Beyond Scalar Affinities for Network Analysis
or
Vector Diffusion Maps and the Connection Laplacian

Amit Singer

Princeton University

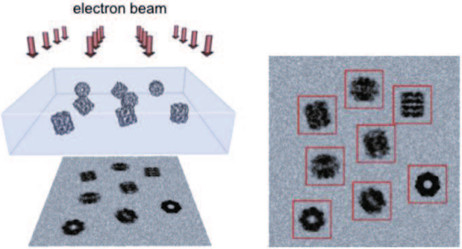
Department of Mathematics and Program in Applied and Computational Mathematics

May 28, 2015

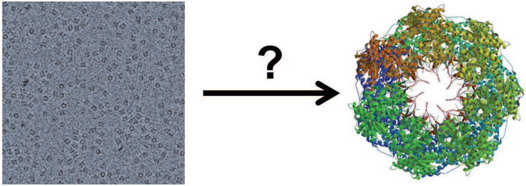
National Institute of General Medical Sciences
Basic Discoveries for Better Health
SIMONS FOUNDATION
GORDON AND BETTY MOORE FOUNDATION

Single Particle Reconstruction using cryo-EM

Schematic drawing of the imaging process:



The cryo-EM problem:



tomographic image

3-D structure

Main Algorithmic Challenges

1. Orientation assignment
2. Heterogeneity (resolving structural variability)
3. **2D Class averaging (de-noising)**
4. Symmetry detection
5. Motion correction
6. Particle picking

Class Averaging in Cryo-EM: Improve SNR

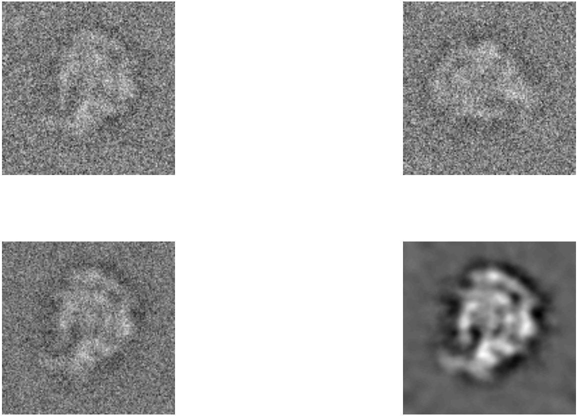
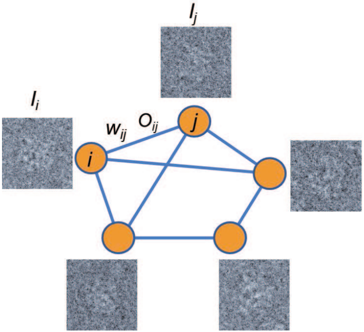
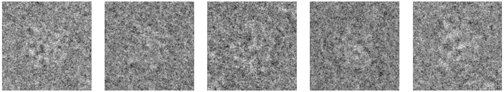


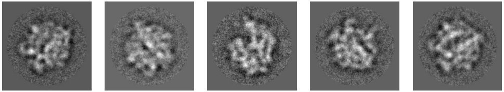
Image denoising by vector diffusion maps

- Generalization of Laplacian Eigenmaps (Belkin, Niyogi 2003) and Diffusion Maps (Coifman, Lafon 2006)
- Introduced the graph Connection Laplacian
- S, Zhao, Shkolnisky, Hadani (SIIMS 2011)
- Hadani, S (FoCM 2011)
- S, Wu (Comm. Pure Appl. Math 2012)
- Zhao, S (J Struct. Bio. 2014)





Experimental images (70S) courtesy of Dr. Joachim Frank (Columbia)



Class averages by vector diffusion maps (averaging with 20 nearest neighbors)

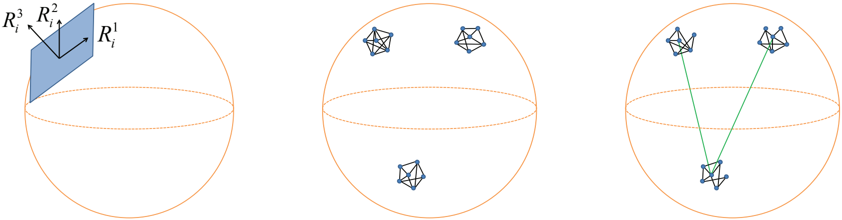
Rotation Invariant Distances

- Projection images $I_1, I_2, \ldots, I_n$ with unknown rotations $R_1, R_2, \ldots, R_n \in SO(3)$
- Rotationally Invariant Distances (RID)

$$d_{RID}(i,j) = \min_{O \in SO(2)} \|I_i - O \circ I_j\|$$

- 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).

