# Vector Diffusion Maps and the Connection Laplacian 

Amit Singer<br>Princeton University, Department of Mathematics and PACM

ICIAM, Vancouver, BC, Canada July 22, 2011

## Motivating Problem: Cryo-Electron Microscopy



- Projection images $P_{i}(x, y)=\int_{-\infty}^{\infty} \phi\left(x R_{i}^{1}+y R_{i}^{2}+z R_{i}^{3}\right) d z+$ "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 $P_{1}, \ldots, P_{n}$.


## 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 $R_{1}, R_{2}, \ldots, R_{n} \in S O$ (3)
- Rotationally Invariant Distances (RID)

$$
d_{R I D}(i, j)=\min _{O \in S O(2)}\left\|P_{i}-O P_{j}\right\|
$$

- 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 $d_{\text {RID }}$ (noise aligns well, instead of underlying signal).


## Outliers: Small World Graph on $S^{2}$

- Define graph $G=(V, E)$ by $\{i, j\} \in E \Longleftrightarrow d_{R I D}(i, j) \leq \varepsilon$.

- Optimal rotation angles

$$
O_{i j}=\underset{O \in S O(2)}{\operatorname{argmin}}\left\|P_{i}-O P_{j}\right\|, \quad i, j=1, \ldots, n .
$$

- Triplet consistency relation - good triangles

$$
O_{i j} O_{j k} O_{k i} \approx I d
$$

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


## 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 $w_{i j}$ describing affinities between data points are accompanied by linear orthogonal transformations $O_{i j}$.

## 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: $O_{i j}=\operatorname{argmin}_{O \in O(d)}\left\|O-O_{i}^{T} O_{j}\right\|_{H S}$ (computed through the singular value decomposition of $O_{i}^{T} O_{j}$ ).



## Parallel Transport

- $O_{i j}$ approximates the parallel transport operator $P_{x_{i}, x_{j}}: T_{x_{j}} \mathcal{M} \rightarrow T_{x_{i}} \mathcal{M}$


## Vector diffusion mapping: $S$ and $D$

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

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

$n \times n$ 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:

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

and

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

## $D^{-1} S$ as an averaging operator for vector fields

- 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

$$
\left(D^{-1} S v\right)(i)=\frac{1}{\operatorname{deg}(i)} \sum_{j:(i, j) \in E} w_{i j} O_{i j} v(j)
$$

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

$$
\left\|\tilde{S}^{2 t}(i, j)\right\|_{H S}^{2}=\frac{\operatorname{deg}(i)}{\operatorname{deg}(j)}\left\|\left(D^{-1} S\right)^{2 t}(i, j)\right\|_{H S}^{2} .
$$

## Embedding into a Hilbert Space

- Since $\tilde{S}$ is symmetric, it has a complete set of eigenvectors $\left\{v_{l}\right\}_{l=1}^{n d}$ and eigenvalues $\left\{\lambda_{i}\right\}_{l=1}^{n d}$ (ordered as $\left|\lambda_{1}\right| \geq\left|\lambda_{2}\right| \geq \ldots \geq\left|\lambda_{n d}\right|$ ).
- Spectral decompositions of $\tilde{S}$ and $\tilde{S}^{2 t}$ :

$$
\tilde{S}(i, j)=\sum_{l=1}^{n d} \lambda_{l} v_{l}(i) v_{l}(j)^{T}, \quad \text { and } \quad \tilde{S}^{2 t}(i, j)=\sum_{l=1}^{n d} \lambda_{l}^{2 t} v_{l}(i) v_{l}(j)^{T}
$$

where $v_{l}(i) \in \mathbb{R}^{d}$ for $i=1, \ldots, n$ and $I=1, \ldots, n d$.

- The HS norm of $\tilde{S}^{2 t}(i, j)$ is calculated using the trace:

$$
\left\|\tilde{S}^{2 t}(i, j)\right\|_{H S}^{2}=\sum_{l, r=1}^{n d}\left(\lambda_{l} \lambda_{r}\right)^{2 t}\left\langle v_{l}(i), v_{r}(i)\right\rangle\left\langle v_{l}(j), v_{r}(j)\right\rangle .
$$

- The affinity $\left\|\tilde{S}^{2 t}(i, j)\right\|_{H S}^{2}=\left\langle V_{t}(i), V_{t}(j)\right\rangle$ is an inner product for the finite dimensional Hilbert space $\mathbb{R}^{(n d)^{2}}$ via the mapping $V_{t}$ :

$$
V_{t}: i \mapsto\left(\left(\lambda_{l} \lambda_{r}\right)^{t}\left\langle v_{l}(i), v_{r}(i)\right\rangle\right)_{l, r=1}^{n d} .
$$

