances $d_{\text{RID}}$. Right: neighbors are post identified using vector diffusion distances.
Convergence Theorem to the Connection-Laplacian

Let \( \iota : \mathcal{M} \hookrightarrow \mathbb{R}^p \) be a smooth \( d \)-dim closed Riemannian manifold embedded in \( \mathbb{R}^p \), with metric \( g \) induced from the canonical metric on \( \mathbb{R}^p \), and the data set \( \{x_i\}_{i=1,\ldots,n} \) is independently uniformly distributed over \( \mathcal{M} \). Let \( K \in C^2(\mathbb{R}^+) \) be a positive kernel function decaying exponentially, that is, there exist \( T > 0 \) and \( C > 0 \) such that \( K(t) \leq Ce^{-t} \) when \( t > T \). For \( \epsilon > 0 \), let \( K_{\epsilon}(x_i, x_j) = K\left(\frac{\|\iota(x_i) - \iota(x_j)\|_{\mathbb{R}^p}}{\sqrt{\epsilon}}\right) \). Then, for \( X \in C^3(\mathcal{T}\mathcal{M}) \) and for all \( x_i \) almost surely we have

\[
\lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{\epsilon} \left[ \frac{\sum_{j=1}^{n} K_{\epsilon}(x_i, x_j) O_{ij} X_j}{\sum_{j=1}^{n} K_{\epsilon}(x_i, x_j)} - x_i \right] = \frac{m_2}{2dm_0} \left( \langle \iota_* \nabla^2 X(x_i), e_l \rangle \right)_{l=1}^d,
\]

where \( \nabla^2 \) is the connection Laplacian, \( X_i \equiv \left( \langle \iota_* X(x_i), e_l \rangle \right)_{l=1}^d \in \mathbb{R}^d \) for all \( i \), \( \{e_l(x_i)\}_{l=1,\ldots,d} \) is an orthonormal basis of \( \iota_* T_{x_i} \mathcal{M} \), \( m_l = \int_{\mathbb{R}^d} \|x\|^l K(\|x\|)dx \), and \( O_{ij} \) is the optimal orthogonal transformation determined by the algorithm in the alignment step.
Example: Connection-Laplacian for $S^d$ embedded in $\mathbb{R}^{d+1}$

The connection-Laplacian commutes with rotations and the eigenvalues and eigen-vector-fields are calculated using representation theory:

\[
S^2 : 6, 10, 14, \ldots ,
\]
\[
S^3 : 4, 6, 9, 16, 16, \ldots .
\]
\[
S^4 : 5, 10, 14, \ldots .
\]
\[
S^5 : 6, 15, 20, \ldots .
\]

Figure: Bar plots of the largest 30 eigenvalues of $D^{-1}S$ for $n = 8000$ points uniformly distributed over spheres of different dimensions.
More applications of VDM: Orientability from a point cloud

Encode the information about reflections in a symmetric $n \times n$ matrix $Z$ with entries

$$Z_{ij} = \begin{cases} \det O_{ij} & (i, j) \in E, \\ 0 & (i, j) \notin E. \end{cases}$$

That is, $Z_{ij} = 1$ if no reflection is needed, $Z_{ij} = -1$ if a reflection is needed, and $Z_{ij} = 0$ if the points are not nearby. Normalize $Z$ by the node degrees.

Figure: Histogram of the values of the top eigenvector of $D^{-1}Z$. 

(a) $S^2$  (b) Klein bottle  (c) $\mathbb{RP}^2$
Orientable Double Covering

Embedding obtained using the eigenvectors of the (normalized) matrix

$$\begin{bmatrix}
  Z & -Z \\
  -Z & Z
\end{bmatrix} = \left( \begin{array}{cc}
  1 & -1 \\
  -1 & 1
\end{array} \right) \otimes Z,$$

Figure: Left: the orientable double covering of $\mathbb{R}P(2)$, which is $S^2$; Middle: the orientable double covering of the Klein bottle, which is $T^2$; Right: the orientable double covering of the Möbius strip, which is a cylinder.
Ongoing Research in cryo-EM

- Molecules with symmetries
- Heterogeneity problem
- Signal/Image processing
VDM is a generalization of diffusion maps: from functions to vector fields.

A way to globally connect local PCAs.

Vector diffusion distance: a new metric for data points

Noise robustness: random matrix theory (noise model – orthogonal transformations average to 0).

Other higher order Laplacians from point clouds (e.g., the Hodge Laplacian).

Revealing the topology of the data (e.g., orientability).

Diffusion on orbit spaces $\mathcal{M}/G$.

More applications.
References

- A. Singer, H.-T. Wu, “Vector diffusion maps and the connection Laplacian”, submitted.
- A. Singer, H.-T. Wu, “Orientability and Diffusion Maps”, *Applied and Computational Harmonic Analysis, Applied and Computational Harmonic Analysis*, 31 (1), pp. 44–58 (2011).
- A. Singer, Z. Zhao, Y. Shkolnisky, R. Hadani, “Viewing Angle Classification of Cryo-Electron Microscopy Images using Eigenvectors”, *SIAM Journal on Imaging Sciences, SIAM Journal on Imaging Sciences*, 4 (2), pp. 723–759 (2011).
Thank You!

Acknowledgements:

Students:
- Hau-tieng Wu
- Zhizhen Zhao

Collaborators:
- Ronny Hadani (UT Austin)
- Yoel Shkolnisky (Tel Aviv University)
- Fred Sigworth (Yale Medical School)

Funding:
- NIH/NIGMS R01GM090200
- Sloan Research Foundation