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



- Optimal rotation angles

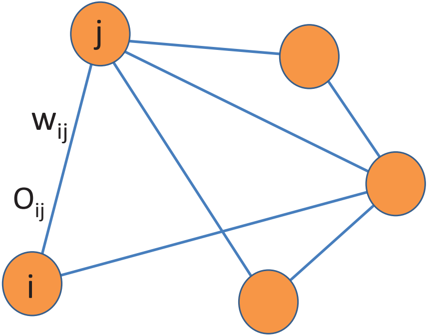
$$O_{ij} = \operatorname*{argmin}_{O \in SO(2)} \|I_i - O \circ I_j\|, \quad i, j = 1, \ldots, n.$$

- Triplet consistency relation – good triangles

$$O_{ij} O_{jk} O_{ki} \approx I_{2\times 2}.$$

- How to use information of optimal rotations in a systematic way? Vector Diffusion Maps
- "Non-local means with rotations"

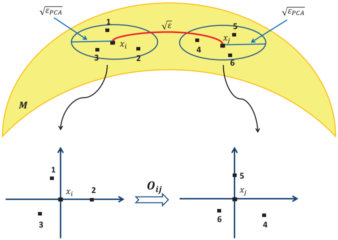
Vector Diffusion Maps: Setup

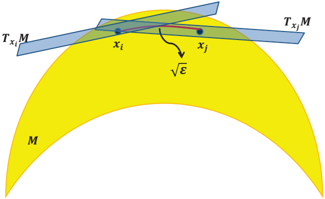


In VDM, the relationships between data points (e.g., cryo-EM images) are represented as a weighted graph, where the weights w_{ij} describing affinities between data points are accompanied by linear orthogonal transformations O_{ij}.

Manifold Learning: Point cloud in $\mathbb{R}^p$

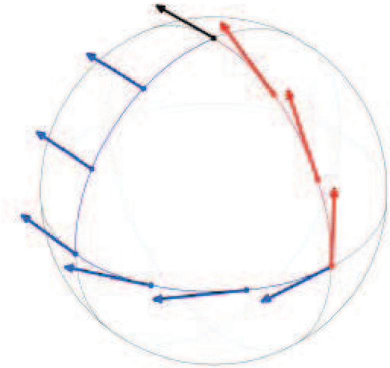
- $x_1, x_2, \ldots, x_n \in \mathbb{R}^p$.
- Manifold assumption: $x_1, \ldots, x_n \in \mathcal{M}^d$, with $d \ll p$.
- Local Principal Component Analysis (PCA) gives an approximate orthonormal basis O_i for the tangent space $T_{x_i}\mathcal{M}$.
- O_i is a $p \times d$ matrix with orthonormal columns: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





Parallel Transport

- O_{ij} approximates the parallel transport operator $P_{x_i,x_j} : T_{x_j}\mathcal{M} \to T_{x_i}\mathcal{M}$



Laplacian Eigenmap (Belkin and Niyogi 2003) and Diffusion Map (Coifman and Lafon 2006)

- Symmetric $n \times n$ matrix W_0:

$$W_0(i,j) = \begin{cases} w_{ij} & (i,j) \in E, \\ 0 & (i,j) \notin E. \end{cases}$$

- Diagonal matrix D_0 of the same size:

$$D_0(i,i) = \deg(i) = \sum_{j:(i,j)\in E} w_{ij}.$$

- Graph Laplacian, Normalized graph Laplacian and the random walk matrix:

$$L_0 = D_0 - W_0, \quad \mathcal{L}_0 = I - D_0^{-1/2} W_0 D_0^{-1/2}, \quad A_0 = D_0^{-1} W_0$$