## Vector Diffusion Distance

- The vector diffusion mapping is defined as

$$
V_{t}: i \mapsto\left(\left(\lambda_{l} \lambda_{r}\right)^{t}\left\langle v_{l}(i), v_{r}(i)\right\rangle\right)_{l, r=1}^{n d} .
$$

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

$$
d_{\mathrm{VDM}, t}^{2}(i, j)=\left\langle V_{t}(i), V_{t}(i)\right\rangle+\left\langle V_{t}(j), V_{t}(j)\right\rangle-2\left\langle V_{t}(i), V_{t}(j)\right\rangle
$$

- 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

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

for any $v \in \mathbb{R}^{n d}$. Therefore, $\lambda_{I} \in[-1,1]$. 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 $d_{V D M, t=2}$

Figure: $\mathrm{SNR}=1 / 64$ : 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 $d_{\text {RID }}$. Right: neighbors are post identified using vector diffusion distances.

## The Hairy Ball Theorem

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



## 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 $\left\{x_{i}\right\}_{i=1, \ldots, n}$ is independently uniformly distributed over $\mathcal{M}$. Let $K \in C^{2}\left(\mathbb{R}^{+}\right)$be a positive kernel function decaying exponentially, that is, there exist $T>0$ and $C>0$ such that $K(t) \leq C e^{-t}$ when $t>T$. For $\epsilon>0$, let $K_{\epsilon}\left(x_{i}, x_{j}\right)=K\left(\frac{\left\|\iota\left(x_{i}\right)-\iota\left(x_{j}\right)\right\|_{\mathbb{R} p}}{\sqrt{\epsilon}}\right)$. Then, for $X \in C^{3}(T \mathcal{M})$ and for all $x_{i}$ almost surely we have

$$
\lim _{\epsilon \rightarrow 0} \lim _{n \rightarrow \infty} \frac{1}{\epsilon}\left[\frac{\sum_{j=1}^{n} K_{\epsilon}\left(x_{i}, x_{j}\right) O_{i j} X_{j}}{\sum_{j=1}^{n} K_{\epsilon}\left(x_{i}, x_{j}\right)}-X_{i}\right]=\frac{m_{2}}{2 d m_{0}}\left(\left\langle\iota_{*} \nabla^{2} X\left(x_{i}\right), e_{I}\right\rangle\right)_{l=1}^{d},
$$

where $\nabla^{2}$ is the connection Laplacian, $X_{i} \equiv\left(\left\langle\iota_{*} X\left(x_{i}\right), e_{\rho}\right\rangle\right)_{l=1}^{d} \in \mathbb{R}^{d}$ for all $i,\left\{e_{l}\left(x_{i}\right)\right\}_{l=1, \ldots, d}$ is an orthonormal basis of $\iota_{*} T_{x_{i}} \mathcal{M}$, $m_{l}=\int_{\mathbb{R}^{d}}\|x\|^{I} K(\|x\|) \mathrm{d} x$, and $O_{i j}$ 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:$$
\begin{aligned}
& S^{2}: 6,10,14, \ldots \\
& S^{3}: 4,6,9,16,16, \ldots \\
& S^{4}: 5,10,14, \ldots \\
& S^{5}: 6,15,20, \ldots
\end{aligned}
$$


(a) $S^{2}$

(b) $S^{3}$

(c) $S^{4}$

(d) $S^{5}$

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$$
Z_{i j}=\left\{\begin{array}{cl}
\operatorname{det} O_{i j} & (i, j) \in E \\
0 & (i, j) \notin E
\end{array}\right.
$$

That is, $Z_{i j}=1$ if no reflection is needed, $Z_{i j}=-1$ if a reflection is needed, and $Z_{i j}=0$ 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 $D^{-1} Z$.

## Orientable Double Covering

Embedding obtained using the eigenvectors of the (normalized) matrix

$$
\left[\begin{array}{rr}
Z & -Z \\
-Z & Z
\end{array}\right]=\left(\begin{array}{rr}
1 & -1 \\
-1 & 1
\end{array}\right) \otimes Z
$$



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.

3D Reconstruction (not optimal)

- 27000 images of the 50 S ribosomal subunit were class averaged into 3000 classes using available software (Imagic).



## 3D Reconstruction (not optimal)



Refined

$N=3000$

$N=1500$

$N=750$

Figure: The refined model is from an Imagic reference-based alignment of the 27,000 particle data set used in this study and refined to $15 \AA$ resolution ( 0.5 Fourier shell correlation threshold criteria). The structures were also flipped about the $z$-axis such that their handedness is consistent with the X -ray structure of T . Steitz.

## Ongoing Research in cryo-EM

- Molecules with symmetries
- Heterogeneity problem
- Signal/Image processing


## Summary and Outlook

- 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, 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