- The diffusion map Φ_t is defined in terms of the eigenvectors of A_0:

$$A_0 \phi_l = \lambda_l \phi_l, \quad l = 1, \ldots, n$$

$$\Phi_t : i \mapsto (\lambda_l^t \phi_l(i))_{l=1}^n.$$

Vector diffusion mapping: W_1 and D_1

- Symmetric $nd \times nd$ matrix W_1:

$$W_1(i,j) = \begin{cases} w_{ij} O_{ij} & (i,j) \in E, \\ 0_{d\times d} & (i,j) \notin E. \end{cases}$$

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

- Diagonal matrix D_1 of the same size, where the diagonal $d \times d$ blocks are scalar matrices with the weighted degrees:

$$D_1(i,i) = \deg(i) I_{d\times d},$$

  and

$$\deg(i) = \sum_{j:(i,j)\in E} w_{ij}$$

$A_1 = D_1^{-1} W_1$ is an averaging operator for vector fields

- The matrix A_1 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 $A_1 = D_1^{-1} W_1$ is an averaging operator for vector fields, since

$$(A_1 v)(i) = \frac{1}{\deg(i)} \sum_{j:(i,j)\in E} w_{ij} O_{ij} v(j).$$

  This implies that the operator A_1 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.
- $A_1 = D_1^{-1} W_1$ is similar to the symmetric matrix $\tilde{W}_1$

$$\tilde{W}_1 = D_1^{-1/2} W_1 D_1^{-1/2}$$

- We define the affinity between i and j as

$$\|\tilde{W}_1^{2t}(i,j)\|_{HS}^2 = \frac{\deg(i)}{\deg(j)} \|(D_1^{-1} W_1)^{2t}(i,j)\|_{HS}^2.$$

Embedding into a Hilbert Space

- Since $\tilde{W}_1$ 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 $\tilde{W}_1$ and $\tilde{W}_1^{2t}$:

$$\tilde{W}_1(i,j) = \sum_{l=1}^{nd} \lambda_l v_l(i) v_l(j)^T, \quad \text{and} \quad \tilde{W}_1^{2t}(i,j) = \sum_{l=1}^{nd} \lambda_l^{2t} v_l(i) v_l(j)^T,$$

where $v_l(i) \in \mathbb{R}^d$ for $i = 1, \ldots, n$ and $l = 1, \ldots, nd$.
- The HS norm of $\tilde{W}_1^{2t}(i,j)$ is calculated using the trace:

$$\|\tilde{W}_1^{2t}(i,j)\|_{HS}^2 = \sum_{l,r=1}^{nd} (\lambda_l \lambda_r)^{2t} \langle v_l(i), v_r(i)\rangle \langle v_l(j), v_r(j)\rangle.$$

- The affinity $\|\tilde{W}_1^{2t}(i,j)\|_{HS}^2 = \langle V_t(i), V_t(j)\rangle$ is an inner product for the finite dimensional Hilbert space $\mathbb{R}^{(nd)^2}$ via the mapping V_t:

$$V_t : i \mapsto \left((\lambda_l \lambda_r)^t \langle v_l(i), v_r(i)\rangle\right)_{l,r=1}^{nd}.$$

Vector Diffusion Distance

- The vector diffusion mapping is defined as

$$V_t : i \mapsto \left((\lambda_l \lambda_r)^t \langle v_l(i), v_r(i) \rangle \right)_{l,r=1}^{nd}.$$

- The vector diffusion distance between nodes i and j is denoted $d_{\mathrm{VDM},t}(i,j)$ and is defined as

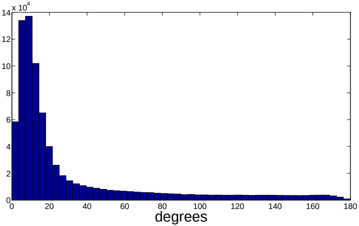
$$d^2_{\mathrm{VDM},t}(i,j) = \langle V_t(i), V_t(i) \rangle + \langle V_t(j), V_t(j) \rangle - 2\langle V_t(i), V_t(j) \rangle.$$

- Other normalizations of the matrix W_1 are possible and lead to slightly different embeddings and distances (similar to diffusion maps).
- The matrices $I - \tilde{W}_1$ and $I + \tilde{W}_1$ are positive semidefinite, because

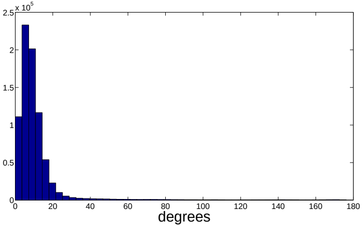
$$v^T (I \pm D_1^{-1/2} W_1 D_1^{-1/2}) v = \sum_{(i,j) \in E} \left\| \frac{v(i)}{\sqrt{\deg(i)}} \pm \frac{w_{ij} O_{ij} v(j)}{\sqrt{\deg(j)}} \right\|^2 \geq 0,$$

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 (S, Zhao, Shkolnisky, Hadani 2011)



(a) Neighbors are identified using d_{RID}



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

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

Computational Aspects

Zhao, S J. Struct. Biol. 2014

- Naïve implementation requires $O(n^2)$ rotational alignments of images
- Rotational invariant representation of images: “bispectrum”
- Dimensionality reduction using a randomized algorithm for PCA (Rokhlin, Liberty, Tygert, Martinsson, Halko, Tropp, Szlam, ...)
- Randomized approximated nearest neighbors search in nearly linear time (Jones, Osipov, Rokhlin 2011)

- There is no non-vanishing continuous tangent vector field on the sphere.
- Cannot find O_i such that $O_{ij} = O_i O_j^{-1}$.
- No global rotational alignment of all images.

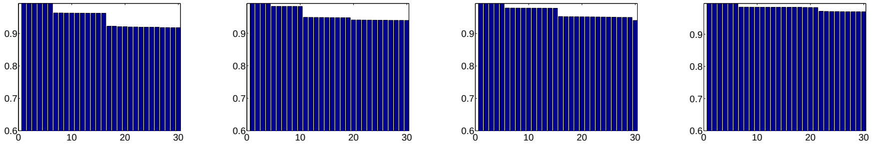
Let $\iota : \mathcal{M} \hookrightarrow \mathbb{R}^p$ be a smooth d-dim closed Riemannian manifold embedded in $\mathbb{R}^p$, with metric g induced from the canonical metric on $\mathbb{R}^p$, and the data set $\{x_i\}_{i=1,\ldots,n}$ is independently uniformly distributed over $\mathcal{M}$. Let $K \in C^2(\mathbb{R}^+)$ 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 $X \in C^3(T\mathcal{M})$ 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 ∇^2 is the connection Laplacian, $X_i \equiv (\langle \iota_* X(x_i), e_l\rangle)_{l=1}^d \in \mathbb{R}^d$ for all i, $\{e_l(x_i)\}_{l=1,\ldots,d}$ is an orthonormal basis of $\iota_* T_{x_i}\mathcal{M}$, $m_l = \int_{\mathbb{R}^d} \|x\|^l K(\|x\|)\mathrm{d}x$, and O_{ij} is the optimal orthogonal transformation determined by the algorithm in the alignment step.

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

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



(a) S^2 (b) S^3 (c) S^4 (d) S^5

Figure : Bar plots of the largest 30 eigenvalues of A_1 for $n = 8000$ points uniformly distributed over spheres of different dimensions.

Spectral graph theory with vector fields

- The Graph Connection Laplacian

$$L_1 = D_1 - W_1$$

- The Normalized Graph Connection Laplacian

$$\mathcal{L}_1 = I - D_1^{-1/2} W_1 D^{-1/2} = I - \tilde{W}_1$$

- Averaging operator / random walk matrix for vector diffusion:

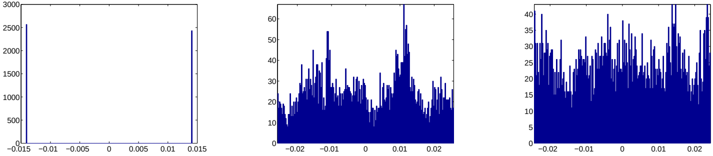
$$A_1 = D_1^{-1} W_1$$

- A Cheeger inequality for the graph connection Laplacian (Bandeira, S, Spielman SIAM Matrix Analysis 2013)

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.



(a) S^2 (b) Klein bottle (c) $\mathbb{R}P^2$

Figure : Histogram of the values of the top eigenvector of $D_0^{-1} Z$.

Orientable Double Covering

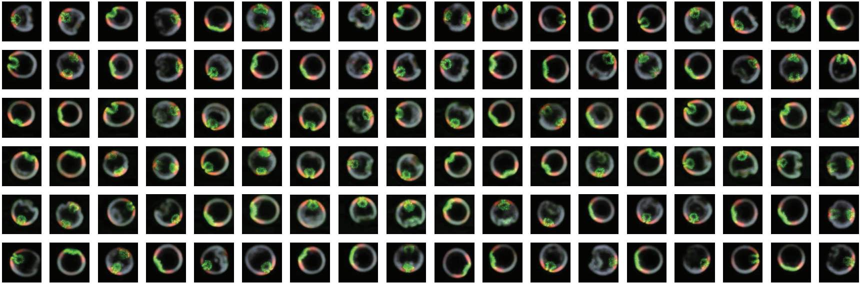
Embedding obtained using the eigenvectors of the (normalized) matrix

$$\begin{bmatrix} Z & -Z \\ -Z & Z \end{bmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \otimes Z,$$

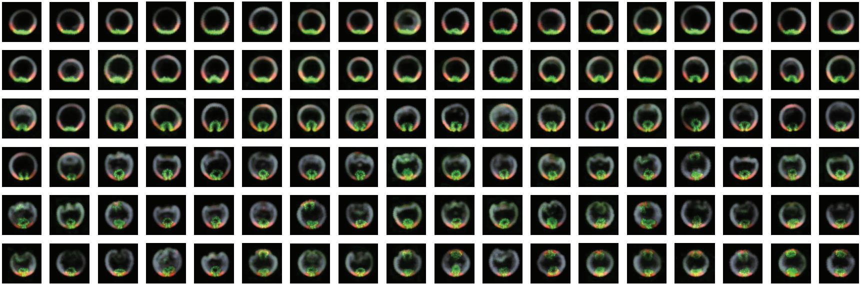
Figure : Left: the orientable double covering of $\mathbb{R}P(2)$, which is S^2; Middle: the orientable double covering of the Klein bottle, which is T^2; Right: the orientable double covering of the Möbius strip, which is a cylinder.

Registration and Ordering of *Drosophila* Embryogenesis Images

- Dsilva, Lim, Lu, S, Kevrekidis, Shvartsman, Development 2015



Registered and Ordered using VDM



References

- A. Singer, Z. Zhao, Y. Shkolnisky, R. Hadani, "Viewing Angle Classification of Cryo-Electron Microscopy Images using Eigenvectors", *SIAM Journal on Imaging Sciences*, **4** (2), pp. 723–759 (2011).
- A. Singer, H.-T. Wu, "Orientability and Diffusion Maps", *Applied and Computational Harmonic Analysis*, **31** (1), pp. 44–58 (2011).
- R. Hadani, A. Singer, "Representation theoretic patterns in three dimensional Cryo-Electron Microscopy II – The class averaging problem", *Foundations of Computational Mathematics*, **11** (5), pp. 589–616 (2011).
- A. Singer, H.-T. Wu, "Vector Diffusion Maps and the Connection Laplacian", *Communications on Pure and Applied Mathematics*, **65** (8), pp. 1067–1144 (2012).
- A. S. Bandeira, A. Singer, D. A. Spielman, "A Cheeger inequality for the graph connection Laplacian", *SIAM Journal on Matrix Analysis and Applications*, **34** (4), pp. 1611–1630 (2013).
- Z. Zhao, A. Singer, "Rotationally Invariant Image Representation for Viewing Direction Classification in Cryo-EM", *Journal of Structural Biology*, **186** (1), pp. 153–166 (2014).
- C. J. Dsilva, B. Lim, H. Lu, A. Singer, I. G. Kevrekidis, S. Y. Shvartsman, "Temporal ordering and registration of images in studies of developmental dynamics", *Development*, **142**, 1717–1724 (2015).